



Atatürk Üniversitesi
Açık ve Uzaktan Öğretim Fakültesi

İSTATİSTİĞE GİRİŞ



Sizler için içeriklerimizi sürekli güncelliyoruz.
En son güncellemeleri görmek için okutunuz.

Bu kitabın, basım, yayım ve saş hakları Atatürk Üniversitesi'ne aİ r. Bireysel öğrenme yaklaşımıyla hazırlanan bu kitabın bütün hakları saklıdır. Atatürk Üniversitesi'nin izni alınmaksızın kitabın tamamı veya bir kısmı mekanik, elektronik, fotokopi, manyeak kayıt veya başka şekillerde çoğalĞlamaz, basılamaz ve dağıĞlamaz.

Copyright ©Ekim 2025

The copyrights, publicaøons and sales rights of this book belong to Atatürk University. All rights reserved of this book prepared with an individual learning approach. No part of this book may be reproduced, printed, or distributed in any form or by any means, technical, electronic, photocopying, magneac recording, or otherwise, without the permission of Atatürk University.



ATATÜRK ÜNİVERSİTESİ
AÇIK VE UZAKTAN ÖĞRETİM FAKÜLTESİ

İSTATİSTİĞE GİRİŞ

ISBN: 978-625-8738-15-5



ATATÜRK ÜNİVERSİTESİ AÇIK VE UZAKTAN ÖĞRETİM FAKÜLTESİ YAYINI

ERZURUM

İÇİNDEKİLER

1. Temel Kavramlar <i>Prof.Dr. MEHMET SUPHİ ÖZÇOMAK</i>	<u>4</u>
2. İstatistik Verileri <i>Dr. EMRAH TALAŞ</i>	<u>21</u>
3. İstatistik Serileri <i>Dr. EMRAH TALAŞ</i>	<u>38</u>
4. Grafikler <i>Prof.Dr. MEHMET SUPHİ ÖZÇOMAK</i>	<u>60</u>
5. Parametrik Merkezî Eğilim Ölçüleri <i>Dr. EMRAH TALAŞ</i>	<u>79</u>
6. Parametrik Olmayan Merkezî Eğilim Ölçüleri <i>Dr. EMRAH TALAŞ</i>	<u>99</u>
7. Parametrik Değişkenlik Ölçüleri <i>Prof.Dr. MEHMET SUPHİ ÖZÇOMAK</i>	<u>117</u>
8. Parametrik Olmayan Değişkenlik Ölçüleri <i>Prof.Dr. MEHMET SUPHİ ÖZÇOMAK</i>	<u>138</u>
9. Şekil Ölçüleri <i>Prof.Dr. MEHMET SUPHİ ÖZÇOMAK</i>	<u>154</u>
10. İndeksler <i>Prof.Dr. ERKAN OKTAY</i>	<u>178</u>
11. İhtimal Teorisi <i>Prof.Dr. ERKAN OKTAY</i>	<u>196</u>
12. Kesikli Tesadüfi Değişkenler ve Dağılımları <i>Prof.Dr. ERKAN OKTAY</i>	<u>211</u>
13. Sürekli Tesadüfi Değişkenler ve Normal Dağılım <i>Prof.Dr. ERKAN OKTAY</i>	<u>229</u>
14. Kesikli İhtimal Dağılımların Normal Dağılıma Yaklaşımı <i>Prof.Dr. ERKAN OKTAY</i>	<u>246</u>

Editör

Prof.Dr. ERKAN OKTAY (1.,2.,3.,4.,5.,6.,7.,8.,9.)

TEMEL KAVRAMLAR



İÇİNDEKİLER

- İstatistiğin Tanımı
- İstatistiğin Konusu
- İstatistiğin Önemi
- İstatistiğin Kullanıldığı Yerler
- İstatistiğin Tarihsel Gelişimi
- Temel Kavramlar



HEDEFLER

- Bu üniteyi çalıştıktan sonra;
- İstatistiğin tanımlayabilecek,
- İstatistiğin kullanıldığı yerleri açıklayabilecek,
- Anakütle ve örnek kavramlarını açıklayabilecek,
- Tam sayım ile örnekleme arasındaki farkı açıklayabilecek,
- Anakütle ve örnek hacmi kavramlarını kıyaslayarak ifade edebilecek,
- Parametre ve istatistik kavramları hakkında bilgi sahibi olabileceksiniz.



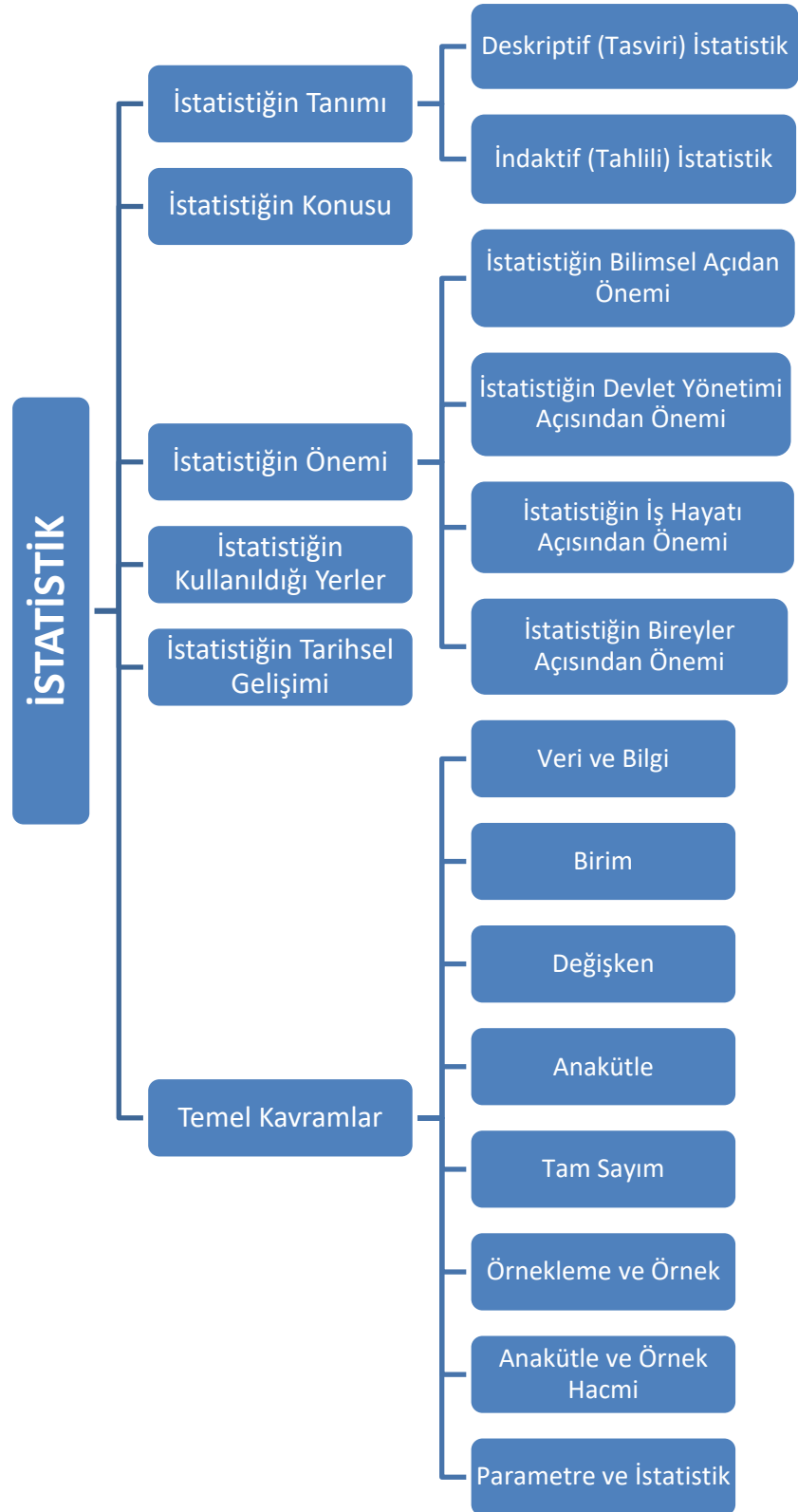
Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ

**Prof. Dr. Mehmet
Suphi ÖZÇOMAK**

ÜNİTE

1



GİRİŞ

İnsanlar topluluklar hâlinde yaşamaya başlayıp devletler kurulunca, onları yönetenlerin işlerini daha düzenli biçimde ele alıp yürütebilmeleri için bazı bilgilere ihtiyaç duyulmuştur. Bu amaçla, en ilkel toplumlarda bile toplum hayatına ilişkin bazı bilgiler toplanmış ve toplanan bu bilgiler başlangıçta toplumdaki birey sayısı, asker sayısı, hayvan sayısı, toplum için tehlikeli sayılabilecek düşmanların sayısı vb. konuları kapsamıştır. Zamanla bu bilgilere yenileri eklenmiş, kayıt şekilleri geliştirilmiş ve bu işleri özel olarak görevlendirilmiş kişiler veya kurumlar yürütmeye başlamıştır. Kısacası, toplum hayatına ilişkin bilgi sahibi olmadan devlet yönetmek son derece güç olduğundan, istatistik uygulamaları da devlet kadar eski bir geçmişe sahiptir.

İstatistik kelimesinin kökeninin latince devlet ya da durum anlamına gelen “status” kelimesinden geldiği yaygın bir görüş olmakla beraber kelimenin İtalyanca devlet adamı anlamına gelen “statista” kelimesinden ve Yunanca gözlem anlamına gelen “statizein” kelimesinden geldiğini ifade edenler de bulunmaktadır. 17. Yüzyılda Almanya’da devlet işleri ile ilgili toplanan sayısal veriler için ilk kez “statistik” kelimesi kullanılmıştır.

İstatistik kelimesi günlük hayatta farklı anlamlarda kullanılmaktadır. Televizyonda bir futbol müsabakasını izleyen bir taraftar için istatistik, maç esnasında yapılan faul sayısı, atılan korner sayısı, topa sahip olma oranları gibi değerleri; bir aile reisi için açıklanan aylık enflasyon oranlarını; başka bir birey için ülke nüfusu, ihracat değerleri, ithalat değerleri, inşa edilen konut sayıları gibi rakamları ifade ederken, akademik çalışma yapan bir bilim adamı için ise sayısal analizleri ifade etmektedir.

Aslında günlük hayatta kullandığımız istatistik kelimesi, Türkiye İstatistik Kurumu (TÜİK) ve/veya başka kurumlar tarafından herhangi bir konuyla ilgili toplanmış verileri ifade etmektedir. Bilindiği üzere TÜİK; nüfus, enflasyon rakamları, ithalat ve ihracat değerleri gibi birçok alanla ilgili veri toplayarak bültenler hâlinde yayınlamaktadır. Aslında istatistik, ziraattan iktisada, tıptan sosyolojiye, dış hekimliğinden eğitim bilimlerine kadar pek çok alanda yaygın kullanım alanı olan bir bilim dalıdır.



İstatistik; herhangi bir konuyla ilgili verilerin toplanması, düzenlenmesi, özetlenmesi, sunulması, uygun yöntemlerle analizi ve bu analizlerle elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilenir.

İSTATİSTİĞİN TANIMI

Uygulama alanı çok geniş olan istatistiğin farklı tanımları yapılmaktadır. İstatistik genel anlamı ile sayısal analizleri ifade etmektedir. İstatistiğin genel bir tanımı şu şekilde yapılabilir: *İstatistik; herhangi bir konuyla ilgili verilerin toplanması, düzenlenmesi, özetlenmesi, sunulması, uygun yöntemlerle analizi ve bu analizlerle elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilenir.* Tanımdan da anlaşıldığı üzere istatistikten söz edebilmek için ilk önce veriye ihtiyaç duyulmaktadır. Veriler elde edildikten sonra analize uygun hâle getirilmesi için düzenlenmesi gerekebilir. Veriler düzenlendikten sonra analiz için uygun istatistiksel yöntem veya yöntemler seçilir.

Daha dar anlamda ise istatistik terimi; verileri veya verilerden elde edilen aritmetik ortalama, standart sapma gibi değerleri ifade etmek için kullanılır. Bu şekilde; istihdam istatistikleri, kaza istatistikleri, ithalat ve ihracat istatistikleri gibi istatistiklerden söz edilebilir. Yukarıdaki genel tanıma göre istatistik, deskriptif (tasviri) ve indaktif (tahlili) istatistik olmak üzere ikiye ayrılmaktadır.



İstatistik, deskriptif (tasviri) ve indaktif (tahlili) istatistik olmak üzere ikiye ayrılmaktadır.

Deskriptif (Tasviri) İstatistik

Tasviri istatistik olarak da adlandırılan deskriptif istatistik, herhangi bir konuyla ilgili verilerin toplanması, düzenlenmesi, özetlenmesi ve söz konusu verilerin tablo ve grafikler hâlinde gösterilmesi ile ilgilenir. Frekans dağılımları, merkezî eğilim ölçüleri (aritmetik ortalama, mod, medyan, ...), dağılma ölçüleri (standart sapma, varyans, değişim aralığı, ...), çarpıklık ve basıklık ölçüleri gibi konular verilerin özetlenmesi ve tasviri ile ilgili olduğundan, deskriptif istatistiğin konularını teşkil etmektedir.

İndaktif (Tahlili) İstatistik

İndaktif istatistik, ilgilenilen konuyla ilgili tüm veriler arasından seçilen alt veriler kullanılarak analizlerin yapılması ve bu analizler ile elde edilen sonuçlar kullanılarak tüm birimler hakkında yorum yapılması ve bir karara bağlanması ile ilgilenir. Bu tanımdan yola çıkarak indaktif istatistik, tahlili istatistik olarak da adlandırılmaktadır. Örneklem teorisi, hipotez testleri, regresyon ve korelasyon analizleri gibi konular ise indaktif istatistiğin konularını teşkil etmektedir [1].

İSTATİSTİĞİN KONUSU

Rakamları eğip bükerek gerçeği farklı yansıtmak hiç de zor değildir. Bu nedenle istatistik konularının anlaşılabilmesi için olaylar arasındaki farklılıkların bilinmesi gerekir. Bir olayın kendi türünden olayları tam anlamıyla temsil edip edemediği bir ölçüt olarak kabul edildiğinde bu ölçüte göre olaylar “tipik olay” ve “kollektif olay” şeklinde ikiye ayrılabilir.

Birbirinin tam benzeri olan olaylara “tipik olay” adı verilmektedir. Olaylar birbirinin tam benzeri olunca bunlardan sadece biri bile oluşturdukları topluluğu temsil edebilir. Bu durumda çok sayıda olayı incelemeye gerek kalmaz. Sadece bir olayı ele alıp incelemek yeterli olmaktadır. Örneğin, belirli bir ısı derecesinde ki gazların hacimleriyle bunlara uygulanan basınç arasında belirli bir oran vardır. Laboratuvar koşulları değişmediği taktirde, bu oran bir deneyden diğerine değişmez. Deney ne kadar tekrar edilirse edilsin, sonuç hep aynı çıkar. Basınçla hacim arasındaki bu oran, bir deneyden diğerine değişmediği ve kesinlik taşıdığı için matematiksel bir formülle ifade edilebilmiştir.

Birbirine benzemeyen, bazı ortak yönleri olmakla birlikte genelde aralarında önemli farklılıklar da bulunan olaylara ise “kollektif (toplu) olay” denir. Olaylar birbirine benzemediği için bunlardan sadece biri oluşturdukları topluluğu temsil edemez. Örneğin, nüfus kollektif olaylar topluluğuna güzel bir örnektir. Bilindiği gibi, nüfusu oluşturan bireylerin, insan olmak ve aynı bölgede veya ülkede yaşamak vb. ortak özellikleri vardır. Buna karşılık; cinsiyet, yaş, medeni durum.

meslek, doğum yeri, boy uzunluğu, göz rengi, vb. diğer birçok bakımdan birimler arasında farklılıklar da görülür. Bu farklılıklar nedeniyle, bütün nüfusu temsil edebilecek tek bir insan bulunamaz. Durum diğer canlı veya cansızlar içinde farklı değildir. Nitekim bir buğday tanesi bir diğerine benzemez. Bir ağacın yaprakları arasında da şekil, renk, büyüklük, ağırlık ve su oranı bakımından farklılıklar vardır. Öte yandan bir işletmenin ürününü aynı kalitede ve aynı fiyatta satmasına ve reklamında herhangi bir değişiklik yapmamasına karşın satışlarının aynı olmadığı gözlemlenebilir. Çünkü ürünlerin satışı, çok sayıda ve kontrol altında tutulamayacak faktörlere bağlıdır.

Teoride tipik ve kolektif olaylar birbirinden kolayca ayırt edilebilirse de, gerçekte bu iki çeşit olay arasında kesin bir çizgi çizmek olanaksızdır. Kuşkusuz hukuk, iktisat, sosyoloji vb. sosyal bilimlerin alanındaki olaylar genellikle kolektif olaylardır. Ancak doğa bilimleri alanındaki olayların bütünüyle tipik olay olduğu söylenemez. Nitekim kolektif olaylara; fizik, astronomi, meteoroloji ve teknik alanlarda da rastlanmaktadır. Canlı varlıkları konu alan biyoloji, psikoloji ve tıp alanlarındaki olaylar da kolektif olaylardır [2].

İSTATİSTİĞİN ÖNEMİ



Tüm bilimsel araştırmaların en önemli aracı istatistiktir.

Hangi kararlar için hangi verilerin toplanacağı, hangi varsayımların yapılacağı, verilerin ne şekilde işleneceği, hangi tekniklerin uygulanacağı ve elde edilen sonuçların nasıl yorumlanacağını bilmek için istatistiğin öneminin incelenmesi gerekir. Bu amaçla bu bölümde; bilimsel açıdan, devlet yönetimi açısından, iş hayatı açısından ve son olarak bireyler açısından istatistiğin önemi anlatılacaktır.

İstatistiğin Bilimsel Açıdan Önemi

Çok çeşitli alanlara uygulanabilen istatistiğin bilimsel açıdan büyük önemi vardır. Örneğin; sosyoloji, intiharlar ve suçlar vb. konuları incelerken istatistikten geniş ölçüde yararlanır. Psikolojide test sonuçları istatistik yardımıyla analiz edilir. Filoloji ve edebiyatta bile istatistik kullanılmaktadır. Nitekim dillerin fonetik vasıflarını belirlemek amacıyla sesli harfler incelenirken, yazarların üslubundaki özellikleri kavramak için de belirli kelimelerin tekrarlama sayıları istatistik yardımıyla tespit edilebilir. Bu yöntem yardımıyla sahibi bilinmeyen edebî eserlerin sahipleri tespit edilebilir. Öte yandan, hangi toprak türünde en verimli hangi ürünün yetişeceğini, hangi besleme yönteminin tavukçulukta et veya yumurta verimini arttıracığı vb. konularda doğru bir sonuca varmak, bilimsel araştırmalara başvurmayı gerektirmektedir. *Tüm bilimsel araştırmaların en önemli aracı ise istatistiktir.* Herhangi bir alanda araştırmalar düzenlemek, yürütmek ve bu araştırmaların sonuçlarını uygun biçimde yorumlamak istatistik bilgisi gerektirir.

Bazı bilimlerin için istatistik o kadar büyük önem kazanmıştır ki, istatistiğin o bilimlere özgü ayrı kolları meydana gelmiştir. Örneğin; *sosyolojiye özgü istatistiğe sosyometri, antropolojiye özgü istatistiğe antropometri, biyolojiye özgü istatistiğe ise biyometri (biyoistatistik) adı verilmektedir.*



Sosyolojiye özgü istatistiğe sosyometri, antropolojiye özgü istatistiğe antropometri, biyolojiye özgü istatistiğe ise biyometri veya biyoistatistik adı verilmektedir.

Sosyal bilimler alanında istatistiğin ayrı bir önemi vardır. Örneğin; sosyal olay niteliğindeki köyden şehire göç, gecekondu sorunu, işsizlik problemleri ile hukuksal olay niteliğindeki suçlar hakkında yapılacak bilimsel araştırmalar, istatistiğe dayanmak, doğru sonuçlara ulaşabilmek için de gözlem ve sayısal analizlere yer vermek zorundadır. Öte yandan iktisadi olayların gelişimini, nedenlerini ve aralarındaki ilişkileri belirleyip buna göre gerekli önlemleri almak ve iktisat politikasını düzenlemek istatistik sayesinde olabilmektedir. Bu örnekler çoğaltılabilir. Özetle saymaya /ölçmeye veya kategorize etmeye elverişli konularda istatistiğin önemi günden güne artmaktadır. Bunun nedeni verilerin toplanmasına ilişkin, değişkenler arasındaki ilişkileri ortaya çıkarmayı amaçlayan ve elde edilen sonuçları değerlendirmeye yarayan bilimsel araştırma yöntem ve tekniklerini istatistiğin sağlamasıdır. Kısacası istatistik hem yöntem ve teknikler geliştiren bilim dalı hem de yöntem ve teknikler topluluğu sayılabilir.

İstatistiğin Devlet Yönetimi Açısından Önemi

İstatistiksel veriler devlete idari, sosyal ve ekonomik hayatın izlenmesinde rehberlik görevi yapmaktadır. Örneğin; hayat pahalılığının olup olmadığı, dış ticaretin gidişatından ülkenin yararlanıp yararlanmadığı, millî gelirin sosyal gruplar arasında ne şekilde bölündüğü istatistiksel veriler yardımıyla incelenir.

Devlet açısından büyük bir öneme sahip olan istatistiğe “devlet hayatının aynası” denildiği de görülmüştür. Devlet ile istatistik arasındaki bu sıkı ilişki en eski zamanlarda bile devletlerin kurulur kurulmaz arazi yazımları ve nüfus sayımlarına başvurmaları da bu durumu ortaya koymaktadır. Devletin ekonomiye müdahalesi arttıkça istatistiğe olan ihtiyacı da kuşkusuz artmaktadır. Örneğin planlı ekonomi, devlet yönetiminde çok daha çeşitli ve ayrıntılı istatistikî verilerin varlığını gerektirmektedir.

İstatistiğin İş Hayatı Açısından Önemi

Önceleri daha çok kamu kuruluşlarının bazı kamu hizmetlerinin yerine getirilebilmesi amacıyla yararlandığı istatistiğin uygulama alanı, zamanla hem kamu sektöründe hem de özel sektörde hızla gelişmiş ve genişlemiştir. İş hayatında öteden beri istatistikten yararlanıldığı hemen herkes tarafından bilinen bir gerçektir. Günümüzde birçok endüstriyel veya ticari büyük işletmeler çeşitli istatistiksel tekniklerden büyük ölçüde yararlanmaktadır. Çünkü modern girişimler, kendi işlemleri veya piyasa hakkında istatistiksel araştırmalara başvurmadığında rasyonel bir şekilde yönetilemez.

İyi bir satış politikası izleyebilmek için arz ve talebin gelişmelerini ve bu gelişmeler üzerinde etki yapan faktörleri ortaya koyacak istatistiksel araştırmalara ihtiyaç vardır. Bu araştırmalar arasında müşterileri çeşitli ürün ve hizmetleri tercih etmeye yönelten nedenleri belirlemek amacıyla yapılan istatistiksel araştırmalar büyük önem arz etmektedir. Öte yandan ücretler, ham madde, nakliye, depo masrafları, üretim vb. konular hakkında toplanan kayıtlardan yararlanılarak düzenlenecek bazı özel istatistiksel verilerin kıyaslanması ve analizi, işletmelerin yönetimi için nerdeyse zorunlu hale gelmiştir. Ayrıca yöneticiler, ülkenin genel

iktisadi koşullarıyla fiyat ve ücret düzeylerinin nasıl bir gelişme gösterdiğini izlemeye her zaman ihtiyaç duymaktadırlar.

Yöneticiler verecekleri kararlarda kendilerine yardımcı olacak istatistiksel tekniklere her zaman ihtiyaç duymaktadırlar. İstatistiksel teknikler birçok konuda yöneticilere yol göstermektedir. İstatistiksel tekniklerin kullanılmasındaki en önemli amaç ise belirsizlik altında karar verebilmektir.

İstatistiğin Bireyler Açısından Önemi

İstatistiğin devlet ve iş hayatı bakımından olduğu gibi bireyler ve hatta aileler açısından da önemi vardır. Örneğin; aile bütçesinin düzenlenmesi zaman zaman vücut ağırlığının, hastalık durumunda vücut sıcaklığının, nabız ve tansiyonun ölçülmesi, gereken önlemlerin alınması ve bunların sonuçlarının görülmesi için istatistiki bilgilere ihtiyaç duyulur. Aslında bunlar, bireyin günlük hayatta karşılaştığı “Ne kadar?”, “Ne zaman?”, “Nerede?”, “Nasıl?”, ve “Kaç tane?” vb. soruları çoğu kez sayısal ifadeler kullanarak cevaplanmasından başka bir şey değildir. İstatistik, en azından bireye çevresinde olup bitenleri anlama ve bunları başkalarına anlatmada yardımcı olmaktadır. Öte yandan hayat pahalılığı hakkında fikir veren fiyat istatistikleri, hastalık ve ölümleri, trafik kazalarını, bu kazalarda ölen ve yaralananları ve seçim sonuçlarını gösteren istatistikler, siyasal ve sosyal bazı sorunlar hakkındaki kamuoyu araştırmaları ve bunların sonuçları geniş halk kitleleri tarafından büyük bir ilgi ile izlenir [2].

İSTATİSTİĞİN KULLANILDIĞI YERLER



İstatistik; pazarlama, muhasebe, işletme, ekonomi, sosyoloji, mühendislik, tıp, ziraat vb. tüm bilim dallarında kullanılmaktadır.

İstatistik; pazarlama, muhasebe, işletme, ekonomi, sosyoloji, mühendislik, tıp, ziraat vb. tüm bilim dallarında kullanılmaktadır. Bir işletmede, geçmişteki etkinliklerle ilgili veriler toplanıp, bu verilerden gelecekle ilgili tahminler yapılabilir. Örneğin; geçmişteki ürün, fiyat, dağıtım, satış miktarları, gelir, gider vb. ilgili istatistiksel veriler toplanarak, bu konularla ilgili tahminler yapılabilir.

Bir işletmede, makinelerin ürettiği ürünler ölçülüp tartılarak standartlara uygun olup olmadığı belirlenir ve sonuca göre gerekirse makineler üzerinde ayarlamalar yapılır. Örneğin, bir dolum makinesinin doldurduğu şişelerin doluluk oranları ölçülerek standartlara uygun doldurulup doldurulmadığı belirlenerek, gerek görüldüğü takdirde dolum makinesinin ayarları yapılır.

Bir işletmenin pazarlama bölümünde, yeni ürün tasarımlarından hangisinin veya hangilerinin daha çok satılacağı ile ilgili veriler toplanıp, bu verilere uygun pazarlama teknikleri oluşturulur. İşletmenin üreteceği ürünlere benzer ürünlerin pazar fiyat istatistikleri belirlenip, fiyatlandırma yapılırken bu veriler göz önünde bulundurulur.

Siyasi partiler, seçimleri kazanmak için yaptırdıkları araştırmalarda seçmenlerin partilerle, liderlerle, ülkenin ekonomik ve toplumsal konularıyla ilgili görüşlerini öğrenerek, bu bulgulara uygun stratejiler belirlerler [3].

Tıp alanında, bir hastalık türü için yeni geliştirilecek bir ilaç öncelikle deney hayvanları üzerinde denenerek elde edilen veriler istatistiksel olarak analiz edildikten sonra söz konusu ilacın etkili olup olmayacağına karar verilir.

Yine ziraat alanında çalışma yapan bir araştırmacı üretilen bir ürünün verimini arttırmak için arazide uygun ekim zamanı, gübreleme, sulama tekniği vb. durumları deneyerek elde ettiği verileri istatistiksel olarak analiz ettikten sonra uygulanacak en uygun ekim zamanı, uygun gübre miktarı, uygun sulama tekniği vb. konulara karar verebilir.



İlk istatistiki çalışmaların Mısır, Çin ve Roma'da yapıldığı yapılan araştırmalarla belirlenmiştir.

İSTATİSTİĞİN TARİHSEL GELİŞİMİ

İstatistiğin tarihsel gelişimine bakıldığında; devletlerin savaşabilecek insan gücünü belirlemek amacıyla nüfus sayımları yapılarak ilk istatistiki çalışmaların temeli atılmıştır. Bunun yanında arazi ve servet sayımlarının yapıldığına da rastlanmaktadır. Ancak yapılan tüm sayımların bugünkü anlamda bir sayım olarak kabul edilmesi beklenemez.

İlk istatistiki çalışmaların Mısır, Çin ve Roma'da yapılan araştırmalarla belirlenmiştir. Örneğin, Mısır'da piramitlerin yapımı ile ilgili iş gücünün belirlenmesi için nüfus sayımı yapıldığı bilinmektedir. Ancak nüfus sayımlarına en fazla önemi Romalılar vermiştir. Orta çağa gelindiğinde devletlerin otoritesinin zayıflamasına paralel olarak yapılan sayımlarda da gerilemeler baş göstermiştir. Roma imparatorluğunun yıkılması ile beraber istatistiki çalışmalar da durmuştur. Orta çağda duraklayan istatistiki çalışmalar, sonraki çağlarda güçlü devletlerin kurulması ile yeniden hız kazanmıştır. 17. yüzyılda önce Fransa'da özel bir büro kurularak, maliye ve dış ticaret alanında istatistikler toplanmıştır. Daha sonra 1790 yılında dünyada ilk kez bugünkü anlamda nüfus sayımı Amerika Birleşik Devletleri'nde yapılmıştır. Yine ciddi manada ilk sanayi sayımı 1810 yılında, ilk tarım sayımı da 1840 yılında yapılmıştır. Avrupa'da ise ilk kez bugünkü anlamda nüfus sayımı 1801 yılında Fransa'da yapılmıştır.

Almanya'da bazı üniversitelerde 17. yüzyılın ilk yarısında "Devletlerin Özellikleri" adı ile yeni bir ders okutulmaya başlanmış ve bu derste ülkelerin idari, mali ve askeri örgütleri, demografik ve ekonomik alanlardaki rakamlar ile ilgili bilgiler verilmiştir. Burada istatistik kelimesi kullanılmakla beraber bu kullanım daha çok istatistiğin rakam anlamındaki tanımına uygun olarak kullanılmıştır.

İlhanlılar, Selçuklular ve Osmanlılar döneminde bugünkü anlamından uzak olsa da zaman zaman nüfus, arazi ve hayvan sayımları yapılmıştır. Bu sayımların hemen hemen hepsi kısmi nitelikte sayımlar olup, hepsi de savaşacak asker gücünü ve toplanacak vergiyi belirlemeye yönelik yapılmıştır. Askere alınabilecek nüfusu belirlemeye yönelik olarak ilk nüfus sayımı 1831 yılında yapılan kısmi nüfus sayımıdır. Daha sonra vatandaşlara kimlik belgesi verilmesi amacıyla 1844 yılında bir nüfus sayımı yapılmıştır.

Cumhuriyetin ilanı ile birlikte birçok alanda yapılan çalışmalara istatistik konusunda da rastlanmaktadır. Cumhuriyet tarihinin ilk nüfus sayımı 1927 yılında yapılmıştır. 25 Nisan 1926 tarihinde Başbakanlığa bağlı olarak "Merkezi İstatistik



Veri, incelenen birimlerin çeşitli özelliklerine ait sembolik değerlerdir.

Dairesi” kurulmuştur. 1930 yılında Merkezi İstatistik Dairesi’nin tüm görev ve sorumlulukları yeniden düzenlenerek “İstatistik Umum Müdürlüğü’ne” bırakılmıştır. 1962 yılında Başbakanlığa bağlı olarak “Devlet İstatistik Enstitüsü (DİE)” kurulmuştur. Devlet İstatistik Enstitüsü’nün adı 2005 yılında yapılan kanun değişikliği ile “Türkiye İstatistik Kurumu (TÜİK) olarak değiştirilmiştir [4].

TEMEL KAVRAMLAR

Bu bölümde istatistikte çok fazla karşılaşılabilecek temel kavramlar ele alınmıştır. İstatistiğin iyi anlaşılması için istatistikte çok sık kullanılan bazı kavramların anlamının, birbirleri ile ilişkilerinin ve farklılıklarının iyi bilinmesi gerekir. Bazı kavramlar eşit veya aynı anlamda olmasalar da çoğu zaman birbirlerinin yerine kullanılmaktadırlar.

Veri ve Bilgi

İstatistikte birbirinin yerine çok kullanılan iki kavram; Veri ve bilgi kavramlarıdır. *Veri, incelenen birimlerin çeşitli özelliklerine ait sembolik değerlerdir. Semboller yerine çoğunlukla rakamlar kullanılır. Bilgi ise birimlerden elde edilen verilerin işlenerek anlamlı hale getirilmiş halidir. Bir başka ifade ile bilgi, ham verilerin işlenmiş halidir.* Elde edilen veriler birtakım işlemlere tabi tutulduktan sonra bilgi elde edilmektedir. Dolayısıyla veriler hacim olarak (çok sayıda) büyük, değer olarak küçük iken; bilgi, aksine hacim olarak küçük iken değer olarak büyüktür.



Bilgi, birimlerden elde edilen verilerin işlenerek anlamlı hale getirilmiş halidir.

Örneğin, bir üniversitede okuyan öğrencilere bitirdikleri lise türü sorularak elde edilen cevaplar veriyi ifade etmektedir. Cevapları inceleyerek sınıftaki öğrencilerin örneğin %60’ının kolej mezunu olduğu sonucuna varıldığında, elde edilen bu sonuç ise bilgiyi ifade etmektedir. Bu örnekte her öğrenciden alınan bir cevap olduğuna göre elde öğrenci sayısı kadar veri olacaktır. Bu çok sayıdaki veri ile veriler üzerinde bir düzenleme yapmadan bir sonuca ulaşmanın imkânı yoktur. Bu nedenle de eldeki verilerin fazlaca bir değeri de olmayacaktır. Ancak bu verilerin düzenlenmesi ile tüm verilerin kolej mezunu olan veya olmayan gibi iki gruba ayrıldığında; öğrencilerin %60’ının kolej mezunu olması sonucu elde edildiğinde hacim olarak çok sayıdaki veriden tek bir sonuç elde edilecektir. Böylece çok sayıdaki veri tek bir bilgiye dönüştürülmüş olunacaktır. Burada anlam ifade etmeyen ve bu nedenle de değersiz olan veriler anlamlı hale gelmiş ve değeri de artmış olacaktır [4].

Birim

Birim, hakkında ölçüm yaparak bilgi elde ettiğimiz kişi ya da eşya biçimindeki nesnelerdir. Bir başka ifade ile birim, istatistiki bilgilerin toplandığı insan, bitki, hayvan, eşya vb. yapıların her biridir.

Herhangi bir konuda araştırma yapılırken ilk olarak istatistiki verilerin kimden, nereden, ne zaman toplanacağına karar verilir. Bu aşamada verilerin hangi birimden toplanacağına karar verilir. Örneğin, öğrencilerin ders çalışma alışkanlıkları üzerinde bir araştırma yapılacaksa burada öğrencilerin her biri birimi



Birim, hakkında ölçüm yaparak bilgi elde ettiğimiz kişi ya da eşya biçimindeki nesnelerdir.

oluşturmaktadır. Öte yandan buğdayın verimi ile ilgili araştırma yapan bir araştırmacı istatistiki birim olarak buğdayı almaktadır.

Değişken



Bir gözlem biriminden diğerine farklılık gösteren nesne özelliklerine değişken denir.

Bir gözlem biriminden diğerine farklılık gösteren nesne özelliklerine değişken denir. Başka bir tanımla, farklı birimler için farklı değerler alabilen ve incelenen tüm özelliklere değişken adı verilir. Değişken; bir şahsın, grubun veya çevrenin bir durumundan diğerine değişebilen veya bir farklılık gösterebilen herhangi bir özelliğidir. Cinsiyet, ağırlık, meslek, boy, gelir, not vb. örnekler değişken olarak adlandırılır.

Değişken kavramından kasıt, araştırma yapılan o özelliğin (değişkenin) sürekli değişmesi değildir. Burada ifade edilmek istenen değişkenin birimden birime farklılık göstermesidir. Örneğin, öğrencilerin boyları üzerinde yapılan bir araştırmada değişkenimiz boy olacağına göre tüm öğrencilerin boyları aynı değildir. Öğrenciden öğrenciye boy uzunluğu değişmektedir. Aynı şekilde tüm öğrencilerin cinsiyetleri, isimleri, memleketleri, kiloları, düşünceleri farklılık göstermektedir [5].

Anakütle



Bir istatistiki araştırmada, araştırmaya konu olan bütün birimlere anakütle denir.

Bir istatistiki araştırmada, araştırmaya konu olan bütün birimlere anakütle denir. Anakütlenin çerçevesi yapılacak araştırmadan araştırmaya değişiklik göstermektedir. İktisadi ve İdari Bilimler Fakültesi'nin birinci sınıfında okuyan öğrencilerinin sınav notları ilgili bir araştırma yapıldığında söz konusu fakültede okuyan birinci sınıf öğrencilerinin tamamı anakütleyi oluştururken, fakültede okuyan tüm öğrencilerinin sınav notları ilgili bir araştırma yapıldığında ise fakültede okuyan tüm öğrenciler anakütleyi oluşturmaktadır.

Çerçeve daha da genişletilerek üniversitede okuyan tüm öğrenciler bir anakütle olabileceği gibi Türkiye'deki üniversitelerin tamamında okuyan öğrenciler de bir anakütleyi teşkil edebilir. Üniversite öğrencileri bir anakütleyi ifade ederken, ilkokul öğrencileri, liselerde okuyan öğrenciler, herhangi bir kitabı okuyan bireyler, herhangi bir konuya ait belli bir düşünceye sahip şahıslar, devlet dairelerinde çalışan memurlar, bir ilde yaşayan hane halkları ve daha birçok birim anakütleyi ifade edebilir.

Bazı durumlarda üzerinde araştırma yapılan anakütle sayılamayacak kadar birim ihtiva edebilir. Örneğin, Karadeniz'deki hamsiler üzerinde bir araştırma yapılacaksa Karadeniz'deki tüm hamsiler anakütleyi ifade etmektedir ki hamsilerin tamamını saymamız imkânsızdır. Bu durumda karşımıza sınırlı ve sınırsız anakütle kavramları çıkmaktadır.

Araştırma konusu ile ilgili tüm birimlerin sayısı bilinen anakütlelere "sınırlı anakütle" denir. Araştırma konusu ile ilgili birimlerin çerçevesi çizilebiliyorsa bu anakütle sınırlı anakütledir. Örneğin İktisadi ve İdari Bilimler Fakültesinin birinci sınıfında okuyan öğrenciler ile ilgili bir araştırma yapılacaksa birinci sınıfta okuyan öğrencilerin sayısı bilindiği için söz konusu anakütle sınırlı anakütledir.



Araştırma konusu ile ilgili birimlerin çerçevesi çizilebiliyorsa bu anakütle sınırlı anakütledir.



Araştırma konusu ile ilgili birimlerin çerçevesi çizilemediği durumlar, sınırsız anakütleyi ifade etmektedir.

İktisadi ve İdari Bilimler Fakültesi'nin birinci sınıfında okuyan öğrenciler, bir bölümde okuyan öğrenciler, fakültede okuyan tüm öğrenciler, Atatürk Üniversitesi'nde okuyan tüm öğrenciler ve biraz daha genişletilirse Türkiye'deki üniversitelerin tamamında okuyan öğrencilerin sayıları bilindiğinden sınırlı anakütleye örnek teşkil ederler. Bu örnekler genişletilebilir. Bir ildeki bina sayısı, trafik kazalarının sayısı, doğum ve ölüm sayıları vb. birçok konu sınırlı anakütleye örnek teşkil etmektedir.

Araştırma konusu ile ilgili tüm birimlerin sayısı tespit edilemeyen anakütlelere "sınırsız anakütle" denir. Araştırma konusu ile ilgili birimlerin çerçevesi çizilemediği durumlar, sınırsız anakütleyi ifade etmektedir. Örneğin Karadeniz'deki hamsiler üzerine yapılan bir araştırmada; Karadeniz'deki hamsilerin tamamı sayılamayacağı için söz konusu anakütle sınırsız anakütledir.

Karadeniz'deki hamsilerin sayısı sınırsız anakütle için örnek teşkil ettiği gibi, bir ülkede roman okuyan şahısların sayısı, bir ildeki herhangi bir yemeği seven insanların sayısı, bir ülkede seyahat etmeyi sevenlerin sayısı gibi konularda anakütleyi oluşturan birimlerin sayısı tespit edilemediği için sınırsız anakütle için örnek teşkil etmektedir [6].



Anakütle ile ilgili bilgi toplanmak istendiğinde tüm birimlerin teker teker incelenmesi gerekmektedir. Bu işleme tam sayım adı verilmektedir.

Tam Sayım

Bir önceki başlıkta bir istatistiki araştırmada, araştırmaya konu olan bütün birimler anakütle ile ifade edilmişti. Birim ise anakütleyi oluşturan en küçük parçadır. Örneğin; bir ilde yaşayan ailelerin mutfak giderleri ile ilgili bir araştırmada, söz konusu ilde yaşayan ailelerin tamamı anakütleyi oluştururken; bu ilde yaşayan her bir aile ise anakütlenin birimlerini oluşturmaktadır.

Anakütle ile ilgili bilgi toplanmak istendiğinde tüm birimlerin teker teker incelenmesi gerekmektedir. Bu işleme tam sayım adı verilmektedir. Anakütle birim sayısı çok fazla olması durumunda tüm birimlerin incelenmesi fazla zaman alabileceği gibi aynı zamanda masraflı da olacaktır. İstatistiki araştırmalar genellikle bir fayda elde etmek için yapılır ve yapılan her bir araştırma belirli bir sürede bitirilemiyor ve yapılan masraflar faydayı aşıyorsa araştırma yapmanın anlamı kalmamaktadır. İfade edilen bu nedenlerden dolayı tam sayım işlemi çoğu zaman yapılamamaktadır [7].

Örnekleme ve Örneklem

Araştırmaya konu olan bütün birimlerin tamamına ulaşmak mümkün veya gerekli olmayabilir. Yukarıda ifade edildiği gibi mümkün olsa bile zaman ve maliyet gibi bazı kısıtlayıcılar nedeniyle tüm anakütleye ulaşmak mümkün olmayabilir. Bu gibi durumlarda *anakütleden çeşitli yöntemlerle anakütle birim sayısından daha az sayıda birimlerin seçilme işlemine örnekleme denir. Anakütleden örnekleme yardımıyla seçilen birimler ise "örneklem" olarak ifade edilmektedir.* Örneğin; bir üniversitede okuyan öğrenciler üzerinde bir araştırma yapıldığında bu üniversitedeki tüm öğrenciler anakütleyi teşkil ederken, bu öğrenciler arasından tesadüfi olarak seçilen 100 öğrenci ise örneklemi teşkil etmektedir [1].



Anakütleden çeşitli yöntemlerle anakütle birim sayısından daha az sayıda birimlerin seçilme işlemine örnekleme denir.

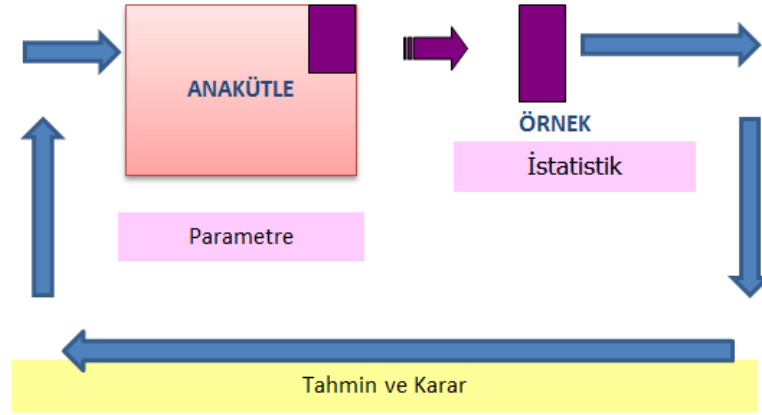
Anakütle ve Örneklem Hacmi



Anakütle hacmi, anakütleyi oluşturan birimler topluluğudur ve genellikle N ile gösterilir. Örneklem hacmi ise örneğe seçilen birim sayısıdır ve n ile gösterilir.

Anakütle hacmi, anakütleyi oluşturan birimler topluluğudur ve genellikle N ile gösterilir. Örneklem hacmi ise örneğe seçilen birim sayısıdır ve n ile gösterilir. Örneğin; Atatürk Üniversitesi'nde okuyan öğrencilerin kitap okuma alışkanlıkları ile ilgili bir araştırma yapılacaksa, üniversitede okuyan yaklaşık 496000 öğrenci anakütle hacmini (N) ifade ederken, bu öğrenciler arasından tesadüfi yöntemlerle seçilen 1000 öğrenci ise örneklem hacmini (n) ifade etmektedir. Eğer araştırma İktisadi ve İdari Bilimler Fakültesi için yapılıyorsa, fakültede okuyan 8000 öğrenci anakütle hacmini, bu öğrenciler içerisinde tesadüfi olarak seçilen 100 kişilik öğrenci grubu ise örneklem hacmini ifade etmektedir. Bu tip bir örneğe *şans örneği* denir.[1]

Örnekleme yapmanın temel amacı anakütleden seçilen örnekler yardımıyla anakütle hakkında bilgi elde etmektir. Örnekten elde edilen veriler kullanılarak elde edilen sonuçlar bütün anakütleye genelleştirilebilir. Bu durumunun daha iyi anlaşılması için anakütle ve örnek kavramları Görsel 1.1 üzerinde açıklanmaya çalışılacaktır.



Görsel 1.1. Anakütle ve Örneklem Kavramları

Görsel 1.1 incelendiğinde anakütleden tesadüfi yöntemlerle anakütle birim sayısından daha az sayıda seçilen örneklem yardımıyla tahminler yapılarak anakütle hakkında karar verilir. Şekilde verilen parametre ve istatistik kavramları aşağıda açıklanmıştır.

Parametre ve İstatistik

Anakütlerdeki bütün birimler kullanılarak hesaplanan ölçülere parametre adı verilir. Örneğin; bir fakültede okuyan öğrencilerin istatistik dersinden aldıkları notlar ile ilgili bir araştırma yapıldığında, istatistik dersini alan tüm öğrenciler anakütleyi temsil etmektedir. Anakütleyi temsil eden tüm öğrencilerin istatistik dersinden aldıkları notların ortalaması hesaplandığında elde edilen değer parametre değerini ifade etmektedir.

Anakütleyi temsil etme gücüne sahip bir örneklemdeki verilerden hesaplanan ölçülere istatistik adı verilir. Yukarıda verilen örnekte istatistik dersini alan tüm öğrenciler arasından tesadüfi olarak seçilen 30 öğrencinin not ortalaması



Anakütlerdeki bütün birimler kullanılarak hesaplanan ölçülere parametre adı verilir.



Anakütleyi temsil etme
gücüne sahip bir
örneklemden
verilerden hesaplanan
ölçülere istatistik adı
verilir.

istatistik değeri ifade etmektedir. İstatistiki bilgilerin hesaplanması daha çok tasviri istatistiğin konusudur. Eldeki istatistik değerlerini kullanarak anakütle parametreleri hakkında bir kısım yargılara varmak tahlili istatistik veya istatistik analizin konusunu teşkil etmektedir [8].



Bireysel Etkinlik

- İstatistiğin bilimsel açıdan önemini tartışınız.
- Tipik ve kolektif olaylara örnekler vererek karşılaştırınız.



Özet

- İstatistik kelimesinin kökeni Almanca olup “devlet” anlamına gelmektedir. İstatistik kelimesi günlük hayatta farklı anlamlarda kullanılmaktadır.
- İstatistik; herhangi bir konuyla ilgili verilerin toplanması, düzenlenmesi, özetlenmesi, sunulması, uygun yöntemlerle analizi ve bu analizlerle elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilidir. İstatistik, deskriptif (tasviri) ve indaktif (tahlili) istatistik olmak üzere ikiye ayrılmaktadır.
- Tasviri istatistik olarak da adlandırılan deskriptif istatistik, herhangi bir konuyla ilgili verilerin toplanması, düzenlenmesi, özetlenmesi ve söz konusu verilerin tablo ve grafikler hâlinde gösterilmesi ile ilgilidir. Frekans dağılımları, merkezî eğilim ölçüleri (aritmetik ortalama, mod, medyan,), dağılım ölçüleri (standart sapma, varyans, değişim aralığı...), çarpıklık ve basıklık ölçüleri gibi konular verilerin özetlenmesi ve tasviri ile ilgili olduğundan, deskriptif istatistiğin konularını teşkil etmektedir.
- İndaktif istatistik, ilgilenilen konuyla ilgili tüm veriler arasından seçilen alt veriler kullanılarak analizlerin yapılması ve bu analizler ile elde edilen sonuçlar kullanılarak tüm birimler hakkında yorum yapılması ve bir karara bağlanması ile ilgilidir. Bu tanımdan yola çıkarak indaktif istatistik, tahlili istatistik olarak da adlandırılmaktadır. Örneklem teorisi, hipotez testleri, regresyon ve korelasyon analizleri gibi konular ise indaktif istatistiğin konularını teşkil etmektedir.
- Olaylar “tipik olay” ve “kollektif olay” şeklinde ikiye ayrılabilir. Birbirinin tam benzeri olan olaylara “tipik olay” adı verilmektedir. Birbirine benzemeyen, bazı ortak yönleri olmakla birlikte genelde aralarında önemli farklılıklar da bulunan olaylara ise “kollektif (toplu) olay” denir.
- Veri incelenen birimlerin çeşitli özelliklerine ait sembolik değerlerdir. Semboller yerine çoğunlukla rakamlar kullanılır. Bilgi ise birimlerden elde edilen verilerin işlenerek anlamlı hale getirilmiş halidir.
- Birim, hakkında ölçüm yaparak bilgi elde ettiğimiz kişi yada eşya biçimindeki nesnelerdir. Bir başka ifade ile birim, istatistiki bilgilerin toplandığı insan, bitki, hayvan, eşya, öğrenci, seçmen, masa, ağaç vb. yapıların her biridir.
- Bir gözlem biriminden diğerine farklılık gösteren nesne özelliklerine değişken denir. Başka bir tanımla, farklı birimler için farklı değerler alabilen ve incelenen tüm özelliklere değişken adı verilir. Değişken; bir şahsın, grubun veya çevrenin bir durumundan diğerine değişebilen veya bir farklılık gösterebilen herhangi bir özelliğidir. Cinsiyet, ağırlık, meslek, boy, gelir, not vb. örnekler değişken olarak adlandırılır.
- Bir istatistiki araştırmada, araştırmaya konu olan bütün birimlere anakütle denir. Sınırlı anakütle, İktisadi ve İdari Bilimler Fakültesi’nin birinci sınıfında okuyan öğrencilerinin sınav notları ilgili araştırma örneğinde olduğu gibi bu öğrencilerin sayısı tespit edilebildiği anakütleleri ifade etmektedir. Kısacası araştırma konusu ile ilgili birimlerin çerçevesi çizilebiliyorsa bu anakütle sınırlı anakütledir. Sınırsız anakütle kavramı ise, Karadeniz’deki hamsiler ile ilgili yapılan araştırma örneğindeki gibi hamsilerin tamamının sayısı tespit edilemediği durumlarda karşımıza çıkmaktadır. Araştırma konusu ile ilgili birimlerin çerçevesi çizilemediği durumlar sınırsız anakütleyi ifade etmektedir.
- Anakütle hacmi, anakütleyi oluşturan birimler topluluğudur ve genellikle N ile gösterilir. Örnek hacmi ise örneğe seçilen birim sayısıdır ve n ile gösterilir.
- Anakütledeki bütün birimler kullanılarak hesaplanan ölçülere parametre adı verilir. Anakütleyi temsil etme gücüne sahip bir örnekteki verilerden hesaplanan ölçülere istatistik adı verilir.

DEĞERLENDİRME SORULARI

1. İstatistiğin yaptığı işler arasında aşağıdakilerden hangisi bulunmaz?
 - a) Verilerin toplanması
 - b) Verilerin düzenlenmesi
 - c) Verilerin özetlenmesi
 - d) Verilerin analizi
 - e) Verilerin saklanması
2. Deskriptif (tasviri) istatistiğin konusu aşağıdakilerden hangisi değildir?
 - a) Medyan
 - b) Mod
 - c) Hipotez testi
 - d) Standart sapma
 - e) Değişim aralığı
3. İndaktif (tahlili) istatistiğin konusu aşağıdakilerden hangisidir?
 - a) Örneklem teorisi
 - b) Aritmetik ortalama
 - c) Varyans
 - d) Değişim aralığı
 - e) Mod
4. İstatistiki çalışmalarda, hakkında bilgi elde edilmek istenen ve araştırmaya konu olan bütün birimlere ne ad verilir?
 - a) Varyans
 - b) Anakütle
 - c) Mod
 - d) Örneklem
 - e) Örneklem
5. Bir üniversitedeki öğrenciler içerisinde cinsiyet ve kitap okuma alışkanlığı arasında bir ilişkinin bulunup bulunmadığının sınındığı bir çalışmada anakütle aşağıdakilerden hangisidir?
 - a) Üniversitedeki okuyan erkek öğrenciler
 - b) Üniversitedeki kitap okuyan kız öğrenciler
 - c) Bir fakülte'deki kitap okuyan öğrenciler
 - d) Üniversitedeki okuyan tüm öğrenciler
 - e) Üniversitedeki kitap okumayan öğrenciler

6. Sınırlı anakütleye aşağıdakilerden hangisi örnek olabilir?
- a) Bir ilde A marka deterjan kullanan ailelerin sayısı
 - b) Bir ülkede roman okuyanların sayısı
 - c) Bir sınıftan seçilen 10 öğrenci
 - d) Doğu Anadolu Bölgesi'nde kayıtlı araç sayısı
 - e) Doğu Anadolu Bölgesi'nde yaşayan canlıların sayısı
7. Aşağıdakilerden hangisi sınırsız anakütleye örnek olmaz?
- a) Bir ormandaki canlılar
 - b) Bir göldeki balıklar
 - c) Bir okulda okuyan birinci sınıf öğrencileri
 - d) Bir ilde bilgisayar kullanan şahıslar
 - e) Bir ülkede seyahat etmeyi seven şahıslar
8. Anakütle ile ilgili bilgi toplamak istendiğinde bütün birimlerin teker teker incelenmesine ne ad verilir?
- a) Örneklem
 - b) Tam sayım
 - c) Örnek
 - d) Anakütle
 - e) Parametre
9. Anakütleden tesadüfi yöntemlerle anakütle birim sayısından daha az sayıda birimlerin seçilme işlemine ne ad verilir?
- a) Tam sayım
 - b) Anakütle
 - c) Örneklem
 - d) Örneklem
 - e) İstatistik
10. Anakütleden örneklem yardımıyla seçilen alt birimlere ne ad verilir?
- a) Anakütle alt birimi
 - b) Tam sayım
 - c) Örneklem
 - d) Parametre
 - e) İstatistik

Cevap Anahtarı

1.e, 2.c, 3.a, 4.b, 5.d, 6.d, 7.c, 8.b, 9.d, 10.c

YARARLANILAN KAYNAKLAR

- [1] Başar, Alaaddin, Oktay, Erkan. (2018). Uygulamalı İstatistik – I: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap ve Kirtasiye.
- [2] Serper, Ö. (2014). Uygulamalı İstatistik (7. baskı). Bursa: Ezgi Kitabevi.
- [3] Tekin, V. N. (2019). İstatistiğine Giriş: Kavramlar-Konu Anlatımları-Çözümlü Problemler (4. baskı). Ankara: Seçkin Yayınları.
- [4] Bülbül, Ş. (2013). Tanımlayıcı İstatistik (3. baskı). İstanbul: Der Yayınları.
- [5] Erilli, N. A. (2017). İstatistik-1 (2. baskı). Ankara: Seçkin Yayınları.
- [6] Berenson, M. L., D. M. Levine (1996). Basic Business Statistics: Concepts and Applications (8. Baskı). New Jersey: Prentice Hall.
- [7] Armutlulu, İ. H. (1999). İşletme İstatistiğine Giriş. İstanbul: Alfa Yayınları.
- [8] Köksal, B. A. (1995). İstatistik Analiz Metodları (4. Baskı). İstanbul: Çağlayan Kitabevi.

İSTATİSTİK VERİLERİ



İÇİNDEKİLER

- Veri Türleri
 - Değişken Sayısına Göre Veriler
 - Ölçüm Türüne Göre Veriler
 - Kayıt Türüne Göre Veriler
- Verilerin Özellikleri
- Veri Kaynakları
 - Canlı Veri Kaynakları
 - Doğal Veri Kaynakları



HEDEFLER

- Bu üniteyi çalıştıktan sonra;
 - İstatistik verilerinin ne olduğunu açıklayabilecek,
 - İstatistik araştırmalara konu olan veri türlerini öğrenebilecek,
 - Verilerin taşınması gereken özellikleri sıralayabilecek,
 - Verilerin hangi kaynaklardan elde edilebileceği hakkında bilgi sahibi olabileceksiniz.

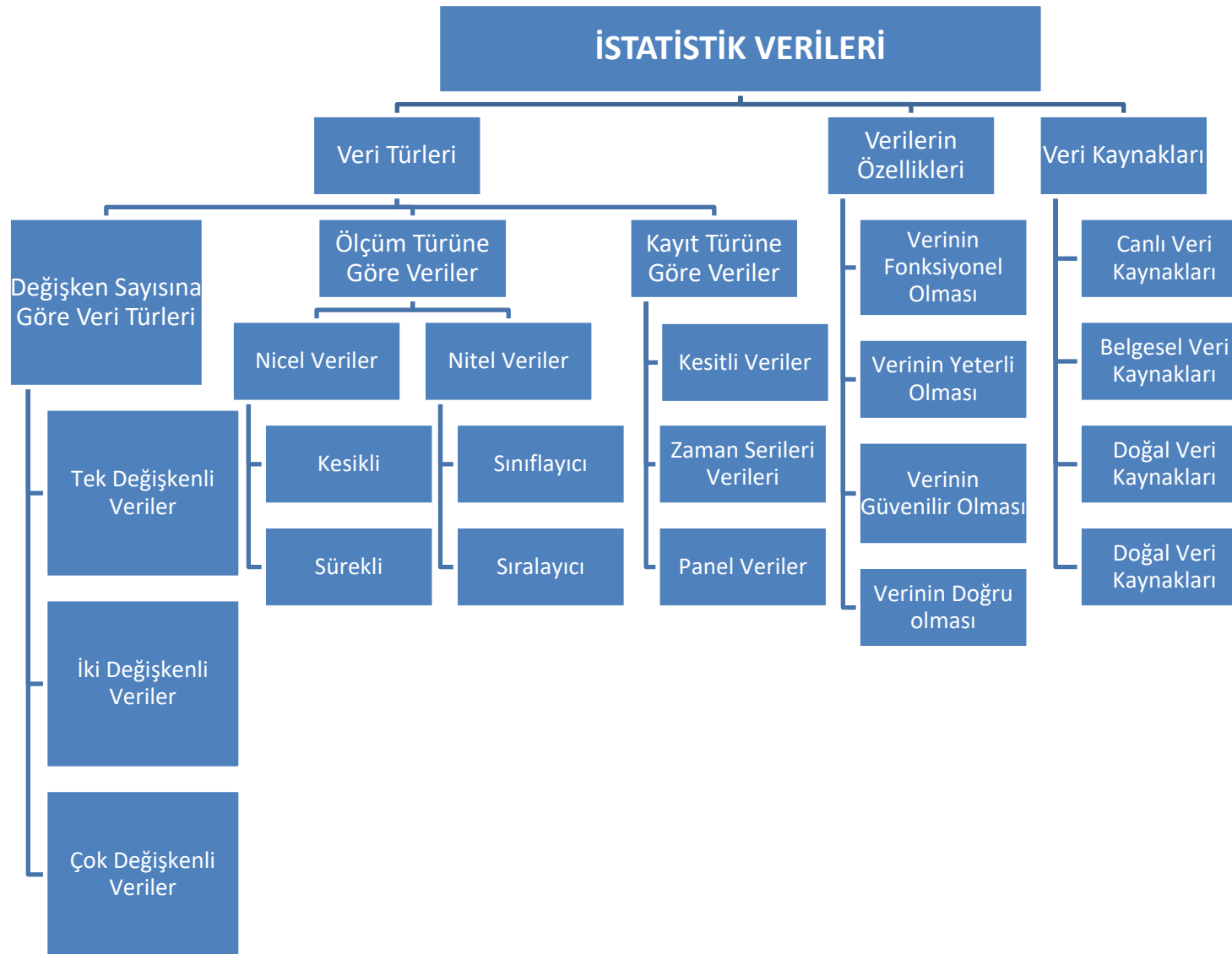


Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ

Dr. Emrah TALAŞ

ÜNİTE 2





Araştırma konusu ile ilgili toplanan işlenmemiş ham bilgiler topluluğuna veri denir.

GİRİŞ

Herhangi bir araştırma konusu ile ilgili toplanan işlenmemiş ham bilgilere *veri* denir. Veri denildiğinde istatistikte genel olarak aklımıza sayılar gelir. Oysa konuya daha geniş bir açıdan baktığımızda, veriler sonuçların çıkarılabileceği olgular ve rakamlardır. Her türlü anlamlı işaret, sayı, sözcük ve nota bizim için veri anlamındadır. Belli bir çalışma için toplanan veriye *veri kümesi* denir [1].

Veriler bir araya getirilerek düzenlenirse enformasyona dönüşürler. Örneğin; notalar veri, şarkılar enformasyondur. Sözcükler veri, anlamlı cümleler enformasyondur. İstatistiksel anlamda, düzenlenerek karar vermede kullanılabilir bir hale dönüştürülen veriler enformasyona dönüşürler [2].

Veri, araştırma konusu ile ilgili istatistiksel çalışmanın temelini oluşturur. Örneğin; iktisatçılar, işsizlik ve enflasyonla ilgili güncel verilerden edindikleri sonuçları kullanarak politik kararlar vermede hükümete yardım eder.

Finansal planlamacılar, borsa ve ekonomideki son trendleri kullanarak yatırım kararlarını verirler.

Muhasebeciler, bir şirketin gerçekleşen satış gelirlerine ait örnek verilerini kullanarak, iddia edilen satış gelirlerinin geçerli olup olmadığını anlayabilirler.

Pazarlama uzmanları, tüketici tercihlerini ortaya çıkaran verileri kullanarak hangi ürünün geliştirilip pazarlanması gerektiği kararını verebilir. Piyasada en fazla tüketilen gıda ile ilgili bir araştırma yapılacaksa gıda maddelerinin tüketimiyle ilgili elde edilen bilgiler veridir ve bu bilgileri kullanarak karar verirler.

Üretim denetçileri, üretim verilerini değerlendirerek ürün kalitesini geliştirme ve kontrolünde kullanır. Politikacılar, kampanya stratejileri hazırlamak ve yasalara son şeklini vermek için kamuoyu görüşünden yararlanır.

Hekimler ve hastaneler, ilaç ve cerrahi işlemler konusundaki verileri kullanarak hastalara mümkün olan en iyi tedavinin geliştirilmesini sağlarlar [1], ya da belli bir bölgede hastalığa yakalananlar ile ilgili araştırmada kullanılabilecek elde edilen her türlü bilgi ile strateji geliştirirler.

Üniversite öğrencilerinin başarılarını etkileyen unsurları araştırmak isteyen bir araştırmacı için öğrenciler hakkında konu ile ilgili elde edilen bilgiler, veridir. Araştırmacılar bu verilere göre sonucu yorumlayabilir.

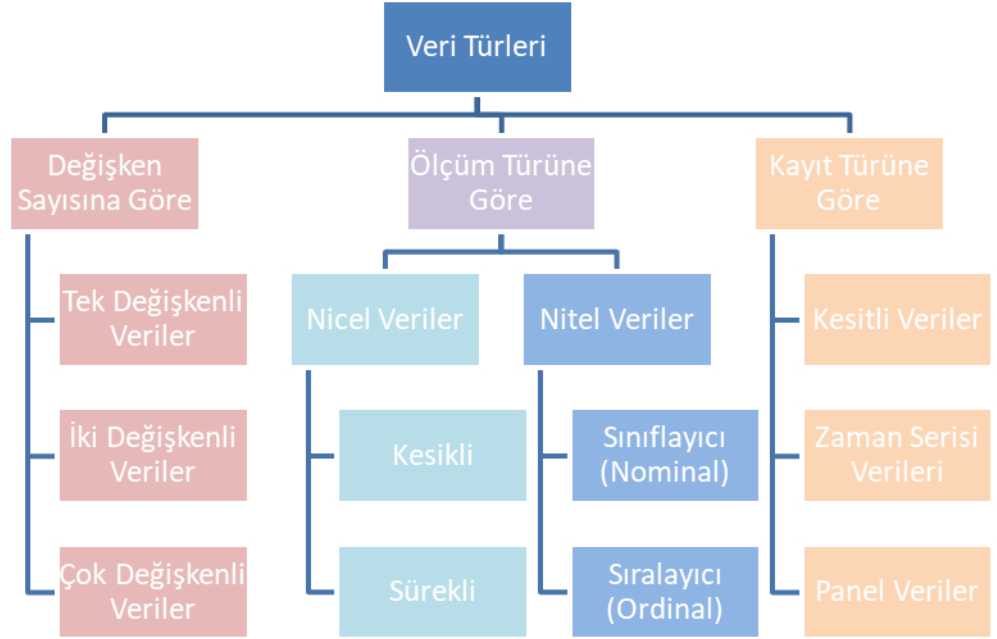
Bu örnekleri daha fazla çoğaltmak mümkündür. Buna göre araştırma konusu ile ilgili toplanan sayısal ya da sayısal olmayan bilgiler, veriyi oluşturmaktadır diyebiliriz.

Veri bir anlamda araştırma konusunun delillerini teşkil eder. İstatistiksel analizler konu ile ilgili toplanan ham bilgilere dayanılarak yapılır. Dolayısıyla istatistiksel analizlerden doğru sonuçların alınması elde edilen verilerin doğruluğuna bağlıdır. Verilerin yanlış ya da hatalı toplanması, sonucun da yanlış veya hatalı çıkmasına neden olacaktır. Veri toplanmadan önce araştırmanın

amacının ne olduğu çok net bir şekilde ortaya konulmalı ve bu amaç çerçevesinde veriler toplanmalıdır.

VERİ TÜRLERİ

Veriler karakterlerine göre farklı şekillerde gruplandırılabilir. Veriler değişken sayısına göre, ölçüm türüne göre ve kayıt türüne göre üç ana başlık altında incelenebilir:



Görsel 2.1 Veri Türleri

Değişken Sayısına Göre Veriler

Gözlemler sonucunda elde edilen veriler birbirinden farklı olabilir. Yani bir konu hakkında değişik veriler elde etmek mümkündür. Dolayısıyla araştırılan konunun değişkenlik özelliği olduğu söylenebilir. Örneğin; bir kamuoyu araştırmasında kişilerin herhangi bir mala olan talepleri incelendiğinde, kişilerin yaşları, cinsiyetleri ya da gelir durumları gibi farklı özelliklerin incelenmesi gerekecektir. Veri ile ilgili bu özellikler değişken olarak adlandırılır. Yani birimlerin incelenen, gözlenen özelliklerine *değişken* denir. Değişken, araştırma sonucunda elde edilen gözlemlerde farklı değerler alabilen nitelikler ya da nesnelerdir. Yaş, gelir gibi sayısal değerler alabildiği gibi cinsiyet, medeni durum gibi sayısal olmayan değerler de alabilir. İstatistiksel araştırma için her bir gözlem değerinin diğerlerinden farklı olması, değişik değerler alabilmesi gerekir. Bu değişim olmadığında bir konunun tanımlanması için tek bir gözlem değeri söz konusu olacaktır ki bu da bir araştırma yapmaya gerek olmadığını gösterir.

Değişken sayısına göre veriler Şekil 2.1’de görüldüğü gibi tek değişkenli veriler, iki değişkenli veriler ve çok değişkenli veriler olmak üzere üçe ayrılmaktadır.



Birimlerin incelenen, gözlenen özelliklerine *değişken* denir.



Tek değişkenli veri, kümelerdeki verilerin değerleri açısından birbirlerinden ne kadar farklı veya benzer olduklarını tespit etmede kullanılır.

Tek değişkenli veriler

Tek değişkenli verilerde araştırmaya konu her birim için tek bir veri elde edilir. Tek değişkenli veri, kümelerdeki verilerin değerleri açısından birbirlerinden ne kadar farklı veya benzer olduklarını tespit etmede kullanılabilir.

Tablo 2.1 Tek Değişkenli Veri Örneği

	Değişken
Birimler	Öğrenci Sayısı
İktisat	120
İşletme	100
Ekonometri	50
Kamu Yönetimi	80
Çalışma Ekonomisi	50
Uluslararası İlişkiler	40
Yönetim Bilgi Sistemleri	30

Tablo 2.1’de İktisat, İşletme, Ekonometri, Kamu Yönetimi, Çalışma Ekonomisi, Uluslararası İlişkiler ve Yönetim Bilgi Sistemleri birimlerine ilişkin öğrenci sayısı değişkeninin verileri görülmektedir. Tek değişkenli veri kümeleri için (özet) veri değerinin hangisi olduğu, veri değerlerinin birbirlerinden ne kadar farklı veya benzer oldukları, özel dikkat gerektiren birimler veya grupların olup olmadığı araştırılabilir.

İki değişkenli veriler

Bu tarz veri kümeleri için birimler ile ilgili iki veri tespit edilir. İki değişkenli veri türlerinde değişkenler arasında bir ilişkinin olup olmadığı, ne yönde olduğu veya değişkenler açısından birimler arasında benzerlikler olup olmadığı araştırılabilir.

Tablo 2.2’de İktisat, İşletme, Ekonometri, Kamu Yönetimi, Çalışma Ekonomisi, Uluslararası İlişkiler ve Yönetim Bilgi Sistemleri birimlerine ilişkin öğrenci sayısı ve not ortalaması değişkenlerinin verileri görülmektedir. İki değişkenli veri kümeleri için iki değişken arasında bir ilişkinin olup olmadığı, varsa bu ilişkinin ne kadar kuvvetli veya zayıf olduğu incelenebilir. Ayrıca bu değişkenlerin biri kullanılarak diğerlerinin kestirilip kestirilmeyeceği ve veri kümesinde özel dikkat gerektiren birimlerin veya grupların olup olmadığı araştırılabilir.

Tablo 2.2 İki Değişkenli Veri Örneği

Değişkenler		
Birimler	Öğrenci Sayısı	Not Ortalaması
İktisat	120	55
İşletme	100	56
Ekonometri	50	57
Kamu Yönetimi	80	65
Çalışma Ekonomisi	50	55
Uluslararası İlişkiler	40	64
Yönetim Bilgi Sistemleri	30	58

Çok değişkenli veriler

Araştırma konusu için üç ya da daha fazla veri elde edilmek istendiğinde çok değişkenli verilerden söz edebiliriz. Çok değişkenli veri türüne aşağıdaki gibi bir örnek verebiliriz:

Tablo 2.3 Çok Değişkenli Veri Örneği

Bölümler	Değişkenler			
	Kız Öğr. Sayısı	Erkek Öğr. Sayısı	Kız Öğr. Not Ortalaması	Erkek Öğr. Not Ortalaması
İktisat	50	70	55	57
İşletme	60	60	56	54
Ekonometri	30	20	57	56
Kamu Yönetimi	10	40	65	62
Çalışma Ekonomisi	25	25	55	54
Uluslararası İlişkiler	20	30	64	66
Yönetim Bilgi Sistemleri	15	35	58	58

Tablo 2.3'te birimlere ilişkin kız öğrenci sayısı, erkek öğrenci sayısı, kız öğrenci not ortalaması, erkek öğrenci not ortalaması değişkenlerinin verileri görülmektedir. Çok değişkenli veri kümeleri için değişkenler arasında ilişkilerin olup olmadığı, varsa bu ilişkilerin ne kadar kuvvetli veya zayıf olduğu incelenebilir. Bu değişkenlerin birini kullanarak diğerinin kestirilip kestirilmeyeceği ve veri kümesinde özel dikkat gerektiren birimlerin veya grupların olup olmadığı araştırılabilir [3].



Nitel veriler (kalitatif) sayısal olmayan veriler, nicel veriler (kantitatif) sayısal verilerdir.

Ölçüm Türüne Göre Veriler

Veriler, ölçüm türlerine göre "nitel veriler (kalitatif) sayısal olmayan veriler", "nicel veriler (kantitatif) sayısal olan veriler" şeklinde genel olarak iki şekilde sınıflandırılır. Araştırmanın konusu gereği bireyleri birbirinden ayırt etmemize yarayan cinsiyet, medeni durum gibi ölçülemeyen ancak sınıflandırılabilen verilere ihtiyaç duyulabileceği gibi sayısal değer alabilen yaş,

boy, gelir gibi değişkenlere ait verilere de ihtiyaç duyulabilir. Değişkenlerin sayısal değer alıp almamasına göre veriler nitel ve nicel olarak sınıflandırılır.

Tablo 2.4 Nicel ve Nitel Veri Örneği

	Değişkenler			
Birimler	Cinsiyet	Yaş	Kilo	Vize
Ali	Erkek	18	62	52
Ayşe	Kız	20	55	70
Ahmet	Erkek	19	65	60
Sevgi	Kız	17	52	55
Mehmet	Erkek	19	58	64
Ayla	Kız	21	50	58

Tablo 2.4’de nicel ve nitel veri örnekleri görülmektedir. Birimlere ilişkin cinsiyet değişkeni nitel veri iken; yaş, kilo ve vize notu değişkenleri ise nicel verilerdir.

Nitel veriler

Nitel veri, değişkenin vasfı ile ilgili sayısal olmayan bilgilerdir. Bir birimin çeşitli kategorilerden hangisinde yer aldığını basit bir şekilde tespit edebildiğimiz veri türüdür. Özellikle sosyal bilimlerde araştırma konuları gereği daha fazla nitel (kalitatif) veri kullanılmaktadır. Nitel veriyle birimlere ait sıfatlar ya da durumlar tespit edilir. Nitel veriler sınıflayıcı (nominal) ve sıralayıcı (ordinal) olmak üzere iki şekilde incelenebilir.

Sınıflayıcı (Nominal) Veriler

Bu tür veriler için elde edilen değerler sayısal bir büyüklük ifade etmezler. Örneğin; araştırma konusuna göre sorulan cinsiyet, medeni durum, saç rengi, göz rengi, mesleği, okuduğu bölüm, doğduğu il, yaşadığı coğrafi bölge gibi değişkenlere verilen cevaplar nominal verilerdir.

Sıralayıcı (Ordinal) Veriler

Sıralayıcı verilerde ilgili değişkenin aldığı değerler açısından birbirlerine üstünlükler ya da önem derecesine göre bir sıralama söz konusudur. Bu tür verilerle üzerinde durulan eşya ve olayın belirli bir sıralamaya tabi tutulması mümkün olur. Bir yarışa katılan koşucular yarışı bitirme zamanlarına göre, birinciden sonuncuya kadar sıralanırken; belirli bir mamül hakkında tüketicilerin beğenileri, “Hiç beğenmedim, az beğendim, kararsızım, beğendim, çok beğendim.” şeklinde sıralanabilir [4]. Belirli bir konu hakkında duyguları hisleri ve fikirleri derecelendirmeye imkân sağlar. Ordinal veriler yalnızca “>” ve “<” işlemleri için sayı gibi değerlendirilebilir, bunlar dışında herhangi bir matematiksel işlem yapılmaz [5].



Nitel veriler, sınıflayıcı (nominal) ve sıralayıcı (ordinal) olmak üzere iki şekilde incelenebilir.



Örnek

•Kanser hastalarının hastalık aşamasının girildiği “AŞAMALAR”adında bir değişken ile ilgili bilgi toplayalım. Bu değişken için aşama 1’den aşama 4’e karşılık gelmek üzere 1, 2, 3 ve 4 değerlerini tespit edebiliriz. Bu değerler belirli bir sıra ifade etmektedir. Örneğin; “Aşama 3, Aşama 2’den daha ileri bir aşamadır.”, “Aşama 1 en iyi, aşama 4 en kötü aşamadır.” Ancak verilerin düzeyleri arasındaki mesafeler belirli değildir. Matematik işlem yapıldığında $2-1=4-3=1$ doğrudur; ama “aşama 2, aşama 1’den ne kadar ileriye, aşama 4 de aşama 3’ten o kadar ileridir.” denilmez.

Nicel Veriler

Nicel veriler, birimlerin sayısal özelliklerini gösteren değerlerdir. Bir değişkenin değerlerinin olası ölçümleri eğer rakamlarla (“ne kadar” veya “kaç” gibi) ifade ediliyor ise o halde verinin nicel olduğu söylenebilir. Örneğin; bir evin satış fiyatı, bir faturanın ödeme süresi ya da bir otomobilin katettiği mesafe nicel verilerdir. Araştırma konusu ile ilgili elde edilen her türlü sayısal değer nicel veri olarak ifade edilir. Nicel veriler kesikli ve sürekli olmak üzere iki şekilde incelenebilir:

Kesikli Veriler

Kesikli veriler, birimlere ait özelliklerin tam sayılarla ifade edildiği veri setleridir. Araştırma konusuna göre sorulacak sorulardan elde edilen değerler tam sayıdır (0, 1, 2, 3, ...vb). Mesela evli bir kadına “Kaç çocuğunuz var?” sorusunu sorduğunuzda alacağınız cevap kesikli veridir. Yine bir öğrenciye “Haftada kaç kez sinemaya gidersin?” sorusunu sorduğunuzda alacağınız cevap kesikli veridir.

Sürekli Veriler

Sürekli veriler, tam sayılar arasında sonsuz değer alabilen ölçü birimi ve kesirli değerler içeren veri setleridir. 86.25, 76.56, 78.89 gibi değerler sürekli verilere örnek verilebilir. Örneğin, bir sınıftaki öğrencilere ait boy ve ağırlık kayıtları sürekli verilerdir. Ölçüm ve tartım aletinin hassasiyetine göre virgülden sonra basamaklar çoğaltılabilir. İki tür sürekli veri vardır.

Aralık ölçeği ile ölçülmüş veriler: Bu ölçekte üzerinde durulan değişken belirli iki değer arasında sonsuz değer alabilir. Bu ölçekteki 0 değeri, ölçülen karakteristiğin olmadığını göstermez. Yine ölçüm karakteristiklerinden biri değerinin katlarıyla ifade edilemez.

Bu ölçeğe verilebilecek en açık örnek, sıcaklık ölçümleridir. 0 °C, sıcaklığın olmadığını ifade etmez. Sadece havanın çok soğuk olduğunu bildirir. Haliyle, niteliği gereği, tanımlanmış bir sıfır değeri yoktur. Bunun yanı sıra sıcaklıkların oranları anlamlı değildir. Örneğin, 8 °C, 4 °C ’nin iki katı daha sıcak olduğunu söylemek mantıklı değildir. Uygulamada sıcaklık dışında çok az sayıda aralık verileri bulunmaktadır. Hemen hemen tüm nicel veriler oran verileridir.



Sürekli veriler, tam sayılar arasında sonsuz değer alabilen ölçü birimi ve kesirli değerler içeren veri setleridir.

Oran ölçeği ile ölçülmüş veriler: Zayıftan kuvvetliye doğru sıraladığımız yukarıdaki ölçeklerin en hassas olanıdır. Ölçülen karakteristiğin sıfır olması o karakteristiğin olmadığını gösterir. Aynı şekilde ölçülen bir karakteristik diğerinin katları ile ifade edilebilir. Ağırlık ve boy ölçümleri bu ölçeğe verilebilecek en iyi örneklerdir. Bir cismin ağırlığının 0 olması o cismin olmadığını gösterirken, 2 kg gelen bir cisim 4 kg gelen aynı özellikteki cismin yarısı kadar ağırdır. Tablo 2.5'de verilerin gruplandırıldığı özet tablo sunulmuştur [9].

Tablo 2.5 Verilerin Gruplandırılması

Veri	Alt Seçenekler	Değer Belirleme	Örnek
1. Nitel (Kalitatif)	Nominal Veri	Özelliklerin isimsel kategorilere göre belirtilmesi	Cins: Erkek, kadın Saç Rengi: Siyah, kumral, sarı vb.
	Ordinal Veri	Özelliklerin sıraya konulmuş biçimde belirtilmesi	Az<Normal<Çok<Küçük<Orta<Uzun
2. Nicel (Kantitatif)	Aralıklı	Değerlerin belirli aralıklara göre belirtilmesi	Yaş Grupları: 0-4, 5-14, 15-24, 25-44, 45-64, 65+
	Oranlı	Fiziksel ölçümlemeye göre değerlerin saptanması	m, cm, mm, kg, gr vb.

Kayıt Türüne Göre Veriler

Kayıt türüne göre veriler; kesit veriler, zaman serisi verileri ve panel veriler olmak üzere üç şekilde incelenebilir.

Kesit Veriler

Kesit veriler, belirli bir anda veya belirli bir zaman aralığında toplanmış verilerdir. Bu tür verilerde veri değerlerinin önemi vardır. Farklı birimlere ait değerlerin aynı zamanda toplanması ile elde edilen veriler kesit verilerdir. Örneğin, farklı gelir gruplarının belli bir zaman döneminde yaptıkları tüketim harcamaları ile ilgili değerler kesit verilerdir veya bir üretim sürecinde belli bir zamanda üretilen ürünler ile ilgili kusurlu-kusursuz şeklinde bir tasnifle elde edilen değerler kesit verilerdir ya da belli bir günde aşağıdaki tabloda gösterildiği gibi döviz kurları ile ilgili elde edilen değerler kesit verilerdir.



Kesit veriler, belirli bir anda veya belirli bir zaman aralığında toplanmış verilerdir.

Tablo 2.6 Kesit Veri Örneği

Döviz Cinsi	TL. Karşılığı
1 Euro	2.320
1 ABD Doları	1.870
1 Japon Yeni	2.278
1 İngiliz Sterlini	2.832
1 Kanada Doları	1.770
1 İsveç Kronu	0.258
1 İsviçre Frangı	1.893

Zaman Serisi Verileri

Araştırma konusu ile ilgili değişkenin zaman içerisindeki değişimini gösteren bilgi, *zaman serisi verisi* olarak ifade edilir. Zaman serileri araştırma konusuna göre günlük, haftalık, aylık veya yıllık şeklinde zaman periyodunda grafikte sunulabilir. Bir zaman serisi grafiği bizlere, bir değişkenin en yüksek ve en düşük seviyelerini gösterir. Bir değişken yükseliyor mu alçalıyor mu bunu takip imkânı verir. Grafiğin eğimi pozitif ise değişkenin yükselmekte, negatif ise düşmekte olduğunu gösterir. Değişkenin ne hızla değiştiğini gösterir. Grafikten değişkendeki yükseliş ya da düşüşün hızlı mı, yavaş yavaş mı olduğunu anlayabiliriz. Eğer grafikteki doğru çok dikse bu değişkendeki düşüş veya yükselişin çok hızlı olduğunu gösterir. Eğer grafikteki doğru yatık ise değişkendeki düşüş veya artış yavaştır. Zaman serileri grafikleri bize değişik zaman aralıklarını karşılaştırma imkânı verir [7]. Örneğin, 2005 yılından 2010 yılına kadar Türkiye’de işsizlik oranlarıyla ilgili elde edilen değerler zaman serisini gösterir.



Araştırma konusuyla ilgili değişkenin zaman içerisindeki değişimini gösteren bilgi, zaman serisi verisidir.

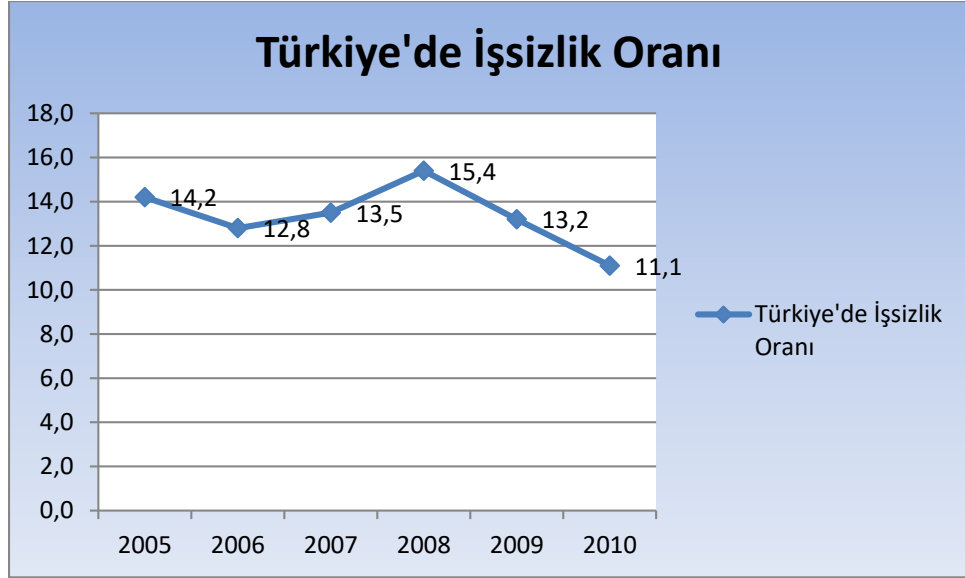


Örnek

- 2005 yılından 2010 yılına kadar Türkiye’de işsizlik oranlarıyla ilgili elde edilen değerler bir zaman serisi verisini göstermektedir. Tablo 2.7 deki veriler kullanılarak, Şekil 2.2’deki zaman serisi grafiği elde edilmiştir.

Tablo 2.7 Yıllara Göre İşsizlik Oranı Verileri

Yıllar	Türkiye’de İşsizlik Oranı
2005	14.2
2006	12.8
2007	13.5
2008	15.4
2009	13.2
2010	11.1



Görsel 2.2 Türkiye'de Yıllara göre İşsizlik Oranı

Panel Veriler

Aile, firma veya birey gibi ele alınan mikro birimlere ait kesit verilerin zaman serisi hâli *panel veridir*. Tüketici, aile veya firma gibi aynı mikro birimlerin zaman içinde gözlenmesi ile elde edilen veriler panel veriler olarak adlandırılır [3]. Hatırlanacağı gibi kesit veriler, bir veya daha fazla birimlerle ilgili belirli bir anda veya belirli bir zaman aralığında toplanmış veriler olarak ifade edilmişti. Zaman serileri ise ilgili değişkenin zaman içerisindeki değişimini gösteren bilgi olarak ifade edilmişti. Panel veriler birimler arasındaki farklılıkların ya da benzerliklerin belirli zaman periyodu içerisinde gözlemlenmesi ile elde edilir [6].



Örnek

- Tablo 2.8'de aylar itibarıyla bir bölgede yaşayan kişilerin gelir ve harcamalarının yaş ve cinsiyete göre dağılımı verilmiştir. Bu dağılımda zaman ve farklı birimlere ait değişken değerleri hem kesit verileri hem de zaman serisini içerisinde barındıran panel verilerdir.

Tablo 2.8 Panel Veri Örneği

Birimler	Aylar	Cinsiyet	Yaş	Gelir	Harcama
1	Ocak	E	38	1500	1190
2	Şubat	E	45	1800	1790
3	Mart	K	35	1550	1000
4	Nisan	E	42	1650	950
5	Mayıs	K	26	1750	1250
6	Haziran	E	28	1550	1000

Tablo 2.8’de görüldüğü gibi birden fazla kişi, birden çok zaman diliminde çeşitli özellikler itibarıyla karşılaştırılmıştır.

VERİNİN TAŞIMASI GEREKEN ÖZELLİKLER

Bir araştırmanın başarısını, konu ile ilgili toplanan verinin taşıması gereken tüm özellikler belirler. Bu özellikler dört ana grupta toplanabilir.

Verinin Fonksiyonel Olması

Fonksiyonel veri toplayabilmek, veri toplama ölçeklerini doğru hazırlamakla mümkündür. Araştırma konusu içinde problem doğru belirlenip sınırı çok net tespit edilmelidir. Veri hangi yöntemle toplanacaksa o yöntemin veri toplama aracı belirlenen sınırın dışına taşmayacak, problemin çözümü için gerekli tüm bilgileri içerecek şekilde hazırlanmalıdır.

Verinin Yeterli Olması

Veri toplama aracının hazırlanması aşamasında araştırma problemi, problemi oluşturan alt problemlere ayrılmalıdır. Hazırlanan her alt problemin altına o alt probleme ilişkin toplanması gereken verileri sağlayacak sorular hazırlanmalı ve hazırlanan her soru alt problem ile ilişkilendirilerek soruların gerekliliği ya da gereksizliği saptanmalıdır. Ayrıca hazırlanan soruların getireceği varsayılan verinin alt probleminin tanımlanması için yeterliliği kontrol edilmelidir.



Verinin taşıması gereken tüm özellikler dört ana grupta toplanabilir.

Verinin Güvenilir Olması

Bir konuda elde edilen verinin aynı koşullar oluşturularak tekrarlandığında aynı verinin elde edilmesi, aynı bireyden aynı yanıtın alınması verinin güvenilir olduğu anlamına gelmektedir. Bilgi doğru ya da yanlış olabilir. Verinin güvenilirliği veri toplanan yer ya da kişi ile de ilgilidir.

Verinin Doğru Olması

Gerçek durumu olduğu gibi yansıtan veri doğru olarak kabul edilmektedir. Tarafli olmadan doğru örneklemden doğru bilgiler elde edilmelidir [9].

VERİ KAYNAKLARI

Canlı Veri Kaynakları

Bitki ve hayvanlar

Bu canlılara ilişkin veriler genelde gözlem ya da deneysel yöntemle toplanmaktadır. Verilerin toplanması uzun bir zaman dilimi gerektirmektedir. Bu canlılara ilişkin araştırmalar daha çok o canlıların yaşam biçimleri ile canlılar ve doğa dengesi ilişkisini belirlemeye yönelik olmalıdır.

İnsanlar

En çok kullanılan veri kaynağı insanlardır. İnsanlar günlük yaşantılarında çeşitli konulara ilişkin önemli görüşler geliştirmekte, sorunları görmekte ve hatta bireysel çözümler üretebilmektedirler.

Bireysel yaşantılar; bireyin inançlarının, gelenek ve göreneklerinin, alışkanlıklarının, toplumun değer kalıplarının etkisi altında kazanılmaktadır. Bu etkilerin insanlardan elde edilen verilerde de görülmesi söz konusu olmaktadır. Aile, sosyal ve ekonomik açıdan toplumun en önemli birimidir ve önemli bir veri kaynağıdır. Ürettiği aile kültürü, doğruları ve yanlışları, etik anlayışı ile toplumun kültür yapısını etkilemektedir [8].

Belgesel Veri Kaynakları

Yayınlanmış belgesel veri kaynakları

Bunlar kitap, ansiklopedi, gazete, dergi, araştırma, istatistikler vb. veri kaynaklarıdır. Her araştırma çalışmasında konuya ilişkin yeterli miktarda belge taranması gerekmektedir. Özellikle önceki zamanlara ilişkin olay ve olguların araştırılmasında ya da problemin geçmişle olan ilişkisi yönünden incelenmesinde yayınlanmış belgesel veri kaynakları çok kullanılmaktadır.

Yayınlanmamış belgesel veri kaynakları

Yayınlanmamış belge, bulgu, arşiv evrakları ve diğer dokümanlar birer veri kaynağıdır. Bu veriler ilgili olay ve olgularla ilişkilendirilerek araştırmayı belirli sonuçlara götürebilecek nitelikte olabilmektedir.

Doğal Veri Kaynakları

Yaşayan doğal veri kaynakları

Doğada bulunan çeşitli varlıklar ve olaylar çeşitli alet ve yöntemlerle incelendiğinde araştırma için gereken veriler elde edilebilir. Toprak bilimi, deniz bilimi, gök bilimi, çevre bilimi vb. doğaya ilişkin bilim dallarında çalışan araştırmacılar, ilgili alanda doğadan bol miktarda veri toplayabilir.

Kalıntı doğal veri kaynakları

Bunlar toprak altında ve üstünde olan, daha önceki zamanlara ilişkin doğa kaynaklarından elde edilen verilerdir. Yapı kalıntıları, fosiller, mezarlar, tapınaklar, yaşanan döneme ilişkin araçlar vb. veri kaynaklarını oluşturmaktadır.

Daha çok arkeoloji, sanat tarihi, uygarlık tarihi gibi daha önceki zamanlara ilişkin kültür yapısını ortaya koymaya çalışan bilim dallarındaki araştırmacılar tarafından başvurulanan veri kaynaklarıdır [10].



Doğada bulunan çeşitli varlıklar ve olaylardan çeşitli alet ve yöntemlerle araştırma için gereken veriler elde edilir.



Bireysel Etkinlik

- Veri ve veri türlerini açıklayınız.
- Veri türlerine örnek veriniz.
- Kalitatif ve kantitatif veriler arasındaki farkı açıklayınız.
- Ölçüm türlerine göre verilere örnek veriniz. Veriler arasındaki farka dikkat ediniz.



Özet

- Veri denildiğinde istatistikte genel olarak aklımıza sayılar gelir. Oysa konuya daha geniş bir açıdan baktığımızda veriler, sonuçların çıkarılabileceği olgular ve rakamlardır. Herhangi bir araştırma konusu ile ilgili toplanan işlenmemiş ham bilgilere veri denir. Belli bir çalışma için toplanan veri'ye ise veri kümesi denir. Gözlemler sonucunda elde edilen veriler birbirinden farklı olabilir. Yani bir konu hakkında değişik veriler elde etmek mümkündür. Dolayısıyla araştırılan konunun değişkenlik özelliği olduğu söylenebilir. Veri ile ilgili bu özellikler değişken olarak adlandırılır. Yani birimlerin incelenen, gözlenen özelliklerine değişken denir. Değişken, araştırma sonucunda elde edilen gözlemlerde farklı değerler alabilen nitelikler ya da nesnelerdir
- Değişken sayısına göre veriler: Tek değişkenli veriler, iki değişkenli veriler ve çok değişkenli veriler olmak üzere üçe ayrılmaktadır.
- Tek Değişkenli Veriler: Tek değişkenli verilerde araştırmaya konu her birim için tek bir veri elde edilir.
- İki Değişkenli Veriler: Bu tarz veri kümeleri için birimler ile ilgili iki veri tespit edilir.
- Çok Değişkenli Veriler: Araştırma konusu için üç ya da daha fazla veri elde edilmek istendiğinde çok değişkenli verilerden söz edebiliriz.
- Ölçüm Türüne Göre Veriler: Veriler, ölçüm türlerine göre nitel veriler (kalitatif) sayısal olmayan veriler nicel veriler (kantitatif) sayısal olan veriler şeklinde genel olarak iki şekilde sınıflandırılır.
- Nitel veri: Değişkenin vasfı ile ilgili sayısal olmayan bilgilerdir. Nitel veriler sınıflayıcı (nominal) ve sıralayıcı (ordinal) olmak üzere iki şekilde incelenebilir:
- Sınıflayıcı (Nominal) Veriler: Bu tür veriler için elde edilen değerler sayısal bir büyüklük ifade etmezler.
- Sıralayıcı (Ordinal) Veriler: Sıralayıcı verilerde ilgili değişkenin aldığı değerler açısından birbirlerine üstünlükler ya da önem derecesine göre bir sıralama söz konusudur.
- Nicel veri: Birimlerin sayısal özelliklerini gösteren değerlerdir. Nicel veriler kesikli ve sürekli olmak üzere iki şekilde incelenebilir: Kesikli veriler; birimlere ait özelliklerin tam sayılarla ifade edildiği veri setidir. Sürekli veriler; tam sayılar arasında sonsuz değer alabilen ölçü birimi ve kesirli değerler içeren veri setleridir. İki tür sürekli veri vardır: Aralık ölçeği ile ölçülmüş veriler, oran ölçeği ile ölçülmüş veriler,
- Kayıt Türüne Göre Veriler:
- Kesit Veriler: Belirli bir anda veya belirli bir zaman aralığında toplanmış verilerdir.
- Zaman Serisi Verisi: Araştırma konusu ile ilgili değişkenin zaman içerisindeki değişimini gösteren bilgi zaman serisi verisi olarak ifade edilir.
- Panel Veri: Aile, firma veya birey gibi ele alınan mikro birimlere ait kesit verilerin zaman serisi hâli panel veridir.
- Verilerin taşıması gereken özellikler: Veriler fonksiyonel olmalı, veriler yeterli olmalı, veriler güvenilir olmalı ve veriler doğru olmalıdır.
- Veri Kaynakları: Canlı veri kaynakları (insanlar, bitki ve hayvanlar), belgesel veri kaynakları (yayınlanmış belgesel veri kaynakları, yayınlanmamış belgesel veri kaynakları), doğal veri kaynakları (toprak altında ve üstünde olan, daha önceki zamanlara ilişkin doğa kaynaklarından elde edilen veriler).

DEĞERLENDİRME SORULARI

1. Ölçüm türlerine göre veriler arasında aşağıdakilerden hangisi bulunmaz?
 - a) Nicel veriler
 - b) Kesikli veriler
 - c) Sürekli veriler
 - d) Sınıflayıcı (Nominal) veriler
 - e) Panel verileri
2. Aşağıdaki ifadelerden elde edilen bilgilerin hangisi nitel veriye örnektir?
 - a) Öğrencilerin cinsiyetleri
 - b) Öğrencilerin sayısı
 - c) Öğrencilerin boyları
 - d) Öğrencilerin ağırlıkları
 - e) Öğrencilerin başarı puanları
3. Aşağıdaki ifadelerden elde edilen bilgilerin hangisi nicel veriye örnektir?
 - a) Cinsiyet
 - b) Medeni durum
 - c) Saç rengi
 - d) Boy
 - e) Uyruk
4. Sıralayıcı (Ordinal) verilere aşağıdaki verilerden hangisi örnektir?
 - a) İstatistik dersini alan sarı saçlı öğrenciler
 - b) İstatistik notu en yüksek olan ilk 5 öğrenci
 - c) İstatistik dersinden 80 puan alan öğrenciler
 - d) İstatistik dersini alan erkek öğrenciler
 - e) İstatistik dersini alan yabancı öğrenciler
5. Hasta memnuniyeti ölçümünde kullanılan sıralayıcı veriler arasında aşağıdakilerden hangisi bulunmaz?
 - a) Hiç beğenmedim
 - b) Az beğendim
 - c) Hastaneye gitmedim
 - d) Beğendim
 - e) Çok beğendim
6. Türkiye istatistik kurumunun sağladığı veri hangi veri kaynağına örnektir?
 - a) Yayınlanmamış veri kaynağı
 - b) Yayınlanmış veri kaynağı
 - c) Yaşayan doğal veri kaynağı
 - d) Kalıntı doğal veri kaynağı
 - e) Deneysel veri kaynağı

7. Aşağıdaki sorulara verilen cevapların hangisi sürekli veriye örnektir?
- a) Bir ayda kaç defa tiyatroya gidersiniz?
 - b) Kilonuz nedir?
 - c) Doğum yılınız nedir?
 - d) En sevdiğiniz rakam nedir?
 - e) Kaç gömleğiniz var?
8. Aşağıdakilerden hangisi sürekli veriye örnek olarak verilebilir?
- a) Medeni hal
 - b) Doğum yılı
 - c) Bir ülkede yaşayanların sayısı
 - d) Ülkelerin yüzölçümü
 - e) Apartmanların daire sayısı
9. Yıllar itibarıyla Erzurum’da satılan doğalgaz miktarı bir seri halinde sunulmuştur. Bu seri aşağıdakilerden hangisine bir örnek olur?
- a) Mekân serileri
 - b) Yatay kesit verileri
 - c) Zaman serisi
 - d) Panel veri seti
 - e) Kesit verileri
10. Verilerin taşınması gereken özellikler arasında aşağıdakilerden hangisi bulunmaz?
- a) Veri fonksiyonel olmalıdır.
 - b) Veri yeterli olmalıdır.
 - c) Veri güvenilir olmalıdır.
 - d) Veri doğru olmalıdır.
 - e) Veri kesin olmalıdır.

Cevap Anahtarı

1.e, 2.a, 3.d, 4.b, 5.c, 6.b, 7.b, 8.d, 9.c 10.e

YARARLANILAN KAYNAKLAR

- [1] Bowerman, Bruce L., O'Connell, Richard T., Murphree, Emily S., Orris J.B. (2013). Essentials of Business Statistics (Fourth Edition). Irwin: McGraw-Hill.
[Orhunbilge, N., Can, M., Er, Ş. (2013). İşletme İstatistiğinin Temelleri (4.basımdan çeviri). Ankara: Nobel Yayın Dağıtım.]
- [2] Çakır, F. (2000). Sosyal Bilimlerde İstatistik (1. baskı). İstanbul: Alfa Yayıncılık.
- [3] Gürsakal, N., Oğuzlar, A, Gürsakal, S. (2022). Betimsel İstatistik (10. Baskı). Bursa: Dora Yayıncılık.
- [4] Başar, Alaaddin, Oktay, Erkan. (2018). Uygulamalı İstatistik-I: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap ve Kirtasiye.
- [5] Büyüköztürk, Ş. (2018). Sosyal Bilimler için Veri Analizi El Kitabı (31. baskı). Ankara: Pegem Akademi Yayıncılık.
- [6] Hayran, M., Özdemir, O. (1996). Bilgisayar İstatistik ve Tıp (2. baskı). Ankara: Hekimler Yayın Birliği.
- [7] Çepni, E. (2019). Ekonomik Göstergeler ve İstatistikler Rehberi (8. baskı). Ankara: Seçkin Yayıncılık.
- [8] Johnson, R., A., Wichern, D., W. (1998). Applied Multivariate Statistical Analysis (4.Baskı). NJ: Prentice-Hall.
- [9] Özdamar, K. (2012). Paket Programlar ile İstatistiksel Veri Analizi (11. baskı). Ankara: Nisan Kitabevi.
- [10] Yazıcıoğlu, Y., Erdoğan, S. (2014). SPSS Uygulamalı Bilimsel Araştırma Yöntemleri (4. baskı). Ankara: Detay Yayıncılık.

İSTATİSTİK SERİLERİ



İÇİNDEKİLER

- Zaman Serileri
- Mekân Serileri
- Bileşik Seriler
- Frekans (Bölünme) Serileri
 - Nitel Seriler
 - Nicel Seriler
 - Mutlak Frekans Serileri
 - Basit Seriler
 - Sınıflandırılmış Seriler
 - Gruplandırılmış Seriler
 - Nispi Frekans Serileri
 - Kümülatif Frekans Serileri



HEDEFLER

- Bu üniteyi çalıştıktan sonra;
 - İstatistiksel serileri gözlemlerden faydalanarak oluşturabilecek,
 - İstatistik serileri ayırt edebilecek,
 - İstatistiki verileri, serilere dönüştürme hakkında detaylı bilgi sahibi olabileceksiniz.



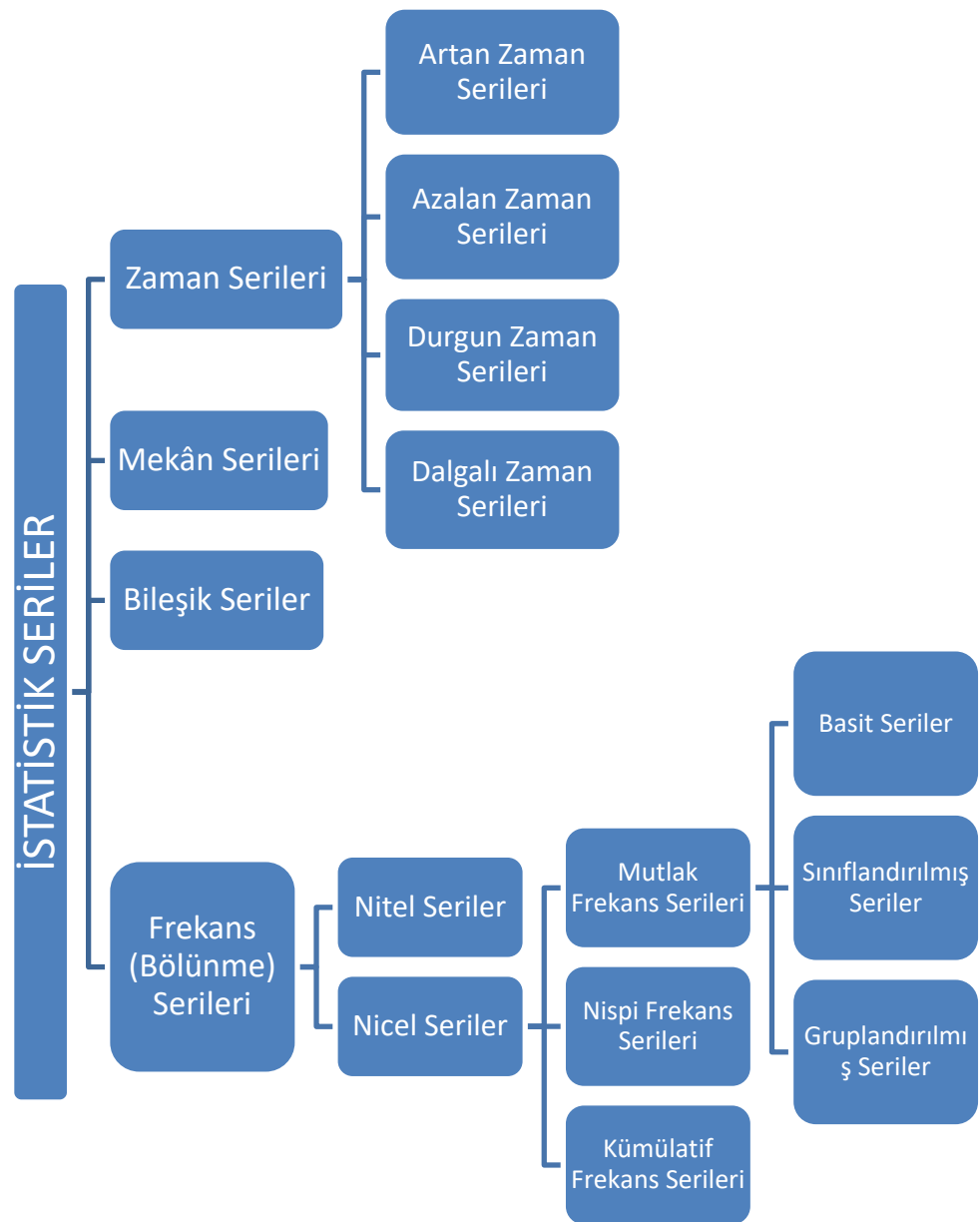
Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ

Dr. Emrah TALAŞ

ÜNİTE

3



GİRİŞ

Bir araştırma sonucunda elde edilen verilerin anlamlı olarak özetlenmesi istatistiğin en önemli fonksiyonudur. İstatistik çok sayıda sayının ortaya koyduğu karmaşıklık içinde sayıların göremediğimiz yüzlerini görmeye, kavrayamadığımız özelliklerini kavramamıza yarar.

İstatistik; sayıları tablolayıp sıralayıp, sınıflayıp, özetleyip grafik çizerek onları başkalarına iletmemizi kolaylaştırır. Başka bir ifadeyle istatistik, verilerin elde edilmesinden başlayarak verilerin bilgiye dönüştürülmeleri ve hacimlerinin azaltılarak (sayfalarca elde edilen veriyi bir sayfalık tablo haline getirerek) kullanım değerlerinin artması sürecinde çok çeşitli alanlarda çalışan insanlara yarar sağlar.

İstatistikte buraya kadar ifade edilen bu işlemleri gerçekleştirebilmek için ham verileri ilk aşamada seri haline getirmek gerekir. Elde edilen ham verilerin anlaşılabilirliğini ve hesaplanabilirliğini kolaylaştırmak için belli bir kurala göre sınıflandırılarak tablolaştırılması sonucu oluşan diziye seri denir. Başka bir ifadeyle seriler, gözlem sonuçlarının belirli bir kritere göre sıralanmasıyla elde edilen rakam dizisidir [2].

Verilerin seri haline dönüştürülmesi, bir seri üzerinde durulan değişkene ait mümkün sonuçları ve bu sonuçların tekrarlanma durumlarını gösterdiği için üzerinde durulan değişkenin anlaşılması bakımından önemli ipuçları içerir. Bu ipuçları, uygun istatistiksel tekniklerle değerlendirilerek üzerinde durulan olayın anlaşılması sağlanabilir.

İstatistik serileri çeşitli kriterlere göre sınıflandırılabilir. Bu ünite de istatistik serileri zaman, mekân ve sıklık kriterlerine göre sınıflanacaktır. Bu şekilde elde edilecek seriler herhangi bir değişkenin zaman içerisinde nasıl değiştiğini gösteren zaman serileri, gözlem sonuçlarını yer vasfının sıklığı itibarıyla gösteren mekân serileri, iki ya da daha fazla değişkenin birlikte değişimini gösteren bileşik serileri ve gözlemlerin maddi bir özelliğe göre sıralanmasıyla elde edilen frekans (bölünme) serileri olmak üzere dört ana başlık altında incelenmiştir [3].

ZAMAN SERİLERİ

İstatistiksel araştırmalarda veriler, araştırma konusunun özelliklerine ve gerçeklerine ait bilgilerden ibarettir. Bu bilgiler çeşitli kaynaklardan elde edilebilirler. İstatistiki araştırmalarda veri toplama farklı biçimde gerçekleştirilir. Bu yöntemlerden biriside önceden toplanmış olan bilgilerden yararlanmaktır. Bunlar genellikle istatistiki bültenler vb. çalışmalar veya istatistik yıllıkları biçimindedir. Bu tür veriler genellikle zaman serisi niteliğinde bilgi veren farklı dönemler itibarıyla toplanan ekonomik verilerdir.

Bu açıklamalar çerçevesinde zaman serilerini şu şekilde tanımlayabiliriz: Gözlemlerin zaman birimlerine göre sınıflandırıldığı serilere zaman serileri denir. Zaman serisi verisi farklı zaman periyodları için toplanan veridir. Zaman serisi verileri, araştırılacak konuya ait gözlem sonuçlarını günlük, haftalık, aylık, üç aylık, altı aylık ve yıllık gibi zaman aralıklarıyla derlenir. Ekonomik ve finansal veriler bu



Ham verilerin anlaşılabilirliğini ve hesaplanabilirliğini kolaylaştırmak için belli bir kurala göre sınıflandırılarak tablolaştırılması sonucu oluşan diziye seri denir.



Gözlemlerin zaman birimlerine göre sınıflandırıldığı serilere zaman serileri denir.

seriler için en sık kullanılan verilerdir. Bir firmanın aylık satış miktarı, bir ülkenin yıllara göre işsizlik oranı gibi vb. veriler zaman serilerine örnek olarak verilebilir.

Zaman serisi verileri, değişkenlerin bir dönemden diğerine ardışık şekilde gözlemlendiği sayısal değerler hakkında bilgiler verir. Düzenli zaman aralıklarında dizinin gelişimini takip etmesi, istatistiki analiz açısından önemlidir. Zaman serisi verisi olarak elde edilen veriler zaman serisi grafiğinde gösterilebilir.

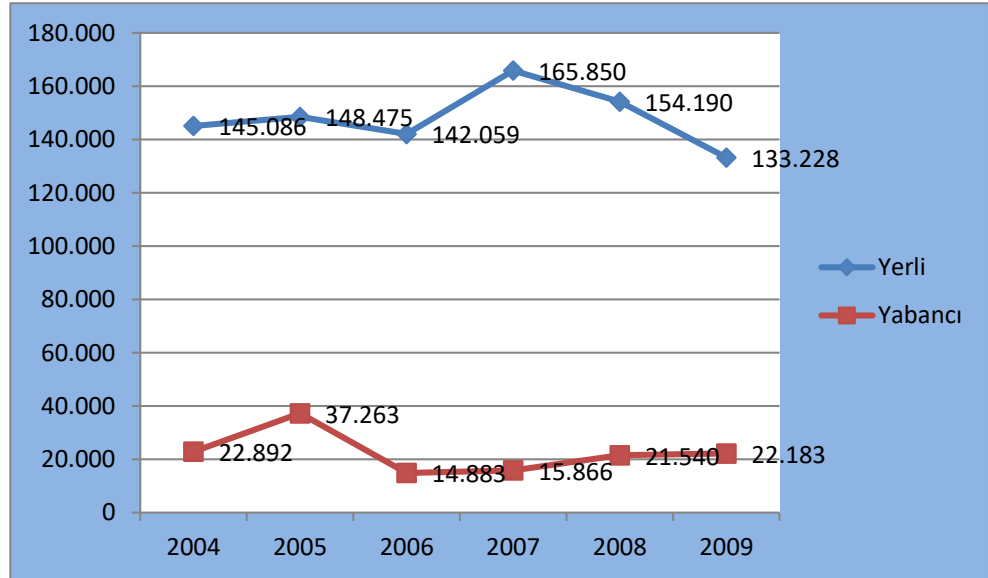


Örnek

- 2004 yılından 2009 yılına kadar Erzurum'a gelen turist sayıları yıllar itibarıyla Tablo.3.1'de verilmiştir. Bu verilere ait zaman serisi grafiği Görsel 3.1'de çizilerek turist sayısındaki değişim dönemler itibarıyla gözlemlenebilir.

Tablo 3.1. 2004-2009 Yılları Arası Erzurum'a Gelen Turist Sayıları

Yıllar	Yerli	Yabancı	Toplam
2004	145.086	22.892	167.978
2005	148.475	37.263	185.738
2006	142.059	14.883	156.941
2007	165.850	15.866	181.685
2008	154.190	21.540	175.730
2009	133.228	22.183	155.411



Görsel 3.1. 2004-2009 Yılları Arası Erzurum'a Gelen Turist Sayısı Zaman Serisi Grafiği

Bir zaman serisi verisinin dönemler itibarıyla ardışık şekilde artma, azalma, sabit kalma ve sürekli değişme göstermesi durumuna göre artan zaman serisi, azalan zaman serisi, durgun zaman serisi ve dalgalı zaman serisinden bahsedebiliriz.

Zaman serisi grafikleri çok fazla bilgiyi bir bakışta izleyebilmenin en kolay yolunu sunarlar [6].

Bir zaman serisi; günler, haftalar, aylar ve yıllar itibarıyla sürekli artıyorsa buna *artan zaman serisi* diyebiliriz.



Örnek

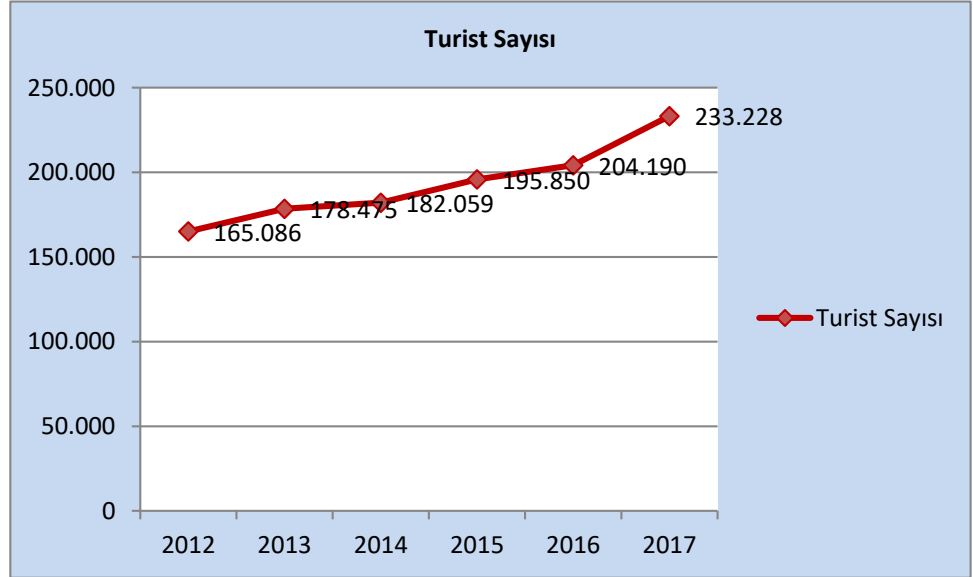
- 2012-2017 yılları arasında Erzurum'a gelen turist sayılarının yıllar itibarıyla Tablo 3.2'deki gibi olduğunu varsayalım. Bu verilere ait zaman serisi grafiği Görsel 3.2' de çizilerek turist sayısındaki sürekli artan değişim, dönemler itibarıyla gözlemlenebilir.

Tablo 3.2. 2012-2017 Yıllarında Erzurum'a Gelen Turist Sayısı

Yıllar	Turist Sayısı
2012	165.086
2013	178.475
2014	182.059
2015	195.850
2016	204.190
2017	233.228



Bir zaman serisi; günler, haftalar ve yıllar itibarıyla sürekli artıyorsa buna artan zaman serisi denir.



Görsel 3.2. 2012-2017 Yıllarında Erzurum'a Gelen Turist Sayısı

Bir zaman serisi; günler, haftalar, aylar ve yıllar itibarıyla sürekli azalıyorsa buna *azalan zaman* serisi diyebiliriz.



Bir zaman serisi; günler, haftalar ve yıllar itibarıyla sürekli azalıyorsa buna azalan zaman serisi denir.

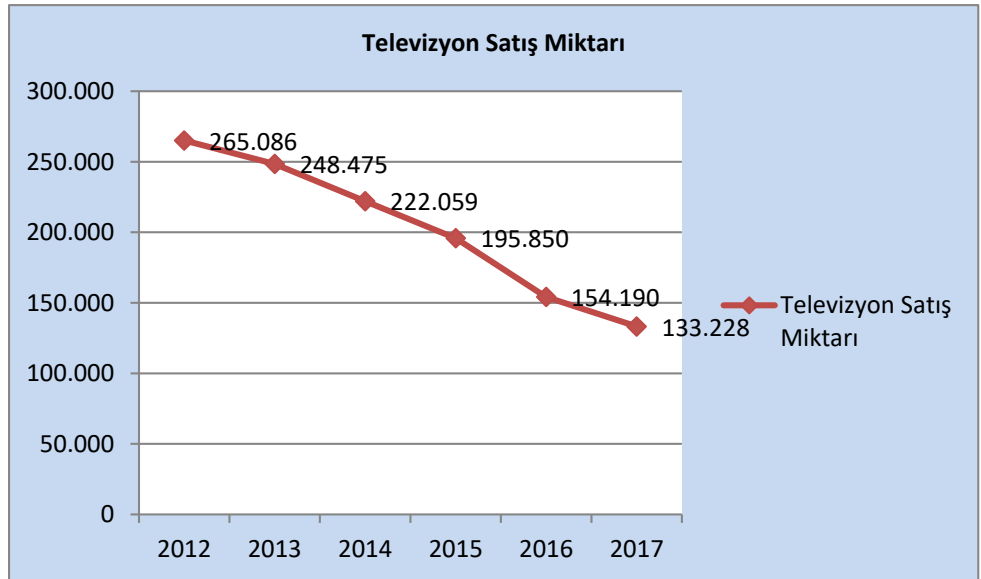


Örnek

- 2012 yılından 2017 yılına kadar televizyon satış miktarının yıllar itibarıyla Tablo 3.3' deki gibi olduğunu varsayalım. Bu verilere ait zaman serisi grafiği Görsel 3.3' de çizilerek televizyon satış miktarındaki sürekli azalan değişim, dönemler itibarıyla gözlemlenebilir.

Tablo 3.3. 2012-2017 Yıllarında Televizyon Satış Miktarı

Yıllar	Televizyon Satış Miktarı
2012	265.086
2013	248.475
2014	222.059
2015	195.850
2016	154.190
2017	133.228



Görsel 3.3. 2012-2017 Yıllarında Televizyon Satış Miktarı Azalan Zaman Serisi Grafiği

Bir zaman serisi; günler, haftalar, aylar ve yıllar itibarıyla hep aynı kalıyorsa buna *durgun zaman serisi* diyebiliriz.



Bir zaman serisi; günler, haftalar ve yıllar itibarıyla hep aynı kalıyorsa buna durgun zaman serisi denir.

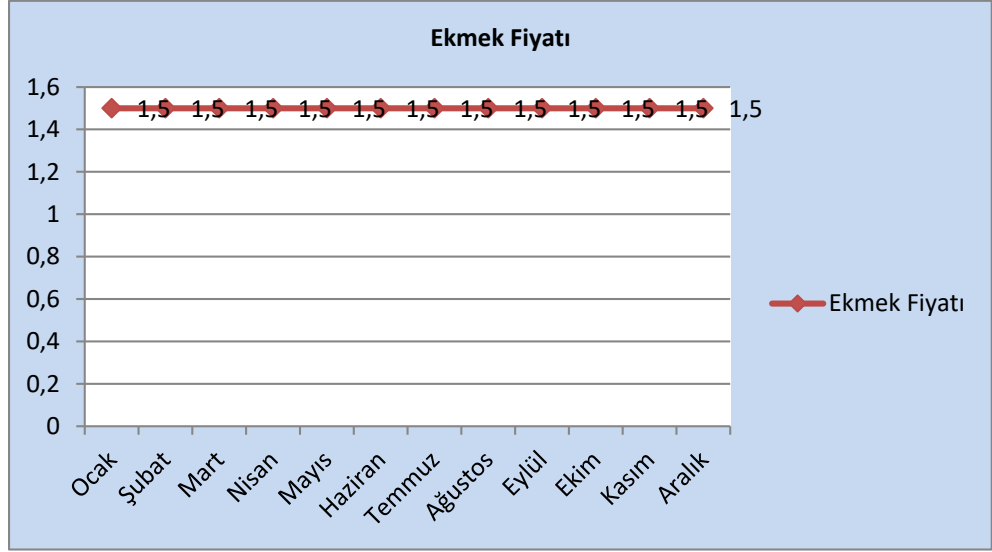


Örnek

- Ocak ayından Aralık ayına kadar ekmek satış fiyatının aylar itibarıyla Tablo 3.4'deki gibi olduğunu varsayalım. Bu verilere ait zaman serisi grafiği Görsel 3.4' de çizilerek ekmek fiyatındaki durgunluk, dönemler itibarıyla gözlemlenebilir.

Tablo 3.4. Ocak-Aralık Arası Ekmek Fiyatı

Aylar	O.	Ş.	M.	N.	M.	H.	T.	A.	Ey	Ek	K	A
Fiyat	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5

**Görsel 3.4.** Ocak Ayından Aralık Ayına Kadar Ekmek Fiyatı Durgun Zaman Serisi Grafiği

Bir zaman serisi; günler, haftalar, aylar ve yıllar itibarıyla tekdüze bir seyir izlemeyip sürekli değişiyorsa buna *dalgalı zaman serisi* diyebiliriz.[1]



Bir zaman serisi; günler, haftalar ve yıllar itibarıyla sürekli değişiyorsa buna dalgalı zaman serisi denir.

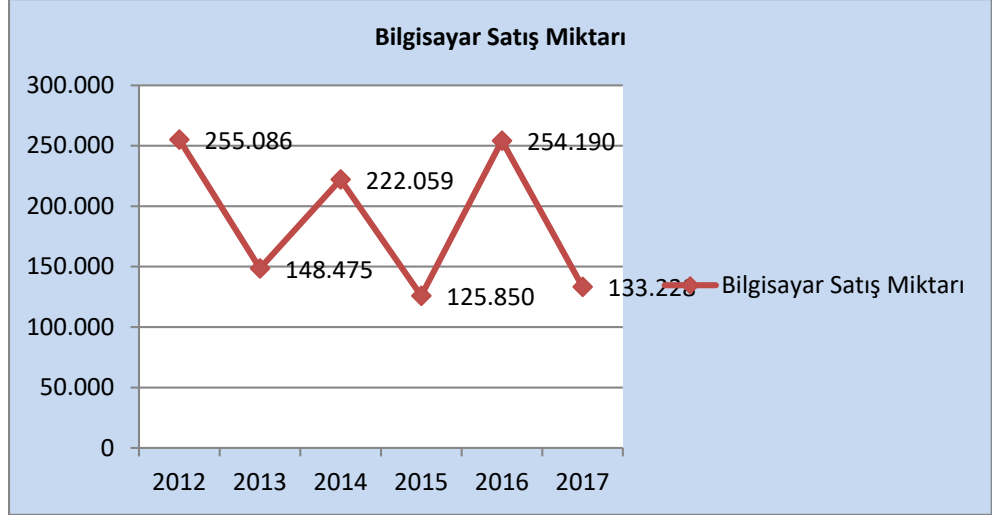


Örnek

- 2012 yılından 2017 yılına kadar bilgisayar satış miktarının yıllar itibarıyla Tablo 3.5'deki gibi olduğunu varsayalım. Bu verilere ait zaman serisi grafiği Görsel 3.5' de çizilerek televizyon satış miktarındaki dalgalı değişim, dönemler itibarıyla gözlemlenebilir.

Tablo 3.5. 2012-2017 Yıllarında Bilgisayar Satış Miktarı

Yıllar	Bilgisayar Satış Miktarı
2012	255.086
2013	148.475
2014	222.059
2015	125.850
2016	254.190
2017	133.228



Görsel 3.5. 2012-2017 Yıllarında Bilgisayar Satış Miktarı Dalgalı Zaman Serisi Grafiği

MEKÂN SERİLERİ



Gözlemler mekâna göre sınıflandırılarak elde ediliyorsa bu serilere mekân serileri denir.

Söz konusu gözlemler mekâna göre sınıflandırılarak elde ediliyorsa bu serilere mekân serileri denir [12]. Mekân (yer) serisi, araştırma konusuna ait gözlem sonuçlarını il, ilçe, bölge, ülke, kıta gibi yer itibarıyla gösteren serilerdir. İllere göre nüfus sayıları, bölgelere göre doğalgaz tüketim miktarları, farklı bölgelerdeki üniversitelerdeki öğrenci sayıları, çeşitli üniversitelerdeki bilimsel yayın sayıları mekân serilerine örnek olarak verilebilir.



Örnek

- 2007 yılına ait bazı ülkelerin ihracat rakamları Tablo 3.6 'da verilmiştir. Tablo 3.6 incelendiğine ikinci sütundaki sayısal veriler ülkeler bazında elde edilmiştir.

Tablo 3.6. Bazı ülkelerin 2007 Yılına Ait Örnek İhracat Rakamları

Ülkeler	İhracat rakamları (Milyon dolar)
Almanya	1.361.000
Çin	1.221.000
ABD	1.140.000
Japonya	799.435
Fransa	558.900
İtalya	474.800



İki ya da daha fazla değişkenin birlikte değişimini gösteren serilere bileşik serileri denir.

BİLEŞİK SERİLER

İki ya da daha fazla değişkenin birlikte değişimini gösteren serilere *bileşik seriler* denir. Gözlem sonuçlarını iki veya daha fazla özelliğe göre düzenleyen seriler, bileşik serilerdir. Bileşik serilerde birden çok özellik ile ilgili bilgiler değerlendirildiği için özellikler arasında bir ilişkinin var olup olmadığı kolaylıkla gözlemlenebilir. Basit bileşik seri iki sütundan meydana gelir. Sütunlardan birinde değişkene ait veriler, diğer sütunda ise ilişki olduğu düşünülen özelliğe ait gözlenen değerler gösterilir.



Örnek

- Öğrencilerin belli bir derse çalışma zamanı ile aldıkları puanlar Tablo 3.7' de verilmiştir. Dikkat edilirse tabloda iki değişken ve bu değişkenlere ait veriler ile iki değişken arasındaki ilişki, farklı istatistik teknikler uygulanarak bulunabilir.

Tablo 3.7. Ders Çalışma Zamanı İle Alınan Notlar

Ders Çalışma Zamanı	Alınan Puan					Toplam
	50 den az	70 den az	80 den az	90 dan az	100 den az	
2 saatten az	1					1
4 saatten az		2				2
6 Saatten az			1			1
7 saatten az				1		1
Toplam	1	2	1	1		5

FREKANS (BÖLÜNME) SERİLERİ

Gözlemlerin maddi bir özelliğe göre sıralanmasıyla *bölünme serisi* elde edilir. Mekân ve zaman özelliğinin dışında kalan tüm özellikler, maddi özellikler olarak kabul edilir.



Örnek

- Araştırma için veri toplanacak kişilerin cinsiyeti, medeni durumu, öğrenim durumu, üretim miktarı gibi özellikler birer maddi özelliktir.

Cinsiyet, medeni durumu, öğrenim durumu, üretim miktarı gibi özellikler birer maddi özelliktir. Dikkat edilirse cinsiyet ve üretim miktarını örnek vermekteyiz. Bu ikisi arasındaki farkı şu şekilde açıklayabiliriz:

Bölünme serileri sınıflama özelliklerine göre nitel bölünme serileri ve nicel bölünme serileri olmak üzere iki grupta incelenmektedir. Bir ülkede yaşayan



Bir ülkede yaşayan nüfusun eğitim, cinsiyet gibi özellikler dikkate alınarak bölünmesi nitel bölünme serisi'ne örnek olarak verilebilir.

nüfusun eğitim, cinsiyet gibi özellikler dikkate alınarak bölünmesi nitel bölünme serisini ifade ederken, aynı nüfusun yaş ve gelir durumları gibi özelliklere göre bölünmesi nicel bölünme serisini ifade etmektedir.[10]

Nitel (Kalitatif) Bölünme Serileri

Gözlemlerin, medeni durum, cinsiyet, eğitim durumu gibi sayısal olmayan değişkenlerin özelliğine göre yani nitel bir özelliğe göre bölünerek oluşturulduğu seri *nitel bölünme serisidir* [8]. Aşağıda herhangi bir iş yerinin personelinin medeni durumu sayısı verilmiştir.

Tablo 3.8. Bir İş Yerinde Çalışanların Medeni Durumlarına Göre Dağılımı

Medeni Durum	Personel Sayısı
Bekâr	9
Evli	32

Nicel (Kantitatif) Bölünme Serileri

Sayısal değer alan özelliklerin sınıflandırılması ile nicel seriler oluşmaktadır. Öğrencilerin boy uzunlukları, yaşları, ağırlıkları, aylık harcamalar, bir iş yerinde çalışan işçilerin ücretleri, çalışan işçi sayısına göre iş yerleri nicel bölünme serilerine örneklerdir. İstatistiksel analizler bakımından oldukça önemli olan bu seriler üç başlıkta incelenebilir: Mutlak frekans serileri, nispi frekans serileri ve kümülatif frekans serileri [1].

Mutlak Frekans Serileri

Mutlak frekans serileri, basit seriler, sınıflandırılmış seriler ve gruplandırılmış seriler olmak üzere üç başlık altında incelenebilir.

Basit seriler

Verilerin küçükten büyüğe veya büyükten küçüğe doğru sıralanmasıyla elde edilen serilere, basit seriler denir. Genelde X_i olarak tanımlanırlar. Basit serilerde gözlem sayısı ile gösterilirse, serideki gözlem değerlerinin toplamı

$$\sum_{i=1}^n X_i = X_1 + X_2 + X_3 \dots X_n$$

şeklinde ifade edilir [11].



Bir ülkede yaşayan nüfusun yaş ve gelir durumları gibi sayısal özellikler dikkate alınarak bölünmesi nicel bölünme serisidir.



Örnek

- Ekonometri bölümü öğrencilerinin İstatistik dersinden aldıkları notlar aşağıda verilmiştir.
- 87, 48, 74, 75, 64, 61, 74, 63, 61, 58, 49, 74, 75, 90, 74, 63, 90, 61, 48
- Bu seriye ait basit seri, aşağıdaki şekilde elde edilecektir.
- Ana küttleden seçilmiş örnek veriler, küçükten büyüğe doğru sıralandığında basit seri elde edilmiş olur. Yukarıda, ham veri setindeki değerlerin küçükten büyüğe doğru sıralanması ile aşağıdaki gibi bir basit seri elde edilir:
- 48, 48, 49, 58, 61, 61, 61, 63, 63, 64, 74, 74, 74, 74, 75, 75, 87, 90, 90

Sınıflandırılmış seriler

Basit serinin daha anlaşılır hâle gelmesi için veriler sınıflandırılır. Sınıflandırılmış serilerde her bir X değerinin karşısına o değer frekansı yazılır. Frekans, tekrarlanma sayısı ya da başka bir ifade ile herhangi bir nitel veya nicel değişken değerinin anakütlede veya örnekleme kaç kez tekrarlandığını gösteren sayıdır [5].



Örnek

- 80 kişilik bir sınıfta 23 farklı not varsa 80 kişinin notu 23 sınıf hâlinde özetlenmiş olur.
- 87, 48, 74, 75, 64, 61, 74, 63, 61, 58, 49, 74, 75, 90, 74, 63, 90, 61, 49, 50, 60, 48, 50, 50, 54, 60, 64, 65, 50, 54, 51, 50, 54, 57, 58, 60, 65, 64, 68, 48, 50, 60, 55, 55, 45, 85, 87, 90, 75, 65, 85, 58, 68, 74, 56, 68, 70, 85, 65, 64, 60, 70, 45, 65, 70, 80, 50, 56, 75, 85, 65, 75, 65, 60, 75, 80, 70, 50, 68, 70



Her gözlem değerinin karşısına bu gözlem değerinin kaç kez tekrarlandığı (frekans) yazılarak elde edilen seriye sınıflandırılmış seri denir.

Yukarıdaki basit serinin sınıflandırılmış seri şeklindeki hâli aşağıda verilmiştir:

45, 45, 48, 48, 48, 49, 49, 50, 50, 50, 50, 50, 50, 50, 50, 51, 54, 54, 54, 55, 55, 56, 56, 57, 57, 58, 58, 58, 60, 60, 60, 60, 60, 60, 61, 61, 61, 63, 63, 64, 64, 64, 64, 65, 65, 65, 65, 65, 65, 68, 68, 68, 68, 70, 70, 70, 70, 70, 74, 74, 74, 74, 75, 75, 75, 75, 75, 80, 80, 85, 85, 85, 85, 87, 87, 90, 90, 90

Kesikli veriler için düzenlenen frekans dağılımı ya da frekans tabloları, iki sütundan oluşmaktadır. Birinci sütunda, değişkenin almış olduğu farklı değerler yer alırken; ikinci sütun, bu değerlerin sıklıklarını yani seri içinde tekrar eden sayılarını gösterir. Her gözlem değerinin karşısına bu gözlem değerinin kaç kez tekrarlandığı (frekans) yazılarak elde edilen seriye **sınıflandırılmış seri** denir. Değişkenin aldığı değerler sütunu X_i , frekansları ise f_i ile belirtilir [4]. Yukarıdaki basit seride yer alan gözlem değerleri ile aşağıdaki gibi bir sınıflandırılmış seri hazırlanır. Sınıflandırılmış seri hazırlandığında herhangi bir bilgi kaybı söz konusu olmaz.

Tablo 3.9. Örnek Sınıflandırılmış Seri

Notlar	Frekans
45	2
48	3
49	2
50	8
51	1
54	3
55	2
56	2
57	2
58	3
60	6
61	3
63	2
64	4
65	7
68	4
70	5
74	4
75	6
80	2
85	4
87	2
90	3

Bir araştırma konusu ile ilgili elimizdeki seride yüzlerce hatta binlerce gözlem değeri olması durumunda, sınıflandırılmış seriler hazırlamak suretiyle özet rakamlar kullanılarak bilgi kaybına neden olmadan verileri takdim etmek mümkün olabilir.

Gruplandırılmış seriler

Verilerin gruplandırılması, özellikle denek sayısı fazla olduğunda, hem veriler üzerinde yapılacak hesaplamaları kolaylaştırır hem de verilerin kolay anlaşılır biçimde okuyucuya sunulmasını sağlar [7]. Sınıflandırılmış serileri bir basamak daha genişleterek gruplandırılmış seriler oluşturulur. Genelde gözlem sayısının çok fazla olduğu durumlarda kullanılır. Böylece durum daha net görülebilir. Gruplamada gruplar genellikle eşit büyüklükte alınır. Gruplandırılmış bir seri altı adımda oluşturulabilir:

Adım 1: Elde edilen veriler basit seri hâline getirilir.

Adım 2: Verilerin en büyüğünden en küçüğü çıkarılarak değişim aralığı bulunur.

Değişim Aralığı (K) = $X_{\max} - X_{\min}$

Adım 3: Olayın kaç grupta inceleneceği (m) belirlenir.

Adım 4: Sınıf büyüklüğü (A) belirlenir. Sınıf büyüklüğü dağılımdaki en büyük değerden en küçük değerin çıkarılıp grup sayısına bölünmesiyle elde edilir. Sınıf büyüklüğü aynı zamanda sınıf üst sınırı ile alt sınırı arasında farktır.

$$A = k/m$$

Adım 5: Serideki en küçük değere sınıf büyüklüğü eklenerek 5. adımda serideki en büyük değere ulaşacak şekilde gruplar oluşturulur.

Adım 6: Her gruba düşen değer sayısı belirlenir [9].



Örnek

- Bir kreşteki 10 çocuğun ağırlıklarına (kg) göre dağılımları aşağıdaki gibidir:
- 13, 11, 11, 12, 15, 17, 16, 12, 14, 15, 14, 17, 10, 12, 14, 15, 14, 13, 15, 13
- Verilen ağırlıklarla eşit aralıklı 5 sınıflı bir gruplandırılmış seri oluşturulur.

1. Adım: Seri, küçükten büyüğe doğru sıralanarak aşağıdaki gibi basit seriye dönüştürülür.

10, 11, 11, 12, 12, 12, 13, 13, 13, 14, 14, 14, 14, 15, 15, 15, 15, 16, 17, 17

2. Adım: Değişim Aralığı (K) = $X_{\max} - X_{\min} = 17 - 10 = 7$

3. Adım: Grup Sayısı = 5

4. Adım: Sınıf Büyüklüğü = $7/5 = 1.4$

5. Adım: Gruplar aşağıdaki Tablo 3.10'daki gibi oluşur:

Tablo 3.10. Örnek Gruplar

Aralıklar
10.0 ile 11,4
11,4 ile 12,8
12,8 ile 14,2
14,2 ile 15,6
15,6 ile 17,0

6. Adım: Her aralığa düşen birim sayısı karşısına yazıldığında gruplandırılmış seri;

Tablo 3.11. Örnek Gruplandırılmış Seri

Ağırlıklar (Kg)	Çocuk Sayısı
10.0 - 11,4	3
11,4 - 12,8	3
12,8 - 14,2	7
14,2 - 15,6	4
15,6 ile 17,0	3



Bir grubun bittiği nokta ile bu grubun peşi sıra gelen grubun başlama noktası arasında boşluk yoksa bu tür gruplandırılmış serilere sürekli gruplandırılmış seri denir.



Örnek

- Bir sınıftaki 20 öğrencinin matematik sınav sonuçları aşağıdaki gibi olursa 4 gruplu bir gruplandırılmış seri nasıl oluşturulur? Yukarıdaki adımları takip ederek 4 gruplu bir seri elde ediniz.
- 45, 50, 82, 70, 50, 60, 84, 76, 90, 70, 80, 40, 60, 36, 80, 60, 90, 50, 70, 80

İki tip gruplandırılmış seri hazırlanabilir: Sürekli gruplandırılmış seriler ve kesikli gruplandırılmış seriler. Yukarıda gruplandırılmış seriye sürekli gruplandırılmış seri denir. Çünkü bir grubun bittiği noktada bir başka grup başlamaktadır. Gruplar arasında boşluk bulunmamaktadır. Ancak bir grubun bittiği nokta ile bu grubun peşi sıra gelen grubun başlama noktası arasında boşluk varsa bu tür gruplandırılmış seriye kesikli gruplandırılmış seri denir. Aşağıda sürekli ve kesikli gruplandırılmış serilere örnek verilmiştir:

Tablo 3.12. Sürekli Gruplandırılmış Seri

Gruplar	Frekans
5-10	3
10-15	4
15-20	2
20-25	3

Tablo 3.13. Kesikli Gruplandırılmış Seri

Gruplar	Frekans
5-10	2
11-16	6
17-22	4
23-28	3



Bir grubun bittiği nokta ile bu grubun peşi sıra gelen grubun başlama noktası arasında boşluk varsa bu tür gruplandırılmış serilere kesikli gruplandırılmış seri denir.

Gruplandırılmış serilerde her grubun alt ve üst değerlerine grup uçları denir. Yukarıda yer alan sürekli gruplandırılmış seride grup uçları; 5-10, 10-15, 15-20, 20-25 değerleridir. Kesikli gruplandırılmış serideki grup uçları ise; 5-10, 11-16, 17-22, 23-28 değerleridir. Bu ikili rakamlardan birinciler alt uçlar, ikinciler ise üst uçlardır.

Kesikli gruplandırılmış serilerde bir grubun üst ucu ile bir sonraki sınıfın alt ucu toplanıp ikiye bölünerek o sınıfın üst sınırı bulunmuş olur. Sürekli gruplandırılmış seride grup uçları aynı zamanda grup sınırlarıdır.

Örneğin, yukarıdaki kesikli gruplandırılmış seride ilk grubun üst sınırı $(10+11)/2 = 10,5$ olarak bulunur. Bu değer aynı zamanda ikinci grubun alt sınırıdır.

Her sınıfın alt ve üst sınıf sınırları veya uçları toplanarak ikiye bölündüğünde sınıfa ait sınıf değeri elde edilmiş olur. Sınıf değeri, ilgili sınıfı temsil eden değerdir.

Nispi Frekans Serileri

Buraya kadar bahsedilen dağılımlar mutlak frekans dağılımlarıdır. Bahse konu olan sınıflandırılmış seriler hazırlanırken her değerin kaç kez tekrarlandığı karşısına yazılır. Aynı şekilde gruplandırılmış seriler hazırlanırken her bir sınıf aralığına düşen değer sayısı da karşısına yazılır. Elde edilen frekans değerleri sayımla elde edilir.

Nispi frekans dağılımları hazırlanırken her değere veya sınıf aralığına karşı gelen frekans, toplam frekansa oranlanmaktadır. Bu şekilde her değerin tüm değerler içindeki nispeti belirlenmektedir. Bazen karşılaştırma kolaylığı olsun diye nispi frekanslar yüz ile çarpılarak yüzde frekanslar da elde edilmektedir.

Tablo 3.14. Gruplandırılmış Seriyeye ait Nispi Frekans Tablosu

Gruplar	Frekans	Nispi Frekanslar	Yüzde Frekanslar
5-10	3	3/12	25
10-15	4	4/12	33
15-20	2	2/12	17
20-25	3	3/12	25

Yukarıdaki gruplandırılmış seride; 5 ile 10 aralığındaki rakamlar %25'i, 10 ile 15 aralığındaki rakamlar %33'ü, 15 ile 20 arasındaki rakamlar %17'sini, 20 ile 25 arasındaki rakamlar %25'i kapsamaktadır.

Kümülatif Frekans Serileri

Kümülatif kelimesi, yığılmış, birikmiş veya toplanmış anlamına gelir. Gerek frekans serisinde, gerekse sınıflı serilerde belli bir değere eşit, küçük veya büyük olan birim sayılarını göstermek için hesaplanan frekanslar kümülatif (birikimli) frekanslar olarak adlandırılır [4]. İki tip kümülatif frekans serisi hazırlamak mümkündür: "...den az kümülatif frekans serileri" ve "...den çok kümülatif frekans serileri".

"...den az kümülatif frekans serileri" oluşturulurken belirli bir değerden az olan değerler sayılır. Gruplandırılmış serilerde ise grup üst sınırından küçük olan değerler sayılır. "...den az kümülatif frekans değerleri" toplam frekansa oranlanarak "...den az nispi kümülatif frekans değerleri" elde edilebilir.

Aşağıda seriye göre ilk grubun "...den az kümülatif frekansı" 10'dan az olan değer sayısıdır. Bu değer ilk grubun mutlak frekansı olan 3'tür. İkinci grubun "...den az kümülatif frekansı" 15'ten küçük değerlerin sayısıdır. Bu da birinci ve ikinci grubun mutlak frekansları toplamıdır. Yani $3 + 4 = 7$ 'dir. Üçüncü grubun "...den az kümülatif frekansı" 20'den küçük değerlerin sayısıdır. Bu da birinci, ikinci ve üçüncü grubun mutlak frekansları toplamıdır. Yani $3 + 4 + 2 = 9$ 'dur. Dördüncü grubun "...den az kümülatif frekansı" 25'ten küçük değerlerin sayısıdır. Bu da

birinci, ikinci, üçüncü ve dördüncü sınıfın mutlak frekansları toplamıdır. Yani $3 + 4 + 2 + 3 = 12$ 'dir. Son grubun "...den az kümülatif frekansı" toplam frekansa eşittir. Yine aşağıdaki seride "...den az kümülatif frekanslar" toplam frekansa bölünerek "...den az kümülatif nispi frekanslar" elde edilmiştir.

Tablo 3.15. ...den Az Kümülatif Frekans Tablosu

Gruplar	Frekans	...den az küm.frekans	...den az küm. nispi frekans
5-10	3	3	3/12
10-15	4	7	7/12
15-20	2	9	9/12
20-25	3	12	12/12

"...den çok kümülatif frekans serileri" oluşturulurken belirli bir değerden büyük olan değerler sayılır. Gruplandırılmış serilerde ise grup alt sınırından büyük olan değerler sayılır. "...den çok kümülatif frekans değerleri" toplam frekansa oranlanarak "...den çok kümülatif nispi frekans değerleri" de elde edilebilir.

Aşağıda seriye göre ilk sınıfın "...den çok kümülatif frekansı" 5'ten büyük olan değer sayısıdır. Bu da birinci, ikinci, üçüncü ve dördüncü grubun mutlak frekansları toplamıdır. Yani $3 + 4 + 2 + 3 = 12$ 'dir. İkinci grubun "...den çok kümülatif frekansı" 10'dan büyük olan değer sayısıdır. Bu da ikinci, üçüncü ve dördüncü grubun mutlak frekansları toplamıdır. Yani $4 + 2 + 3 = 9$ 'dur. Üçüncü grubun "...den çok kümülatif frekansı" 15'ten büyük olan değer sayısıdır. Bu da üçüncü ve dördüncü grubun mutlak frekansları toplamıdır. Yani $2 + 3 = 5$ 'tir. Dördüncü grubun "...den çok kümülatif frekansı" 20'den büyük olan değer sayısıdır. Bu da dördüncü sınıfın mutlak frekansı 3'e eşittir. Yine aşağıdaki seride "...den çok kümülatif frekanslar" toplam frekansa bölünerek "...den çok kümülatif nispi frekanslar" elde edilmiştir.

Tablo 3.16. ...den Çok Kümülatif Frekans Tablosu

Gruplar	Frekans	...den çok küm.frekans	...den çok küm. nispi frekans
5-10	3	12	12/12
10-15	4	9	9/12
15-20	2	5	5/12
20-25	3	3	3/12



Bireysel Etkinlik

- Seri nedir açıklayarak, verilerin nasıl seriye dönüştüğünü inceleyiniz.
- Veriler hangi serilere dönüşebilir? Herbir seri türü ile ilgili kendi örneklerinizi oluşturunuz.
- Etrafınızdaki 30 kişinin ağırlıklarından 5 gruplu gruplandırılmış seri elde ediniz.



Özet

- Bu ünite de seri tarif edilerek, seri türlerinin zaman, mekân, frekans ve kümülatif seriler olduğuna değinilmiştir. Elde edilen ham verilerin anlaşılabilirliğini ve hesaplanabilirliğini kolaylaştırmak için belli bir kurala göre sınıflandırılarak tablolaştırılması sonucu oluşan diziye seri denir.
- Zaman serileri zamana, mekân serileri mekâna göre sınıflandırılarak oluşturulur. Gözlemlerin zaman birimlerine göre sınıflandırıldığı serilere *zaman serileri* denir. Zaman serisi verisi farklı zaman periyodları için toplanan veridir. Söz konusu gözlemler mekâna göre sınıflandırılarak elde ediliyorsa bu serilere *mekân serileri* denir. Mekân (yer) serisi, araştırma konusuna ait gözlem sonuçlarını il, ilçe, bölge, ülke, kıta gibi yer itibarıyla gösteren serilerdir.
- İki ya da daha fazla değişkenin birlikte değişimini gösteren serilere *bileşik seriler* denir. Gözlem sonuçlarını iki veya daha fazla özelliğe göre düzenleyen seriler, bileşik serilerdir.
- Gözlemlerin maddi bir özelliğe göre sıralanmasıyla *frekans (bölünme) serisi* elde edilir. Mekân ve zaman özelliğinin dışında kalan tüm özellikler maddi özellikler olarak kabul edilir. Frekans serileri, nitel ve nicel seriler olmak üzere ikiye ayrılırlar. Gözlemlerin medeni durum, cinsiyet, eğitim durumu gibi sayısal olmayan değişken özelliğine göre yani nitel bir özelliğe göre bölünerek oluşturulduğu seri *nitel bölünme serisi*' dir. Sayısal değer alan özelliklerin sınıflandırılması ile *nicel seriler* oluşmaktadır. Nicel seriler kendi aralarında basit, sınıflandırılmış, gruplandırılmış seriler olmak üzere üçe ayrılırlar.
- Basit seri, verilerin küçükten büyüğe veya büyükten küçüğe doğru sıralanmasıyla elde edilen serilerdir.
- Sınıflandırılmış seri, her bir x değerinin karşısına frekans değeri (tekrarlama sayısı) yazılarak elde edilen serilerdir.
- Gruplandırılmış seri, serideki en büyük değerden en küçük değer çıkarılıp grup sayısına bölünerek elde edilen sınıf büyüklüğüne göre oluşturulan grup alt sınır ve üst sınır aralığına göre frekansları yazmak suretiyle elde edilen serilerdir. İki tip gruplandırılmış seri hazırlanabilir: Sürekli gruplandırılmış seriler ve kesikli gruplandırılmış seriler. *Sürekli gruplandırılmış seri*'de bir sınıfın bittiği noktada bir başka sınıf başlamaktadır. Sınıflar arasında boşluk bulunmamaktadır. Ancak bir sınıfın bittiği nokta ile bu sınıfın peşi sıra gelen sınıfın başlama noktası arasında boşluk varsa bu tür gruplandırılmış seriye *kesikli gruplandırılmış seri* denir. Gruplandırılmış serilerde her sınıfın alt ve üst değerlerine sınıf uçları denir.
- Nispi frekans dağılımları hazırlanırken her değere veya sınıf aralığına karşı gelen frekans, toplam frekansa oranlanmaktadır. Bu şekilde her değer için nispeti belirlenmektedir. Bazen karşılaştırma kolaylığı olsun diye nispi frekanslar yüz ile çarpılarak yüzde frekanslar da elde edilmektedir.

DEĞERLENDİRME SORULARI

1. Yanda verilen tablo hangi seriye bir örnek olabilir?

- a) Zaman serisi
- b) Mekân serisi
- c) Kümülatif seri
- d) Frekans serisi
- e) Bileşik seri

İller	Üretilen Tahıl Miktarı (Ton)
Eskişehir	221000
Elazığ	79435
Erzincan	11400

2. Yanda verilen tablo hangi seriye bir örnek olabilir?

- a) Gruplandırılmış seri
- b) Sınıflandırılmış seri
- c) Zaman serisi
- d) Kümülatif seri
- e) Bileşik seri

Gruplar	Frekanslar
3-5	2
5-7	7
7-9	3

3. Yanda verilen gruplandırılmış serideki üçüncü sınıfın alt sınırı hangisidir?

- a) 8
- b) 8.5
- c) 9
- d) 9.5
- e) 10

Gruplar	Frekanslar
3-5	2
6-8	7
9-11	3

Gruplar	Frekans	... den az kümülatif Frekans
3-5	3	3/18
6-8	7	10/18
9-11	6	?/18
12-14	2	18/18

4. Yukarıda verilen kümülatif nispi frekans serisinde soru işareti olan yere aşağıdakilerden hangisi gelmelidir?

- a) 11
- b) 13
- c) 16
- d) 17
- e) 18

5. Yanda verilen gruplandırılmış serinin sınıf büyüklüğü hangisidir?

- a) 1
b) 1.5
c) 2
d) 2.5
e) 3

Gruplar	Frekanslar
3-5	2
5-7	7
7-9	3
9-11	5

6. Elimizdeki basit seride 2, 2, 2, 3, 3, 5, 7, 7, 7, 7, 10, 12, 12, 12, 12, 12, 12, 12 değerleri yer almaktadır. Bu serideki değerlerle eşit aralıklı beş sınıflı gruplandırılmış bir seri hazırlanırsa, her bir sınıfın sınıf büyüklüğü kaç olur?

- a) 1
b) 1.5
c) 2
d) 2.5
e) 3

7. Elimizdeki basit seride 2, 2, 2, 3, 3, 5, 7, 7, 7, 7, 10, 12, 12, 12, 12, 12, 12, 12 değerleri yer almaktadır. Bu serideki değerlerle eşit aralıklı beş sınıflı gruplandırılmış bir seri hazırlanırsa birinci sınıfın üst sınıf sınırı kaç olur?

- a) 2
b) 3
c) 4
d) 5
e) 6

8. Elimizdeki basit seride 2, 2, 2, 3, 3, 5, 7, 7, 7, 7, 10, 12, 12, 12, 12, 12, 12, 12 değerleri yer almaktadır. Bu serideki değerlerle eşit aralıklı beş sınıflı gruplandırılmış bir seri hazırlanırsa birinci sınıfın mutlak frekansı kaç olur?

- a) 2
b) 3
c) 4
d) 5
e) 6

9. Elimizdeki basit seride 2, 2, 2, 3, 3, 5, 7, 7, 7, 7, 10, 12, 12, 12, 12, 12, 12, 12 değerleri yer almaktadır. Bu serideki değerlerle eşit aralıklı beş sınıflı gruplandırılmış bir seri hazırlanırsa ikinci sınıfın nispi frekansı kaç olur?

- a) 1/18
b) 2/18
c) 3/18
d) 4/18
e) 5/18

10. Yanda verilmiş olan sınıflandırılmış serinin değişim aralığı hangisidir?

- a) 30
- b) 36
- c) 40
- d) 49
- e) 55

Puanlar	Frekans
36	15
45	25
57	12
68	31
72	27
75	19
76	45
79	29
80	11
85	27

Cevap Anahtarı

1.b, 2.a, 3.b, 4.c, 5.c, 6.c, 7.c, 8.d, 9.a, 10.d

YARARLANILAN KAYNAKLAR

- [1] Tekin, A. (2012). Temel İstatistik Dersleri. Konya: Eğitim Yayınları.
- [2] Alpar, R. (2010). Spor, Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik-Güvenirlik. Ankara: Detay Yayıncılık.
- [3] Orhunbilge, N. (2013). İşletme İstatistiğinin Temelleri (4. baskı). Ankara: Nobel Yayınları.
- [4] Çakır, F. (2000), Sosyal Bilimlerde İstatistik. İstanbul: Alfa yayıncılık.
- [5] Gürsakal, N., Oğuzlar, A, Gürsakal, S. (2022). Betimsel İstatistik (10. Baskı). Bursa: Dora Yayıncılık .
- [6] Çepni, E. (2019). Ekonomik Göstergeler ve İstatistikler Rehberi (8. baskı). Ankara: Seçkin Yayıncılık.
- [7] Sümbüloğlu, K., Sümbüloğlu, V. (2009). Biyoistatistik (13. basım). Ankara: Hatipoğlu Yayıncılık.
- [8] Özsoy, O. (2010). İktisatçılar ve İşletmeciler için İstatistik (3. basım). Ankara: Siyasal Kitabevi.
- [9] Serper, Ö. (2004). Uygulamalı İstatistik-I (5.baskı). Bursa: Ezgi Kitabevi.
- [10] Johnson, A. R., Wichern, W. D. (1998). Applied Multivariate Statistical Analysis (4th Edition). London: Prentice-Hall.
- [11] Yıldız, N. T. (2008). Uygulamalı İstatistik (3. baskı). Ankara: Nobel Yayıncılık.
- [12] Yılmaz, B. (2010). İstatistik (3. baskı). Ankara: Nobel Yayıncılık.

GRAFİKLER



İÇİNDEKİLER

- Histogram
- Frekans Poligonu
- Kümülatif Frekans Grafikleri
- Sütun Grafiği
- Daire Grafiği
- Pareto Grafiği
- Serpilme Diyagramı



HEDEFLER

- Bu üniteyi çalıştıktan sonra;
 - Grafik kavramını ifade edebilecek,
 - Gözlem değerlerini görselleştirebilecek,
 - Grafik çizebilecek,
 - Gözlem değerleri için uygun grafik türünü belirleyebilecek,
 - Grafik yorumlama hakkında detaylı bilgi sahibi olabileceksiniz.

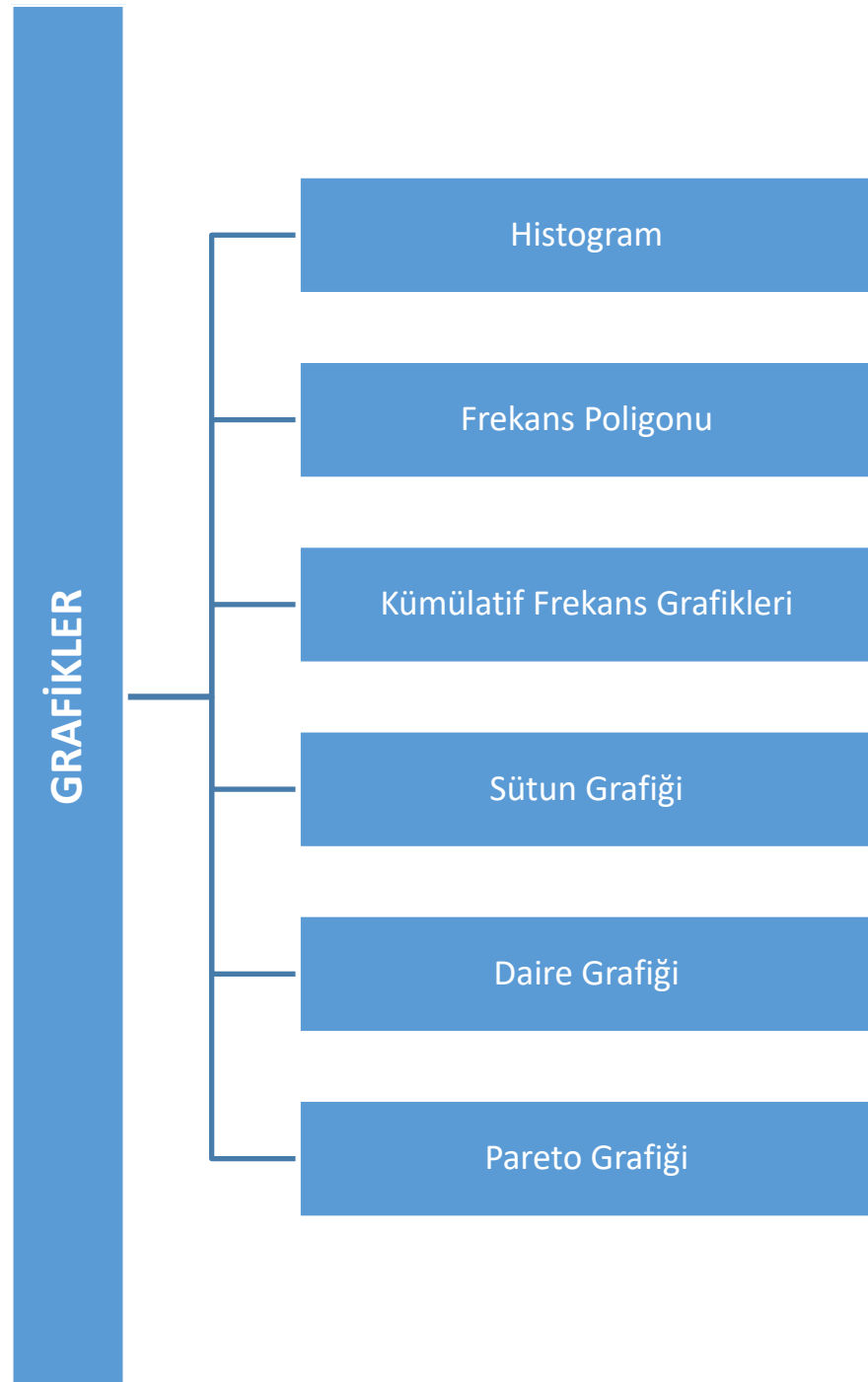


Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ

**Prof. Dr. Mehmet
Suphi ÖZÇOMAK**

ÜNİTE 4





Grafik, gözlem değerlerinin matematiksel ve bilimsel temellere sahip şekiller hâlinde ifade edilmesidir.

GİRİŞ

Bilindiği üzere istatistiğin uğraş alanlarından biri de özetleme ve görselleştirmedir. Frekans dağılımlarının bir tablo şeklinde sunulması bir özetleme şeklidir. Oysa tablolar her zaman kolay anlaşılır olmayabilir. Grafikler gözlem değerlerinin görselleştirilmesinde kullanılır. *Grafik, “gözlem değerlerinin matematiksel ve bilimsel temellere sahip şekiller hâlinde ifade edilmesi” şeklinde tanımlanabilir.*

Grafikler gözlem değerlerinin daha kolay anlaşılmasını sağlamaktadır.

Çünkü grafikler temsil ettikleri olayların bileşimini ve değişimindeki ana eğilimi tüm canlılığı ile ilk bakışta ortaya koymaktadırlar. Bu nedenle rakamlarla uğraşmak istemeyen veya rakamlardan anlam çıkarmakta zorluk çekenler için grafikler daha fazla yardımcı olmaktadır.

Grafik verilmeksizin yalnız rakamlar hâlindeki sonuçları inceleyen bir şahıs, olayın genel eğiliminden çok ayrıntısına takılacak ve bu yüzden olayın genel özelliklerini kavrayamayacaktır. Oysa iyi çizilmiş bir grafik üzerinde gösterilen sonuçlar görsel olduğundan şahıslar tarafından daha anlaşılır olacaktır. Örneğin; yıllar itibarıyla bir ülkenin nüfusunun incelendiği bir araştırmada rakamlardan çok, grafikler daha anlaşılır olacaktır. Grafik üzerinde yıllar itibarıyla nüfusun artıp artmadığı veya nasıl değiştiği ilk bakışta kolaylıkla anlaşılabilir. Oysa rakamlarla bu durum ifade edilmeye çalışıldığında durumun anlaşılması grafiklerde olduğu gibi kolay olmayacaktır [1].

Bu bölümde grafik çeşitlerinden sırasıyla histogram, frekans poligonu, kümülatif frekans dağılımları, daire grafiği, sütun grafiği, pareto grafiği ve serpilme diyagramı ele alınmıştır.

HİSTOGRAM



Histogram, gruplandırılmış seriler için oluşturulan bir diktörtgenler dizisidir.

Histogram, gruplandırılmış seriler için oluşturulan bir diktörtgenler dizisidir. Histogram; yatay ekseninde değişkenin aldığı değerlerin, dikey ekseninde ise frekansların bulunduğu ve her aralığın frekansı ile orantılı yüksekliklere sahip diktörtgenlerin gösterildiği bir yoğunluk grafiğidir. Bu diktörtgenlerin tabanları gruplandırılmış serideki her bir sınıfın sınıf büyüklüğünü, yükseklikleri ise sınıf frekansını göstermektedir.

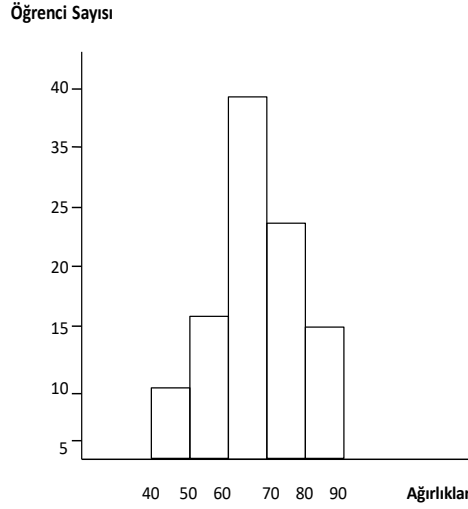
Histogram çizilirken yatay ekseninde gruplandırılmış serinin sınıf sınırları, dikey ekseninde ise frekanslar yer almaktadır. Sınıf aralıkları ve frekans değerleri eksenlerde belirlendikten sonra sınıf sınırlarının alt ve üst sınırlarından frekans değerlerine kadar birer dikme çizilir. Gruplandırılmış serilerde sınıfların frekanslarının sınıf sınırları içerisinde düzgün dağıldığı kabul edildiğinden, çizilen dikmeler yatay eksene paralel bir çizgi ile birleştirilerek diktörtgenler elde edilir. Bu diktörtgenlerin tamamı histogramı oluşturmaktadır [2]. Görsel 4.1’de 100 öğrenciye ait ağırlık dağılımı histogram da yer almaktadır.



Örnek

- Örnek 4.1. Bir sınıftaki 100 öğrencinin ağırlıklarına göre dağılımları aşağıda verilmiştir. Ağırlıkların dağılımının histogramını çizin.

Ağırlıklar	Öğrenci sayısı
40 - 50	10
50 - 60	16
60 - 70	38
70 - 80	22
80 - 90	14



Görsel 4.1. Ağırlık Dağılımı İçin Oluşturulan Histogram

Görsel 4.1'de histogram grafiği incelendiğinde; en yüksek olan dikdörtgen en fazla öğrencinin olduğu 60 – 70 sınıf aralığını, en kısa dikdörtgen ise en az öğrencinin olduğu 40 – 50 sınıf aralığını göstermektedir. Buradan öğrencilerin en fazla 60 – 70 kilo aralığında, en az ise 40 – 50 kilo aralığında oldukları anlaşılmaktadır.

FREKANS POLİGONU

Frekans poligonu, histogramdaki sınıf sınırlarının orta noktalarını apsis, sınıf frekanslarını ordinat kabul eden noktaların doğru parçaları ile birleştirilmesi ile elde edilen grafik çeşididir. Kısacası histogramı oluşturan dikdörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe frekans poligonu denir. *Frekans poligonu, sınıf sınırlarının mümkün olduğunca küçültülmesi durumunda bir eğriye dönüşür. Söz konusu eğriye ise frekans eğrisi adı verilir.*

Frekans poligonunun yatay eksen üzerindeki başlangıç noktası ilk sınıftan bir önceki farazi sınıfın orta noktası iken, bitiş noktası ise son sınıftan bir sonraki farazi



Histogramı oluşturan dikdörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe frekans poligonu denir.

sınıfın orta noktasıdır. Histogramların kapladığı alan ile poligonun altında kalan alan birbirine eşittir [3]. Aşağıdaki örnekte 100 öğrencinin istatistik dersinden aldıkları notlar kullanılarak elde edilen histogram ile frekans poligonu Görsel 4.2’de gösterilmiştir.



Örnek

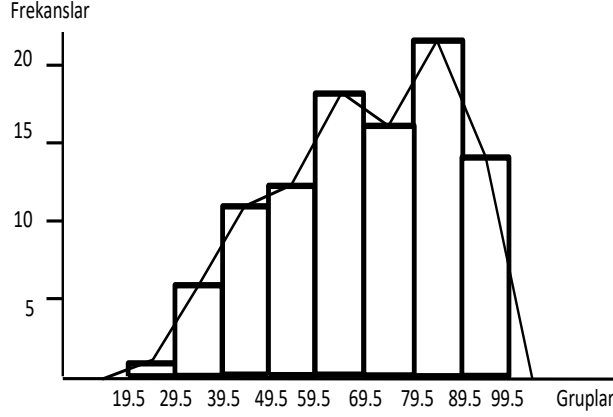
- Örnek 4.2. İktisadi ve İdari Bilimler Fakültesi’nde okuyan 100 öğrencinin istatistik dersinden aldığı notları sorularak aşağıdaki verilere ulaşılmıştır. Öğrencilerin istatistik dersinden aldıkları notları kullanarak histogram ve frekans poligonunu çiziniz?

Notlar	Öğrenci Sayısı
20-29	1
30-39	6
40-49	11
50-59	12
60-69	18
70-79	16
80-89	22
90-99	14

Burada verilen seri kesikli gruplandırılmış seri olduğu için öncelikle serinin sınıf sınırları ve sınıf değerleri elde edilmiş ve aşağıda verilmiştir:

Gruplar	Alt Sınır	Üst Sınır	Sınıf Değeri
20 – 29	19.5	29.5	24.5
30 – 39	29.5	39.5	34.5
40 – 49	39.5	49.5	44.5
50 – 59	49.5	59.5	54.5
60 – 69	59.5	69.5	64.5
70 – 79	69.5	79.5	74.5
80 – 89	79.5	89.5	84.5
90 – 99	89.5	99.5	94.5

Sınıf sınırları ve sınıf değerleri kullanılarak serinin histogramı ve frekans poligonu aşağıdaki gibi elde edilmiştir:



Görsel 4.2. İstatistik Notlarının Histogram ve Frekans Poligonu

KÜMÜLATİF FREKANS GRAFİKLERİ

“...den az kümülatif frekans dağılımları” ve “...den çok kümülatif frekans dağılımları”nın grafikleri koordinat sistemi üzerinde çizilebilir. Grafiklerin çiziminde yatay eksen de değişken değerleri, dikey eksen de ise kümülatif frekans değerleri bulunur. Değişkenin aldığı değerler ile kümülatif frekansların kesiştiği noktaların birleştirilmesi ile kümülatif frekans dağılımlarının grafiği çizilmiş olur. Nispi ve kümülatif frekans dağılımlarına ait grafikleri elde etmek için dikey eksene nispi veya kümülatif frekans değerleri yerleştirilmelidir. Kümülatif frekans poligonlarına ojiv eğrileri de denir [2].



Kümülatif frekans poligonlarına ojiv eğrileri de denir.

Gruplandırılmış serilerin ...den az ve ...den çok kümülatif frekans dağılımlarına ait grafiklerin çiziminde yatay eksen de sınıf sınırları işaretlenir. “...den az kümülatif frekans dağılımı” için üst sınıf sınırları apsisi, frekans değerleri ordinat olacak şekilde düzlemdeki noktalar belirlenir. Belirlenen noktalar çizgilerle birleştirilir. “...den çok kümülatif frekans dağılımları”na ait grafiklerin çiziminde ise gruplandırılmış seride alt sınıf sınırları kullanılmaktadır. Alt sınıf sınırları apsisi, frekans değerleri ordinat olacak şekilde düzlemdeki noktalar çizgilerle birleştirilir. Gruplandırılmış serideki son sınıfın üst sınırında son nokta bulunur ve burada grafik yatay eksene değecek şekilde çizim yapılır. Kısaca gruplandırılmış serilerin “...den az kümülatif frekans dağılımları”na ait grafikler çizilirken üst sınıf sınırları “...den çok kümülatif frekans dağılımları”na ait grafikler çizilirken ise alt sınıf sınırları kullanılmaktadır [4]. Aşağıdaki Örnek 4.2’de verilen öğrencilerin not dağılımları kullanarak oluşturulan “...den az ve ...den çok kümülatif frekans dağılımları”nın grafikleri sırasıyla Görsel 4.3 ve Görsel 4.4’te gösterilmiştir.



Örnek

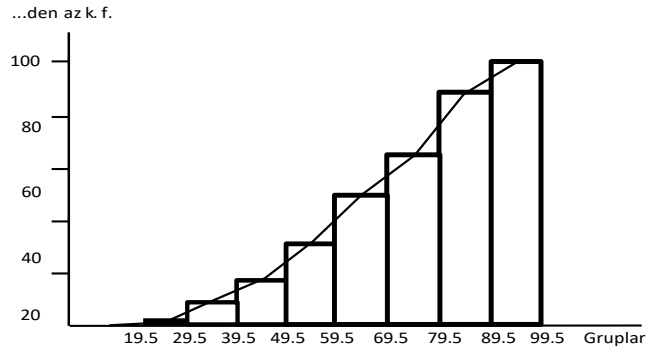
- Örnek 4.3. Örnek 4.2’de verilen öğrencilerin not dağılımını kullanarak, ...den az ve ...den çok kümülatif frekans dağılımlarının grafiklerini çizin.

Notlar	Öğrenci Sayısı
20-29	1
30-39	6
40-49	11
50-59	12
60-69	18
70-79	16
80-89	22
90-99	14

“...den az kümülatif frekans dağılımı”nın grafiğini çizmek için öncelikle “...den az kümülatif frekans değerleri” elde edilir:

Gruplar	Frekanslar	...den az k. f.
20-29	1	1
30-39	6	7
40-49	11	18
50-59	12	30
60-69	18	48
70-79	16	64
80-89	22	86
90-99	14	100

Bu dağılıma göre hazırlanan “...den az kümülatif frekans dağılımı”nın grafiği aşağıdaki gibidir:

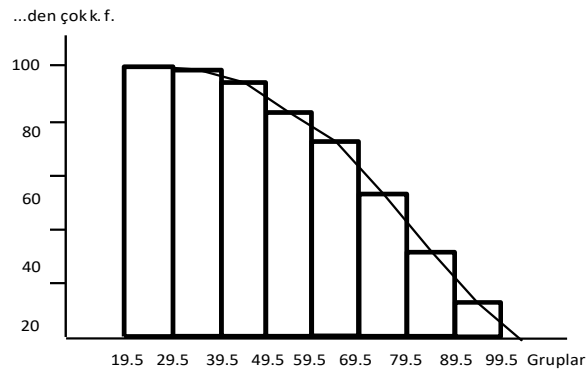


Görsel 4.3. “...den Az Kümülatif Frekans Dağılımı”nın Grafiği

“...den çok kümülatif frekans dağılımı”nın grafiğini çizmek için öncelikle “...den çok kümülatif frekans değerleri” elde edilerek aşağıda verilmiştir:

Gruplar	Frekanslar	...den çok k.f.
20-29	1	100
30-39	6	99
40-49	11	93
50-59	12	82
60-69	18	70
70-79	16	52
80-89	22	36
90-99	14	14

Bu dağılıma göre hazırlanan “...den çok kümülatif frekans dağılımı”nın grafiği aşağıdaki gibidir:



Görsel 4.4. “...den Çok Kümülatif Frekans Dağılımının” Grafiği

...den az ve ...den çok kümülatif frekans dağılımları aynı grafik üzerinde gösterilebilir. Grafikteki iki kümülatif frekans eğrisinin kesiştiği noktadaki X değeri medyan değerini vermektedir.



Örnek

- Örnek 4.4. Bir pil fabrikasında üretilen piller arasından tesadüfi olarak seçilen 400 pilin dayanma süreleri aşağıdaki gibi tespit edilmiştir. ...den az ve ...den çok kümülatif frekans dağılımlarına ait ojiv eğrilerini aynı grafik üzerinde gösteriniz.

Pillerin Dayanma Süresi	Pil Sayısı
300 – 399	14
400 – 499	46
500 – 599	58
600 – 699	76
700 – 799	68
800 – 899	62
900 – 999	48
1000 – 1099	22
1100 – 1199	6

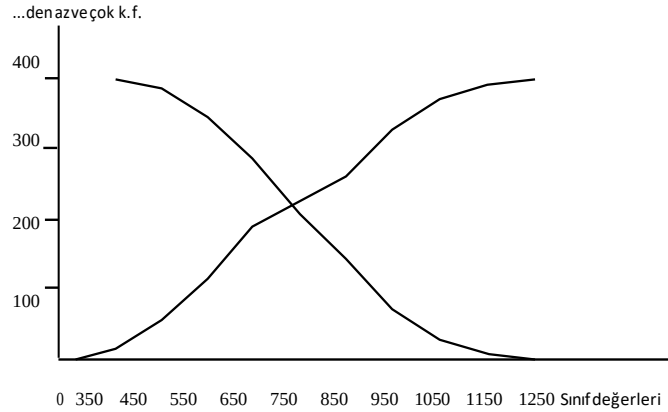
Burada seri kesikli gruplandırılmış seri olduğu için öncelikle serinin sınıf sınırları ve sınıf değerleri elde edilmiş ve aşağıda verilmiştir:

Dayanma Süresi	Pil Sayısı	Alt Sınır	Üst Sınır	Sınıf Değeri
300 – 399	14	299.5	399.5	349.5
400 – 499	46	399.5	499.5	449.5
500 – 599	58	499.5	599.5	549.5
600 – 699	76	599.5	699.5	649.5
700 – 799	68	699.5	799.5	749.5
800 – 899	62	799.5	899.5	849.5
900 – 999	48	899.5	999.5	949.5
1000 – 1099	22	999.5	1099.5	1049.5
1100 – 1199	6	1099.5	1199.5	1149.5

“...den az kümülatif frekans dağılımları” ve “...den çok kümülatif frekans dağılımları”na ait ojiv eğrilerini aynı grafik üzerinde çizmek için öncelikle kümülatif frekans değerleri elde edilerek aşağıda verilmiştir:

Dayanma Süresi	Sınıf Değeri	Pil Sayısı	..den az k.f.	..den çok k.f.
299.5 – 399.5	349.5	14	14	400
399.5 – 499.5	449.5	46	60	386
499.5 – 599.5	549.5	58	118	340
599.5 – 699.5	649.5	76	194	282
699.5 – 799.5	749.5	68	262	206
799.5 – 899.5	849.5	62	324	138
899.5 – 999.5	949.5	48	372	76
999.5 – 1099.5	1049.5	22	394	28
1099.5 – 1199.5	1149.5	6	400	6

Frekans dağılımındaki sınıf değerlerine karşı gelen ...den az ve ...den çok kümülatif frekans değerleri koordinat sisteminde işaretlendikten sonra bu noktalar birleştirilerek ojiv eğrileri aşağıdaki gibi elde edilir. [2]



Görsel 4.5. “...den Az ve ...den Çok Frekans Dağılımlarının” Ojiv Eğrileri

SÜTUN GRAFİĞİ

Nitel veya başka bir ifade ile kategorik veriler değişkenin vasfı ile ilgili sayısal olmayan verilerdir. Özellikle sosyal bilimlerde araştırma konuları gereği daha fazla nitel veriler kullanılmaktadır. Nitel veride birimlere ait sıfatlar ya da durumlar



Sütun grafiği; yatay ekseninde kategorilerin, dikey ekseninde ise bu kategorilerin frekansları gösterilerek elde edilir.

tespit edilir. Nitel verilerin sunumu grafiklerle yapılabilir. Örneğin; ülkelerin nüfus dağılımı, televizyon kanallarının izlenme oranları, siyasi partilerin seçimlerde aldıkları oy oranları, meslek gruplarına göre dağılım grafiklerle gösterilebilir. Bu tip verilerin görsel olarak sunulması için çeşitli grafikler mevcuttur. Bu grafik çeşitlerinden biri de sütun grafiğidir.

Sütun grafiği oluşturulurken, yatay ekseninde kategoriler, dikey ekseninde ise bu kategorilerin frekansları gösterilerek oluşturulur. Çizilen sütunların genişlikleri birbirine eşittir. Sütunlar birbirinden ayrı olarak çizilir. [5]

Aşağıdaki örnekte 100 kişiye ait meslek grupları verilmiş ve bu dağılım kullanılarak oluşturulan sütun grafiği Görsel 4.6'da gösterilmiştir.

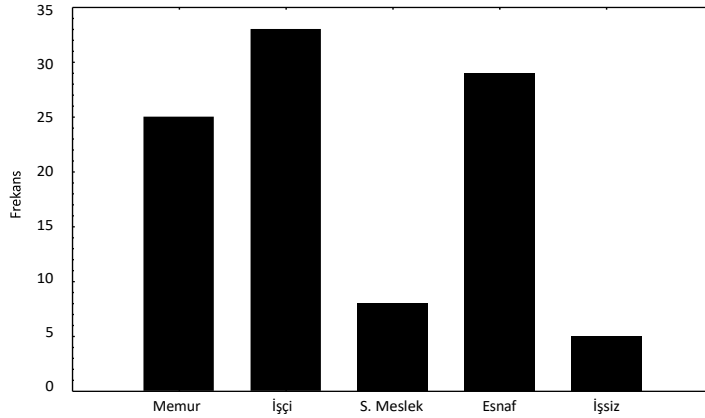


Örnek

- Örnek 4.5. Bir ilde yaşayan 100 şahsın meslekleri aşağıdaki gibi tespit edilmiştir. Bu verileri kullanarak sütun grafiğini çiziniz..

Meslekler	Frekanslar
Memur	25
İşçi	33
Serbest Meslek	8
Esnaf	29
İşsiz	5

Yatay ekseninde meslek grupları dikey ekseninde ise bu meslek gruplarının frekansları gösterilerek sütun grafiği elde edilmiş ve aşağıda verilmiştir:



Görsel 4.6. Meslek Gruplarına Ait Sütun Grafiği

Görsel 4.6 da yer alan sütun incelendiğinde; ilk bakışta en fazla sayıda kişinin işçi olduğu, ikinci sırada ise esnafın olduğu ve en az sayıda kişinin ise işsiz olduğu rahatlıkla görülebilir [2].



Daire grafiği, sütun grafiği gibi kategorik verilerin sunumunda kullanılan bir grafik türüdür.



Daire grafiği, şekil itibarıyla dilimlenmiş pastaya benzediği için bu grafiğe pasta grafiği de denilmektedir.

DAİRE GRAFİĞİ

Daire grafiği yukarıda izah edilen sütun grafiği gibi kategorik verilerin sunumunda kullanılan bir grafik türüdür. Daire grafiği, nominal ölçekle elde edilmiş veriler, kategorik veriler veya az sayıda sınıfa ayrılabilen veriler için kullanılabilecek bir grafik türüdür. Daire grafiğinde her sınıf veya kategori sahip olduğu frekansla orantılı büyüklükteki dilimlerle gösterilmektedir. Şekil itibarıyla dilimlenmiş pastaya benzediği için bu grafiğe pasta grafiği de denilmektedir [4].

Daire grafiğinin çizilmesi için öncelikle nispi frekans değerleri elde edilmektedir. Bir dairenin iç açıları toplamı 360° olduğuna göre nispi frekanslar 360 ile çarpılarak her bir kategorinin 360° den alacağı pay tespit edilir. Toplam veri sayısı 360° ’lik açıya karşılık gelmektedir. Genellikle saatin 12 ’yi gösterdiği noktadan başlanarak her bir kategorinin belirlediği alanın bulunması için, saatin dönüş istikametine göre, kategoriler için belirlenen açılar kadar ilerlenir [2].

Yukarıda ifade edildiği gibi daire grafiği, az sayıda sınıfa veya kategoriye ayrılabilen veriler için kullanılabilecek bir grafik türüdür. Sınıf veya kategori sayısı arttıkça daire grafiğinin çizimi ve anlaşılması zorlaşır. Grafiklerde amaç, verilerin anlaşılır bir şekilde görsel olarak sunulması olduğuna göre grafik karmaşıklıkla amacından uzaklaşmış olacaktır. Dolayısıyla kategorik veriler için grafik çizilirken bu hususun dikkate alınması gerekmektedir [1].

Aşağıdaki örnekte bir ülkedeki çalışma kategorilerine göre çalışanların sayısı verilmiş ve bu veriler kullanılarak oluşturulan daire grafiği Görsel 4.7’de gösterilmiştir.



Örnek

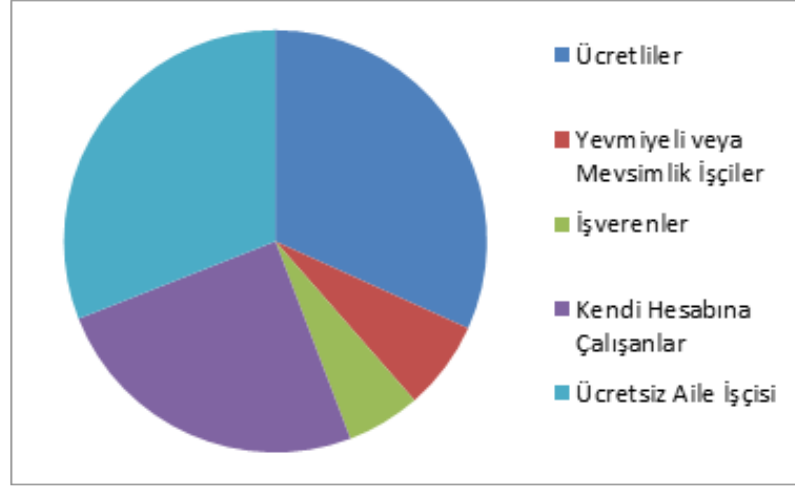
- Örnek 4.6. Bir ülkedeki çalışma kategorilerine göre çalışanların sayısı aşağıdaki gibidir. Söz konusu verileri kullanarak daire grafiğini çizin.

Çalışma Kategorileri	Çalışan Sayısı
Ücretliler	6283355
Yevmiyeli veya Mevsimlik İşçiler	1347728
İşverenler	1118950
Kendi Hesabına Çalışanlar	4905037
Ücretsiz Aile İşçileri	6129021

Daire grafiğinin çizilmesi için ilk olarak nispi frekans değerleri elde edilmiştir. Daha sonra nispi frekans değerleri 360 ile çarpılarak her bir kategorinin 360° den alacağı paylar tespit edilmiştir. Elde edilen değerler aşağıda gösterilmiştir:

Çalışma Kategorileri	Frekanslar	n.f.	360° x n.f
Ücretliler	6283355	0.3176	114.336
Yevmiyeli veya Mevsimlik İşçiler	1347728	0.0681	24.516
İşverenler	1118950	0.0566	20.376
Kendi Hesabına Çalışanlar	4905037	0.2479	89.244
Ücretsiz Aile İşçileri	6129021	0.3098	111.528
Toplam	19784091		360

Gerekli hesaplamalar yapıldıktan sonra saatin 12'yi gösterdiği noktadan başlanarak her bir kategorinin belirlediği alanın bulunması için saatin dönüş istikametine göre, kategoriler için belirlenen açılar kadar ilerlenir. Böylece daire grafiği aşağıdaki gibi elde edilir:



Görsel 4.7. Çalışma Kategorilerine Ait Daire Grafiği

Grafik incelendiğinde; en fazla payı %31.76 ile ücretliler alırken onu %30.98 ile ücretsiz aile işçileri ve %24.79 ile kendi hesabına çalışanlar takip etmektedir. En az payı ise %5.66 ile işverenler almaktadır [2].

PARETO GRAFİĞİ



Pareto grafiği, hata ve maliyet analizleri için kullanılan basit bir yöntemdir.

İlk kez İtalyan Ekonomist Pareto (1848-1923) tarafından bulunan Pareto grafiği, söz konusu bilim adamının adıyla anılmaktadır. *Pareto grafiği gelir dağılımları ile ilgili çalışmalarda tespit edilmiştir. Grafik hata ve maliyet analizleri için kullanılan basit bir yöntemdir.*

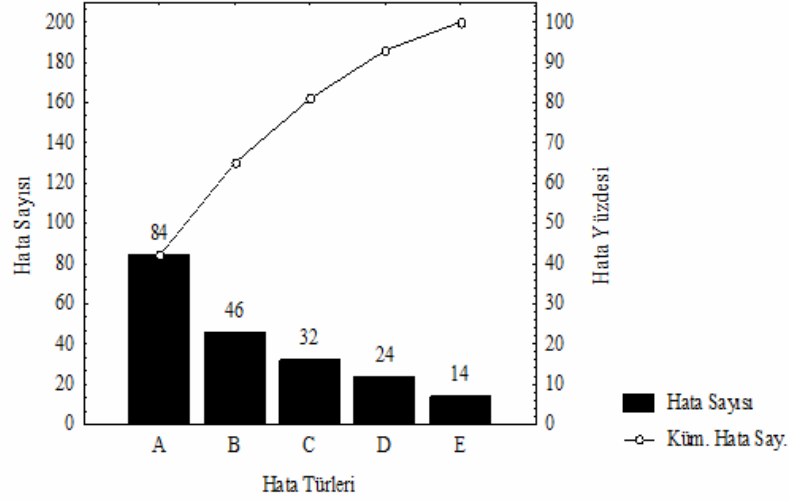
Pareto grafiği; bozuk ürünler, tamirler, arızalar, talepler, noksanlıklar veya kazalar ile mali kayıplar ve bunların sebepleri gibi olayların görsel olarak meydana gelme frekanslarını gösteren bir tür frekans dağılım grafiğidir.

Pareto grafiği bir sorunu oluşturan nedenleri önem sırasına göre sıralayarak, önemlileri önemsizlerden ayırt etmeye ve dikkatleri önemli nedenler üzerinde toplamaya imkân vermektedir. Grafikte soldan sağa doğru gidildikçe küçülen sütunlara sahip veri sınıfları sıralanmaktadır [6].

Pareto grafiğinde en önemli problemler sol tarafta yer alırken, önemsiz olanlar sağ tarafta yer alır. Bazen “Diğerleri” adı altında çok önemsiz durumlar bir

sınıf altında toplanıp birleştirilebilir. “Diğerleri” kategorisi kullanıldığında bu sınıf en sağda yer alır. Maliyet, frekanslar veya % gibi değerler de dikey eksen üzerinde gösterilir.

Bazı durumlarda pareto grafiğinde kümülatif eğri gösterilmektedir. Söz konusu eğri soldan sağa doğru eklenen veri sınıflarının toplamını göstermektedir [4]. Görsel 4.8.’de örnek bir pareto grafiği verilmiştir.



Görsel 4.8. Örnek Bir Pareto Grafiği

Görsel 4.8’ deki grafik incelendiğinde en fazla A türü hatanın yapıldığı onu da B türü hatanın takip ettiği görülmektedir. En az hata türünün ise E türü hata olduğu pareto grafiğinden kolaylıkla anlaşılmaktadır.

SERPİLME DİYAGRAMI



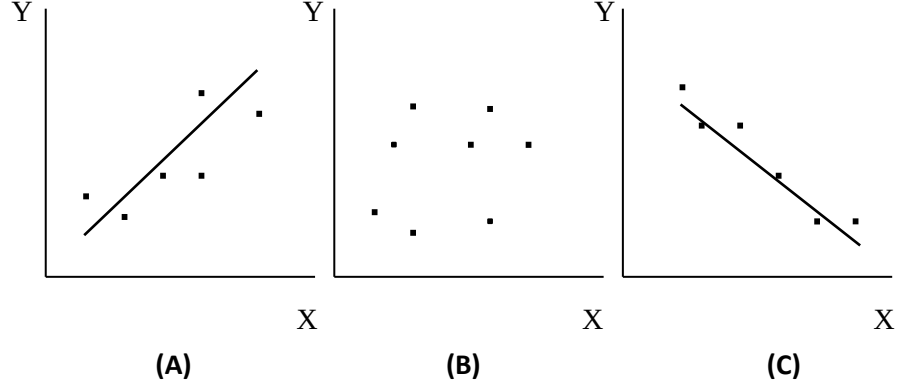
Değişkenler arasındaki ilişkilerin grafik aracılığı ile incelenmesinde serpilme diyagramından yararlanılmaktadır.

Bilimin en temel amaçlarından birisi, değişkenler arasındaki ilişkileri incelemektir. Değişkenler arasındaki ilişkilerin grafik aracılığı ile incelenmesinde serpilme diyagramından yararlanılmaktadır. Serpilme diyagramlarında yatay eksen genellikle neden durumunda olan değişken (X) ve dikey eksen de sonuç durumunda olan değişken (Y) kullanılır. Neden durumunda olan değişkene (X) bağımsız değişken ve sonuç durumunda olan değişkene (Y) de bağımlı değişken adı verilmektedir. Bu iki değişkene ait sıralı ikili çiftleri olan (X, Y) değerler kartezyen koordinat sistemi üzerine yerleştirilir [7].

Boy ile kilo, fiyat ile talep, arz ile talep, gelir ile tasarruf, gelir ile harcama, çalışma süresi ile başarı notu gibi ikili değişkenler arasındaki ilişkileri incelemek için serpilme diyagramı kullanılabilir.

X, Y değerlerinin oluşturduğu noktalara bakılarak ilişkinin yönü ve derecesi göz kararıyla tahmin edilebilir. Bunun için en dışta kalan noktalar birleştirilerek bir şekil elde edilir. Söz konusu şeklin doğrultusuna bakılarak ilişkinin yönü saptandığı gibi, şeklin durumuna göre ilişkinin derecesi hakkında tahminde bulunulur. Şayet şekil oldukça dar bir elipse benziyorsa ilişkinin kuvvetli olduğu ifade edilir. Elips

genişledikçe ilişkinin derecesi zayıflar. Değişkenler arasında ilişki bulunmadığında ise noktaların oluşturduğu şekil bir daireye benzer [4]. Görsel 4.9.'da üç farklı durum için serpilme diyagramı örnekleri verilmiştir.



Görsel 4.9. Serpilme Diyagramı Örnekleri

Görsel 4.9 incelendiğinde; A ve C grafiklerinde şekil elipse benzediği için X ve Y değişkenleri arasında ilişki olduğuna karar verilir. A grafiği X ve Y değişkenleri arasında pozitif ve kuvvetli bir ilişkinin olabileceğini işaret ederken, C grafiği X ve Y değişkenleri arasında ters yönlü kuvvetli bir ilişkinin olabileceğini işaret etmektedir. B grafiğinde ise şekil daireye benzediği için X ve Y değişkenleri arasında ilişki olmadığına karar verilir [2].



Bireysel Etkinlik

- Hangi veri için hangi grafik türünün daha uygun olacağını tartışınız.
- Grafik türleri için birer örnek veriniz.
- Seçim ve referandum sonuçları hangi grafik türü ile gösterilir?



Özet

- Bilindiği üzere istatistiğin uğraş alanlarından biri de özetleme ve görselleştirmedir. Frekans dağılımlarının bir tablo şeklinde sunulması bir özetleme şeklidir. Oysa tablolar her zaman kolay anlaşılır olmayabilir. Grafikler gözlem değerlerinin görselleştirilmesinde kullanılır. Grafik, “gözlem değerlerinin matematiksel ve bilimsel temellere sahip şekiller hâlinde ifade edilmesi” şeklinde tanımlanabilir.
- Grafikler gözlem değerlerinin daha kolay anlaşılmasını sağlamaktadır. Çünkü grafikler temsil ettikleri olayların bileşimini ve değişimindeki ana eğilimi tüm canlılığı ile ilk bakışta ortaya koymaktadırlar. Bu nedenle rakamlarla uğraşmak istemeyen veya rakamlardan anlam çıkarmakta zorluk çekenler için grafikler daha fazla yardımcı olmaktadır.
- Histogram, gruplandırılmış seriler için oluşturulan bir dikkörtgenler dizisidir. Histogram; yatay ekseninde değişkenin aldığı değerlerin, dikey ekseninde ise frekansların bulunduğu ve her aralığın frekansı ile orantılı yüksekliklere sahip dikkörtgenlerin gösterildiği bir yoğunluk grafiğidir. Bu dikkörtgenlerin tabanları gruplandırılmış serideki her bir sınıfın sınıf büyüklüğünü, yükseklikleri ise sınıf frekansını göstermektedir.
- Frekans poligonu, histogramdaki sınıf sınırlarının orta noktalarını apsis, sınıf frekanslarını ordinat kabul eden noktaların doğru parçaları ile birleştirilmesi ile elde edilen grafik çeşididir. Kısacası histogramı oluşturan dikkörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe *frekans poligonu* denir. Frekans poligonu, sınıf sınırlarının mümkün olduğunca küçültülmesi durumunda bir eğriye dönüşür. Söz konusu eğriye ise *frekans eğrisi* adı verilir.
- Frekans poligonunun yatay eksen üzerindeki başlangıç noktası ilk sınıftan bir önceki farazi sınıfın orta noktası iken bitiş noktası ise son sınıftan bir sonraki farazi sınıfın orta noktasıdır. Histogramların kapladığı alan ile poligonun altında kalan alan birbirine eşittir.
- “...den az kümülatif frekans dağılımları” ve “...den çok kümülatif frekans dağılımları”nın grafikleri koordinat sistemi üzerinde çizilebilir. Grafiklerin çiziminde yatay ekseninde değişken değerleri, dikey ekseninde ise kümülatif frekans değerleri bulunur. Değişkenin aldığı değerler ile kümülatif frekansların kesiştiği noktaların birleştirilmesi ile kümülatif frekans dağılımlarının grafiği çizilmiş olur. Nispi ve kümülatif frekans dağılımlarına ait grafikleri elde etmek için dikey eksene nispi veya kümülatif frekans değerleri yerleştirilmelidir. Kümülatif frekans poligonlarına ojiv eğrileri de denir.
- Nitel verilerin sunumu grafiklerle yapılabilir. Örneğin; ülkelerin nüfus dağılımı, televizyon kanallarının prime-time izlenme oranları, siyasi partilerin seçimlerde aldıkları oy oranları, meslek gruplarına göre dağılım grafiklerle gösterilebilir. Bu tip verilerin görsel olarak sunulması için çeşitli grafikler mevcuttur. Bu grafik çeşitlerinden biri de sütun grafiğidir.
- Sütun grafiği, yatay ekseninde kategorilerin, dikey ekseninde ise bu kategorilerin frekansları gösterilerek elde edilir. Çizilen sütunların genişlikleri birbirine eşittir. Çizilen sütunlar birbirine bitişik olabileceği gibi birbirinden ayrıık da olabilir.
- Daire grafiği, nominal ölçekle elde edilmiş veriler, kategorik veriler veya az sayıda sınıfa ayrılabilen veriler için kullanılabilecek bir grafik türüdür. Daire grafiğinde her sınıf veya kategori sahip olduğu frekansla orantılı büyüklükteki dilimlerle gösterilmektedir. Şekil itibarıyla dilimlenmiş pastaya benzediği için bu grafiğe pasta grafiği de denilmektedir.
- Daire grafiği az sayıda sınıfa veya kategoriye ayrılabilen veriler için kullanılabilecek bir grafik türüdür. Sınıf veya kategori sayısı arttıkça daire grafiğinin çizimi ve anlaşılması zorlaşır.



Özet

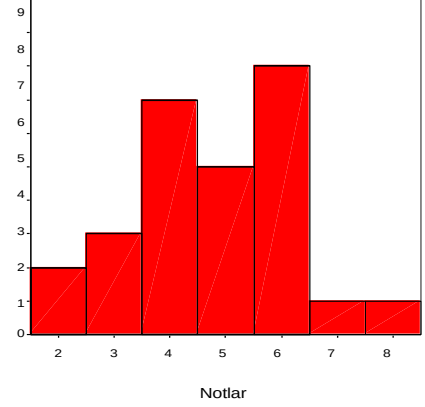
- İlk kez İtalyan Ekonomist Pareto (1848-1923) tarafından bulunan Pareto grafiği, söz konusu bilim adamının adıyla anılmaktadır. Pareto grafiği gelir dağılımları ile ilgili çalışmalarda tespit edilmiştir. Grafik hata ve maliyet analizleri için kullanılan basit bir yöntemdir.
- Pareto grafiği; bozuk ürünler, tamirler, arızalar, talepler, noksanlıklar veya kazalar ile mali kayıplar ve bunların sebepleri gibi, olayların görsel olarak meydana gelme frekanslarını gösteren bir tür frekans dağılım grafiğidir.
- Pareto grafiği bir sorunu oluşturan nedenleri önem sırasına göre sıralayarak, önemlileri önemsizlerden ayırt etmeye ve dikkatleri önemli nedenler üzerinde toplamaya imkân vermektedir. Grafikte soldan sağa doğru gidildikçe küçülen sütunlara sahip veri sınıfları sıralanmaktadır.
- Değişkenler arasındaki ilişkilerin grafik aracılığı ile incelenmesinde serpilme diyagramından yararlanılmaktadır. Serpilme diyagramlarında yatay eksen genellikle neden durumunda olan değişken (X) ve dikey eksen de sonuç durumunda olan değişken (Y) kullanılır.
- Serpilme diyagramında neden durumunda olan değişkene (X) bağımsız değişken ve sonuç durumunda olan değişkene (Y) de bağımlı değişken adı verilmektedir. Bu iki değişkene ait sıralı ikili çiftleri olan (X, Y) değerler kartezyen koordinat sistemi üzerine yerleştirilir.

DEĞERLENDİRME SORULARI

1. Gruplandırılmış serilerin kümülatif frekans dağılımlarına ait grafiklerin çiziminde yatay ekseninde işaretlenir.
Yukarıdaki boşluğa aşağıdakilerden hangisi getirilmelidir?
 - a) Frekanslar
 - b) ...den az frekanslar
 - c) ...den çok frekanslar
 - d) Kümülatif frekanslar
 - e) Sınıf sınırları
2. Gruplandırılmış seriler için oluşturulan dikdörtgenler dizisine ne ad verilir?
 - a) Ojiv eğrisi
 - b) Histogram
 - c) Pareto grafiği
 - d) Frekans poligonu
 - e) Sütun grafiği
3. Histogram ile ilgili aşağıdaki ifadelerden hangisi yanlıştır?
 - a) Histogram bir dikdörtgenler dizisinden oluşmaktadır.
 - b) Yatay ekseninde gruplandırılmış serinin sınıf sınırları yer alır.
 - c) Dikey ekseninde frekanslar yer alır.
 - d) Dikdörtgenlerin tabanları her bir sınıfın sınıf büyüklüğünü göstermektedir.
 - e) Histogramda dikdörtgenler büyüklük sırasına göre yerleştirilir.
4. Histogramı oluşturan dikdörtgenlerin üst kenarlarının orta noktalarını birleştirmek suretiyle elde edilen grafiğe ne ad verilir?
 - a) Frekans poligonu
 - b) Pareto grafiği
 - c) Sütun grafiği
 - d) Daire grafiği
 - e) Kümülatif frekans eğrisi
5. Kümülatif frekans poligonlarına ne ad verilir?
 - a) Pareto grafiği
 - b) Frekans poligonu
 - c) Ojiv eğrileri
 - d) Kümülatif frekans eğrisi
 - e) Frekans dağılımı

6. Yukarıdaki histogramda gösterilen dağılımın gözlem sayısı (n) kaçtır?

- a) 7
- b) 8
- c) 26
- d) 27
- e) 112



7. Grafik türlerinden hangisi kategorik veriler için kullanılan bir grafik türüdür?

- a) Histogram
- b) Daire grafiği
- c) ...den az kümülatif frekans eğrisi
- d) Ojiv eğrisi
- e) Kümülatif frekans eğrisi

8. Aşağıdakilerden hangisi sürekli sayısal veriler için kullanılan bir grafik türüdür?

- a) Sütun grafiği
- b) Daire grafiği
- c) Pareto grafiği
- d) Frekans dağılımı
- e) Histogram

9. Bozuk ürünler, tamirler, arızalar gibi olayların görsel olarak meydana gelme frekanslarını gösteren grafik türü aşağıdakilerden hangisidir?

- a) Sütun grafiği
- b) Daire grafiği
- c) Ojiv eğrisi
- d) Pareto grafiği
- e) Histogram

10. Grafiklerle ilgili aşağıda verilen ifadelerden hangisi yanlıştır?

- a) Histogram bir dikkörtgenler dizisinden oluşmaktadır.
- b) Sütun grafiğinde çizilen sütunların genişlikleri birbirinden farklıdır.
- c) Daire grafiği kategorik verilerin sunumunda kullanılan bir grafik türüdür.
- d) Pareto grafiği, hata ve maliyet analizleri için kullanılan basit bir yöntemdir.
- e) Grafikler, gözlem değerlerinin daha kolay anlaşılmasını sağlamaktadır.

Cevap Anahtarı

1.e, 2.b, 3.e, 4.a, 5.c, 6.d, 7.b, 8.e, 9.d, 10.b

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İsmail Hakkı. (2018). İşletme İstatistiğine Giriş (3.baskı). İstanbul: Alfa Yayınları.
- [2] Başar, Alaaddin, Oktay, Erkan. (2018). Uygulamalı İstatistik – I: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap ve Kirtasiye.
- [3] Berenson, M. L., D. M. Levine (1996). Basic Business Statistics: Concepts and Applications (8. baskı). New Jersey: Prentice Hall Inc.
- [4] Serper, Özer. (2014). Uygulamalı İstatistik I, II (7. baskı), İstanbul: Filiz Kitabevi.
- [5] Turanlı, M., S. Güriş (2000). Temel İstatistik. İstanbul: Der Yayınları.
- [6] Köksal, B. A. (1995). İstatistik Analiz Metodları (4. baskı). İstanbul: Çağlayan Kitabevi.
- [7] Gürsakal, Necmi, Oğuzlar, Ayşe, Gürsakal, Sevdâ. (2022). Betimsel İstatistik (10. Baskı). Bursa: Dora Yayınları.

PARAMETRİK MERKEZÎ EĞİLİM ÖLÇÜLERİ



İÇİNDEKİLER

- Aritmetik Ortalama
- Geometrik Ortalama
- Harmonik Ortalama
- Kareli Ortalama



HEDEFLER

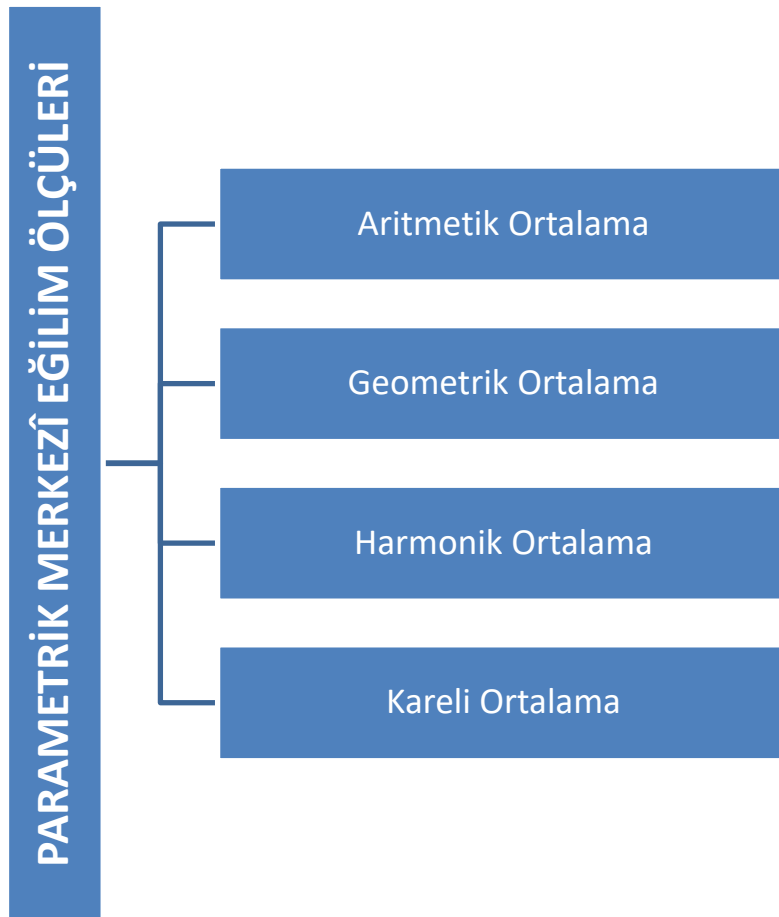
- Bu üniteyi çalıştıktan sonra;
 - Parametrik merkezî eğilim ölçülerini açıklayabilecek,
 - Basit, sınıflandırılmış ve gruplandırılmış serilerden parametrik merkezî eğilim ölçülerini hesaplayabilecek,
 - Parametrik merkezi eğilim ölçüleri arasındaki büyüklük ilişkisi hakkında bilgi sahibi olabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Dr. Emrah TALAŞ

ÜNİTE
5



GİRİŞ

Merkezî eğilim ölçüleri, istatistiğin özetleme görevini en ileri seviyede gören istatistik ölçülerdir. Şöyle ki, 1000 birimlik bir seri, mesela 80 sınıf yahut 15 grup hâlinde özetlenebildiği gibi serinin ortalaması alınmak suretiyle bu 1000 sayı bir tek birimle temsil edilebilmektedir.



Serideki bütün birimlerden etkilenen merkezî eğilim ölçülerine parametrik merkezî eğilim ölçüleri denir.

Merkezî eğilim ölçüleri başlıca iki gruba ayrılır:

- Serinin bütün birimlerine tabi olan merkezî eğilim ölçüleri
- Serinin bütün birimlerine tabi olmayan merkezî eğilim ölçüleri

Serinin bütün birimlerine tabi olan merkezî eğilim ölçülerine *parametrik merkezî eğilim ölçüleri* de denilmektedir. Bu gruptaki merkezî eğilim ölçüleri serideki tek bir rakamın değişmesinden doğrudan doğruya etkilenirler. Bu sebeple parametrik merkezî eğilim ölçülerinin tamamı serideki aşırı uçların etkisinde kalırlar. Sınıf uçları belli olmayan gruplandırılmış serilerde sınıf değerleri hesaplanamayacağı için parametrik merkezî eğilim ölçülerinin hiçbirisi hesaplanamaz. Bu gruptaki merkezî eğilim ölçüleri; aritmetik ortalama ($\bar{X}$), geometrik ortalama (G), harmonik ortalama (H) ve kareli ortalamadır (K).

İkinci grupta incelenecek merkezî eğilim ölçüleri ise serideki her bir değerden direkt olarak etkilenmeyebilir. Medyan, mod, kartiller ve ortalama kartil değerleri bu gruba giren merkezî eğilim ölçülerindendir. Bu ölçüler takip eden üniteye ele alınmıştır.

ARİTMETİK ORTALAMA ($\bar{X}$)

Bu grupta yer alan ortalamalar içinde en basit ve en yaygın olarak kullanılan ortalama aritmetik ortalamadır ve serideki tüm değerlerin toplamının toplam eleman sayısına bölünmesi ile hesaplanır. Frekans ve sınıflı serilerde aritmetik ortalama ise serinin toplam değerinin toplam frekansa (birim sayısına) bölünmesi ile bulunur. Ortalama seride değişkeni ifade eden harfin üzerine çizgi çizilerek gösterilir. [1] Aritmetik ortalama serideki bütün rakamlardan etkilenen bir ortalamadır. İstatistiki uygulamalarda en çok kullanılan merkezî eğilim ölçüsüdür. Basit bir seride aritmetik ortalama serideki birimlerin toplamının birim sayısına bölümüyle elde edilir. Yani,

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

şeklinde hesaplanır. Sınıflandırılmış ve gruplandırılmış serilerde ise aritmetik ortalama,

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{\sum_{i=1}^n f_i}$$



Basit bir seride, serideki birimlerin toplamının birim sayısına bölünmesiyle aritmetik ortalama hesaplanır.

formülüyle hesaplanır. Gruplandırılmış serilerde X değerleri sınıf orta noktalarını yani sınıf değerlerini gösterirler. X değerlerine ait tartılar varsa frekanslar yerine tartılar kullanılarak tartılı aritmetik ortalama hesaplanır. Tartılar t ile gösterilerek, tartılı aritmetik ortalama formülü,

$$\bar{X} = \frac{\sum_{i=1}^n t_i X_i}{\sum_{i=1}^n t_i}$$

şeklinde dir. Tartılı aritmetik ortalama hesaplanırken kullanılan tartılar sınıflandırılmış ve gruplandırılmış serideki frekanslar gibi işlem görür [2].



Örnek

- 10, 12, 18 ve 20 sayılarından oluşan basit serinin aritmetik ortalamasını hesaplayalım.

Basit serilerde aritmetik ortalama hesaplanırken önce serideki değerler toplanır. Örnekte yer alan serideki değerlerin toplamı $\sum_{i=1}^4 X_i = 60$ 'tır. Daha sonra bu toplam serideki değer sayısı olan 4'e bölünür. Böylece serinin aritmetik ortalaması

$$\bar{X} = \frac{\sum_{i=1}^4 X_i}{n} = \frac{60}{4} = 15$$

olarak elde edilir.



Aritmetik ortalama, aritmetik dizi şeklinde artış veya azalış gösteren serileri en iyi temsil eden parametrik merkezî eğilim ölçüsüdür.



Örnek

- Aşağıdaki sınıflandırılmış serinin aritmetik ortalamasını hesaplayalım.

X_i	f_i
4	1
5	2
7	4
9	3

Sınıflandırılmış serilerin aritmetik ortalaması hesaplanırken önce her bir X_i değişken değerinin tekrarlanma sayısını ifade eden f_i değeri ile çarpılır. Çarpım sonuçları aşağıdaki gibidir:

X_i	f_i	$f_i X_i$
4	1	4
5	2	10
7	4	28
9	3	27

Daha sonra çarpım değerini toplarız. Çarpım değerleri toplamı, $\sum_{i=1}^n f_i X_i = 69$ 'dur. Bu toplam serideki toplam değer sayısı olan 10'a bölündüğünde serinin aritmetik ortalaması

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{\sum_{i=1}^n f_i} = \frac{69}{10} = 6.9$$

olarak elde edilir.



Serideki birimlerin aritmetik ortalamadan sapmalarının toplamı sıfırdır.



Örnek

•Aşağıdaki sınıflandırılmış serinin aritmetik ortalamasını hesaplayalım. Derslerin haftalık kredi saatleri onların tartılarıdır.

Dersler	Notu	Kredi Saati	$t_i X_i$
Matematik	85	3	255
Genel İktisat	89	3	267
Temel Hukuk	55	2	110
Muhasebe	60	1	60
Yönetim	60	1	60

Tartılı aritmetik ortalama hesaplanırken önce her bir ders notu kendisine karşı gelen tartı ile çarpılır. Daha sonra çarpım değerlerini toplarız. Çarpım değerleri toplamı $\sum_{i=1}^n t_i X_i = 752$ 'dir. Bu toplam serideki toplam tartı miktarı olan 10'a bölündüğünde serinin aritmetik ortalaması,

$$\bar{X} = \frac{\sum_{i=1}^n t_i X_i}{\sum_{i=1}^n t_i} = \frac{752}{10} = 75.2$$

olarak elde edilir. Bahse konu olan öğrencinin ağırlıklı not ortalaması 75.2'dir.



Örnek

- Aşağıdaki gruplandırılmış serinin aritmetik ortalamasını hesaplayalım.

Gruplar	Frekans
2-4	2
5-7	13
8-10	4
11-13	1

Gruplandırılmış seriden herhangi bir parametrik merkezî eğilim ölçüsü elde edilebilmesi için gruplandırılmış serideki sınıf sınırları veya sınıf uçlarından sınıf değerleri elde edilir. Mesela birinci sınıfın sınıf değeri 2 ile 4 aralığının tam ortası, yani 3'tür. İkinci sınıfın sınıf değeri 6'dır. Üçüncü sınıfın sınıf değeri 9'dur. Dördüncü sınıfın sınıf değeri 12'dir. Sınıf değerleri X_i ile gösterilir. Bundan sonraki işlemler sınıflandırılmış seriden aritmetik ortalamanın hesaplanmasına benzer. Her bir sınıf değeri kendisine karşı gelen sınıf frekansı ile çarpılır.

Gruplar	X_i	f_i	$f_i X_i$
2-4	3	2	6
5-7	6	13	78
8-10	9	4	36
11-13	12	1	12

Sınıf değerleri ile frekanslarının çarpımları toplamı, $\sum_{i=1}^n f_i X_i = 132$ 'dir. Bu toplam frekansı ifade eden 20 değerine bölündüğünde gruplandırılmış serinin aritmetik ortalaması,

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{\sum_{i=1}^n f_i} = \frac{132}{20} = 6.6$$

olarak elde edilir.

Aritmetik Ortalamanın Özellikleri

- Aritmetik dizi şeklinde artış veya azalış gösteren serileri en iyi temsil eden parametrik merkezî eğilim ölçüsüdür. Basit bir sayıya belirli bir sayının katlarının ilave edilmesiyle elde edilen diziye *aritmetik dizi* denir. Mesela 5 sayısına sabit bir sayı olan 3 sayısının sürekli olarak ilave edilmesiyle *5, 8, 11, 14, 17, 20, 23, 26, ...* serisi elde edilir. Bu seri aritmetik dizi şeklinde



Serideki birimlerin aritmetik ortalamadan sapmalarının kareleri toplamı minimumdur.

artışları gösteren bir seridir. Aritmetik ortalama, aritmetik diziye benzeyen serileri de en iyi temsil eden merkezî eğilim ölçüsüdür.

- Serideki birimlerin aritmetik ortalamadan sapmalarının toplamı sıfırdır [3].

X_i	$(X_i - \bar{X})$
10	-5
12	-3
18	3
20	5

Yukarıdaki basit serinin aritmetik ortalaması 15'tir. Serideki değerlerden serinin aritmetik ortalaması çıkarıldığında bazı farklar negatif, bazı farklar pozitif işaretli çıkmaktadır. Bu farklar toplandığında sıfır çıkmaktadır. $\Sigma(X_i - \bar{X}) = 0$

- Serideki birimlerin aritmetik ortalamadan sapmalarının kareleri toplamı minimumdur.

$$\Sigma(X_i - \bar{X})^2 = 68$$

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
10	-5	25
12	-3	9
18	3	9
20	5	25

Eşitlikteki $\bar{X}$ yerine $A = 14$ gibi bir başka değer koyduğumuzda bulacağımız $\Sigma(X_i - A)^2$ toplamı, $\Sigma(X_i - \bar{X})^2$ toplamından büyük çıkacaktır.

X_i	$(X_i - A)$	$(X_i - A)^2$
10	-4	16
12	-2	4
18	4	16
20	6	36

$\Sigma(X_i - \bar{X})^2 = 68$ değeri, $\Sigma(X_i - A)^2 = 72$ değerinden küçüktür.

- Bir serinin aritmetik ortalaması serinin toplam frekansı ile çarpılırsa serideki rakamların toplamı elde edilir. Yani yukarıdaki basit seride $n(\bar{X}) = \Sigma X_i$ olmalıdır. $4(15) = 60$ 'dır. Sınıflandırılmış ve gruplandırılmış serilerde ise $(\Sigma f)\bar{X} = \Sigma fX$ olur.

- Bir serideki rakamlar iki veya daha fazla serinin aynı hizadaki rakamlarının toplamına eşitse, bu serinin aritmetik ortalaması diğer serilerin aritmetik ortalamaları toplamına eşittir.

X	A	B
10	1	11
12	2	12
18	4	22
20	5	25

X serisinin aritmetik ortalaması $\bar{X} = 15$ 'tir. A serisinin aritmetik ortalaması $\bar{X}_A = 3$ 'tür. B serisi, X ve A serisinin aynı hizadaki değerlerinin toplamıdır. Bu sebeple $\bar{X} + \bar{X}_A = 15 + 3 = 18 = \bar{X}_B$ 'dir.

- Bir serideki rakamların her birinin sabit bir sayı ile çarpılması hâlinde bulunacak serinin ortalaması önceki serinin ortalamasının söz konusu sabit sayı ile çarpımına eşittir.

X	A
10	20
12	24
18	36
20	40

X serisinin aritmetik ortalaması $\bar{X} = 15$ 'tir. X serisindeki değerler 2 sayısı ile çarpılarak A serisindeki sayılar elde edilmiştir. A serisinin aritmetik ortalaması $\bar{X}_A = 2(\bar{X}) = 2(15) = 30$ 'dur.

- Serideki rakamlara belirli bir sabit sayının eklenmesi hâlinde bulunacak ortalama, önceki ortalamanın söz konusu sabit sayı ile toplamına eşittir.

X	A
10	12
12	14
18	20
20	22

X serisinin aritmetik ortalaması $\bar{X} = 15$ 'tir. X serisindeki değerlere 2 sayısı eklenerek A serisindeki sayılar elde edilmiştir. A serisinin aritmetik ortalaması $\bar{X}_A = 2 + (\bar{X}) = 2 + 15 = 17$ 'dir.

GEOMETRİK ORTALAMA (G)

Geometrik ortalama da aritmetik ortalama gibi serinin bütün birimlerine tabi bir ortalama çeşididir. Bu ortalama, serideki n tane birimin çarpımının n '



Geometrik ortalama, serideki “n” tane birimin çarpımının “n”inci dereceden kökü alınmak suretiyle hesaplanır.

dereceden kökü alınmak suretiyle hesaplanır. Seride sıfır veya negatif değer varsa geometrik ortalama hesaplanamaz. Geometrik ortalama, geometrik dizi şeklinde artış gösteren serileri en iyi temsil eden parametrik merkezî eğilim ölçüsüdür. Geometrik dizi bir sayının katlanarak değerler alması durumunda oluşan seridir. Mesela 2 değeri katlanarak değerler alırsa 2, 4, 8, 16, 32, ... serisi elde edilir.

Gözlem sonuçları arasındaki oransal (nispi) farkların mutlak farklardan önemli olduğu durumlarda geometrik ortalama başvurulur. Diğer bir ifade ile, gözlem sonuçlarının her biri bir önceki gözlem sonucuna bağlı olarak değişiyorsa ve bu değişimin hızı saptanmak istenirse geometrik ortalama sağlıklı sonuçlar verir.

Basit serilerde geometrik ortalama,

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n}$$

formülü ile hesaplanır. Bu formül seride çok sayıda rakam varsa pek elverişli değildir. Eşitliğin her iki tarafının logaritması alındığında, logaritması alınmış geometrik ortalama,

$$\log G = \frac{1}{n} [\log X_1 + \log X_2 + \dots + \log X_n]$$

formülü ile hesaplanır. logG elde edildikten sonra her iki tarafın anti logaritması alınarak geometrik ortalama hesaplanır. Sınıflandırılmış ve gruplandırılmış serilerde geometrik ortalama,

$$G = \sqrt[n]{X_1^{f_1} \cdot X_2^{f_2} \cdot \dots \cdot X_n^{f_n}}$$

formülüyle bulunur. Formül bu haliyle kullanılamaz. Her iki tarafın logaritması alındığında,

$$\log G = \frac{1}{\sum f_i} [f_1 \log X_1 + f_2 \log X_2 + \dots + f_n \log X_n]$$

olur. logG bulunduktan sonra her iki tarafın anti logaritması alınarak geometrik ortalama bulunur.

Örnek 1: Aşağıdaki basit serinin geometrik ortalamasını hesaplayalım. Bu seri daha önce aritmetik ortalamasının hesaplanışını göstermede kullanılmıştır.

X_i

10

12

18

20

Geometrik ortalamanın hesaplanabilmesi için önce serideki rakamların logaritması alınır. X serisindeki değerlerin logaritmaları, $\log 10 = 1$, $\log 12 = 1.0792$,

$\log 18 = 1.2553$, $\log 20 = 1.301$ 'dur. Serideki $n = 4$ değerin logaritma değerlerinin toplamı, $\sum_{i=1}^4 \log X_i = 4.6355$ 'tir. Bu değer $\log G$ formülünde yerine yazılırsa;

$$\log G = \frac{4.6355}{4} = 1.1589$$

olarak elde edilir. Eşitliğin her iki tarafının anti logaritması alınırsa serinin geometrik ortalaması, $G = 14.42$ olarak elde edilir. Hatırlanacağı üzere aynı serinin aritmetik ortalaması 15'tir. Aynı seriden hem aritmetik ortalama hem de geometrik ortalama elde edilirse $G < \bar{X}$ olduğu gözlenecektir.

Örnek 2: Aşağıdaki sınıflandırılmış serinin geometrik ortalamasını hesaplayalım:

X_i	f_i
4	1
5	2
7	4
9	3

Sınıflandırılmış serilerin geometrik ortalaması hesaplanırken önce her bir X_i değişkenin logaritması alınır. Daha sonra logaritmali değerleri tekrarlanma sayısını ifade eden f_i değerleri ile çarpılır. Çarpım sonuçları aşağıdaki gibidir.

X_i	f_i	$\log X_i$	$f_i \log X_i$
4	1	0.6021	0.6021
5	2	0.6990	1.3980
7	4	0.8451	3.3804
9	3	0.9542	2.8626

Çarpım sonuçları toplamı, $\sum f_i \log X_i = 8.2431$ 'dir. Bu değeri $\log G$ formülünde yerine yazdığımızda,

$$\log G = \frac{\sum f_i \log X_i}{\sum f_i} = \frac{8.2431}{10} = 0.8243$$

olarak elde edilir. Eşitliğin her iki tarafının anti logaritması alındığında sınıflandırılmış serinin geometrik ortalaması, $G = 6.67$ olarak elde edilir. Hatırlanacağı üzere aynı serinin aritmetik ortalaması 6.9'dır. İki parametrik merkezî eğilim ölçüsü arasındaki $G < \bar{X}$ büyüklük ilişkisi yine gözlenmiştir.[4]

Örnek 3: Aşağıdaki gruplandırılmış serinin aritmetik ortalamasını hesaplayalım.

Gruplar	f_i
2-4	2
5-7	13
8-10	4
11-13	1

Gruplandırılmış seriden herhangi bir parametrik merkezî eğilim ölçüsü elde edilebilmesi için gruplandırılmış serideki sınıf sınırları veya sınıf uçlarından sınıf değerleri elde edilir. Bundan sonraki işlemler sınıflandırılmış seriden geometrik ortalamasının hesaplanmasına benzer. Her bir sınıf değerinin logaritması alınır. Daha sonra bu logaritmik değerler kendisine karşı gelen sınıf frekansı ile çarpılır.[5]

Gruplar	f_i	X_i	$\log X_i$	$f_i \log X_i$
2-4	2	3	0.4771	0.9542
5-7	13	6	0.7782	10.1166
8-10	4	9	0.9542	3.8168
11-13	1	12	1.0792	1.0792

Çarpım sonuçları toplamı, $\sum f_i \log X_i = 15.9668$ ’dir. Bu değeri logG formülünde yerine yazdığımızda,

$$\log G = \frac{\sum f_i \log X_i}{\sum f} = \frac{15.9668}{20} = 0.7983$$

olarak elde edilir. Eşitliğin her iki tarafının anti logaritması alındığında sınıflandırılmış serinin geometrik ortalaması, $G = 6.28$ olarak elde edilir. Hatırlanacağı üzere aynı serinin aritmetik ortalaması 6.6’dır. Gruplandırılmış seriden elde edilen iki parametrik merkezî eğilim ölçüsü arasında da $G < \bar{X}$ büyüklük ilişkisi gözlenmiştir.

HARMONİK ORTALAMA (H)

Harmonik ortalama serideki birimlerin çarpmaya göre terslerinin aritmetik ortalamasının tersidir. Seride sıfır veya negatif birim bulunması hâlinde harmonik ortalama kullanılmaz. Basit serilerde harmonik ortalama,

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

formülüyle hesaplanır. Sınıflandırılmış ve gruplandırılmış serilerde harmonik ortalama,

$$H = \frac{\sum f_i}{\sum_{i=1}^n \frac{f_i}{X_i}}$$

formülüyle hesaplanır.

Örnek 1: Aşağıdaki basit serinin harmonik ortalamasını hesaplayalım:

X_i
10
12
18
20



Harmonik ortalama, serideki birimlerin çarpmaya göre terslerinin aritmetik ortalamasının tersidir.

Harmonik ortalamanın hesaplanabilmesi için önce $1/X_i$ değerlerinin bulunması gerekir.

X_i	$1/X_i$
10	0,1000
12	0,0833
18	0,0556
20	0,0500

Daha sonra $1/X_i$ değerlerinin toplamı, $\sum_{i=1}^n \frac{1}{X_i} = 0.2889$ olarak elde edilir. Bu değer formülde yerine yazılırsa basit serinin harmonik ortalaması,

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} = \frac{4}{0.2889} \cong 13.85$$

olarak elde edilir. Bu basit serinin aritmetik ortalaması 15 ve geometrik ortalaması 14.42 olarak hesaplanmıştır. Harmonik ortalama parametrik merkezî eğilim ölçülerinin en küçüğü olarak elde edilir. Şimdiye kadarki parametrik merkezî eğilim ölçüleri arasında $H < G < \bar{X}$ şeklinde büyüklük ilişkisi vardır.[6]

Örnek 2: Aşağıdaki sınıflandırılmış serinin geometrik ortalamasını hesaplayalım:

X_i	f_i
4	1
5	2
7	4
9	3

Sınıflandırılmış serilerin geometrik ortalaması hesaplanırken önce f_i/X_i değerleri bulunur.

X_i	f_i	f_i/X_i
4	1	0,2500
5	2	0,4000
7	4	0,5714
9	3	0,3333

Daha sonra bu değerler toplanarak, $\sum_{i=1}^n \frac{f_i}{X_i} = 1.5547$ değeri bulunur. Bu değer formülde yerine yazılırsa sınıflandırılmış serinin harmonik ortalaması,

$$H = \frac{\sum f_i}{\sum_{i=1}^n \frac{f_i}{X_i}} = \frac{10}{1.5547} \cong 6.43$$

olarak elde edilir. Bu örneğe göre de parametrik merkezî eğilim ölçüleri arasındaki $H < G < \bar{X}$ şeklinde büyüklük ilişkisi gözlenmektedir [7].

Örnek 3: Aşağıdaki gruplandırılmış serinin harmonik ortalamasını hesaplayalım:

Gruplar	f_i
2-4	2
5-7	13
8-10	4
11-13	1

Gruplandırılmış seriden herhangi bir parametrik merkezî eğilim ölçüsünün elde edilebilmesi için gruplandırılmış serideki sınıf sınırları veya sınıf uçlarından sınıf değerleri elde edilir. Bundan sonraki işlemler sınıflandırılmış seriden harmonik ortalamasının hesaplanmasına benzer. Her bir sınıfın frekansını ve değerini kullanarak f_i/X_i değerleri bulunur.

Gruplar	X_i	f_i	f_i/X_i
2-4	3	2	0,6667
5-7	6	13	2,1667
8-10	9	4	0,4444
11-13	12	1	0,0833

Daha sonra bu değerler toplanarak, $\sum_{i=1}^n \frac{f_i}{X_i} = 3.3611$ değeri bulunur. Bu değer formülde yerine yazılırsa sınıflandırılmış serinin harmonik ortalaması,

$$H = \frac{\sum f}{\sum \frac{f_i}{X_i}} = \frac{20}{3.3611} \cong 5.95$$

olarak elde edilir. Daha önce söz konusu gruplandırılmış serinin aritmetik ortalaması 6.6 ve geometrik ortalaması 6.28 olarak elde edilmiştir. Bu örneğe göre de parametrik merkezî eğilim ölçüleri arasında $H < G < \bar{X}$ şeklinde büyüklük ilişkisi vardır.

KARELİ ORTALAMA (K)

Kareli ortalama, serideki birimlerin karelerinin aritmetik ortalamasının kareköküdür [8]. Basit serilerde kareli ortalama,

$$K = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$$

formülüyle hesaplanır. Sınıflandırılmış ve gruplandırılmış serilerde kareli ortalama,

$$K = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i}}$$

formülüyle hesaplanır.

Örnek 1: Aşağıdaki basit serinin kareli ortalamasını hesaplayalım:



Kareli ortalama, serideki birimlerin karelerinin aritmetik ortalamasının kareköküdür.

x_i

10

12

18

20

Basit serilerde kareli ortalama hesaplanırken önce serideki değerlerin kareleri hesaplanır.

 x_i x^2

10 100

12 144

18 324

20 400

Sonra kareli değerleri toplamı, $\sum_{i=1}^4 x_i^2 = 968$ olarak elde edilir. Daha sonra kareli değerlerin toplamı serideki değer sayısı olan 4'e bölündüğünde elde edilen değer karekökü alınarak kareli ortalama,

$$K = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} = \sqrt{\frac{968}{4}} = \sqrt{242} \cong 15.56$$

olarak elde edilir. Bu basit serinin aritmetik ortalaması 15, geometrik ortalaması 14.42 ve harmonik ortalaması 13.85 olarak hesaplanmıştır. Kareli ortalama parametrik merkezî eğilim ölçülerinin en büyüğü olarak elde edilir. Parametrik merkezî eğilim ölçüleri arasında $H < G < \bar{X} < K$ şeklinde büyüklük ilişkisi vardır.[11]

Örnek 2: Aşağıdaki sınıflandırılmış serinin kareli ortalamasını hesaplayalım:

 x_i f_i

4 1

5 2

7 4

9 3

Sınıflandırılmış serilerin kareli ortalaması hesaplanırken önce x_i değerlerinin karesi hesaplanarak f_i değerleri ile çarpılır.

 x_i f_i x^2 fx^2

4 1 16 16

5 2 25 50

7 4 49 196

9 3 81 243

Daha sonra $f_i X_i$ çarpımları toplanarak, $\sum_{i=1}^n f_i X_i^2 = 505$ değeri elde edilir. Bu değer serideki toplam rakam sayısı olan 10'a bölünerek karekökü alınırsa sınıflandırılmış serinin kareli ortalaması,

$$K = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}} = \sqrt{\frac{505}{10}} = \sqrt{50.5} \approx 7.11$$

olarak elde edilir. Bu örnekte de parametrik merkezî eğilim ölçüleri arasındaki büyüklük ilişkisi $H < G < \bar{X} < K$ şeklinde gözlenir [9].

Örnek 3: Aşağıdaki gruplandırılmış serinin kareli ortalamasını hesaplayalım:

Gruplar	f_i
2-4	2
5-7	13
8-10	4
11-13	1

Gruplandırılmış seriden herhangi bir parametrik merkezî eğilim ölçüsü elde edilebilmesi için gruplandırılmış serideki sınıf sınırları veya sınıf uçlarından sınıf değerleri elde edilir. Bundan sonraki işlemler sınıflandırılmış seriden kareli ortalamanın hesaplanmasına benzer. Her bir sınıf değerinin karesi bulunur. Daha sonra bu kareli değerler kendisine karşı gelen sınıfın frekansı ile çarpılır [12].

Gruplar	X_i	f_i	X_i^2	$f_i X_i^2$
2-4	3	2	9	18
5-7	6	13	36	468
8-10	9	4	81	324
11-13	12	1	144	144

Daha sonra bu değerler toplanarak, $\sum_{i=1}^n f_i X_i^2 = 954$ değeri bulunur. Bu değer formülde yerine yazılırsa sınıflandırılmış serinin kareli ortalaması,

$$K = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}} = \sqrt{\frac{954}{20}} = \sqrt{47.7} \approx 6.91$$

olarak elde edilir. Daha önce söz konusu gruplandırılmış serinin aritmetik ortalaması 6.6, geometrik ortalaması 6.28 ve harmonik ortalaması 5.95 olarak elde edilmiştir. Bu gruplandırılmış seriden elde edilen parametrik merkezî eğilim ölçüleri arasında $H < G < \bar{X} < K$ şeklinde büyüklük ilişkisi vardır [13].



Parametrik merkezî eğilim ölçüleri arasında $H < G < \bar{X} < K$ şeklinde büyüklük ilişkisi vardır.



Bireysel Etkinlik

- Parametrik merkezi eğilim ölçülerinin hesaplanmasında serideki hangi birimler dikkate alınır?
- Sosyal etkinlik için biraraya gelenlerin yaşları 50, 55, 46, 40, 35, 60, 65, 45, 50, 55'dir. Yaşların aritmetik ortalamasını hesaplayınız.
- Yukarıdaki örnekte 3. kişinin yaşını kendiniz değiştirip tekrar yaşların aritmetik ortalamasını hesaplayınız. Her seferinde bir kişinin yaşını değiştirip aritmetik ortalama hesaplaması yaptığınızda değişimi gözlemleyiniz.



Özet

- Merkezî eğilim ölçüleri, istatistiğin özetleme görevini en ileri seviyede gören istatistik ölçülerdir.
- Merkezî eğilim ölçüleri başlıca iki gruba ayrılır:
 - 1-) Serinin bütün birimlerine tabi olan merkezî eğilim ölçüleri
 - 2-) Serinin bütün birimlerine tabi olmayan merkezî eğilim ölçüleri
- Birinci gruba giren merkezî eğilim ölçülerine parametrik merkezî eğilim ölçüleri de denilmektedir. Bu gruptaki merkezî eğilim ölçüleri serideki tek bir rakamın değişmesinden doğrudan doğruya etkilenirler. Bu sebeple parametrik merkezî eğilim ölçülerinin tamamı serideki aşırı uçların etkisinde kalırlar. Sınıf uçları belli olmayan gruplandırılmış serilerde sınıf değerleri hesaplanamayacağı için parametrik merkezî eğilim ölçülerinin hiçbirisi hesaplanamaz.
- Bu gruptaki merkezî eğilim ölçüleri; aritmetik ortalama ($\bar{X}$), geometrik ortalama (G), harmonik ortalama (H) ve kareli ortalamadır (K).
- **ARİTMETİK ORTALAMA ($\bar{X}$)**
 - Yukarıda da temas edildiği gibi, aritmetik ortalama serideki bütün rakamlardan etkilenen bir ortalamadır. İstatistiki uygulamalarda en çok kullanılan merkezî eğilim ölçüsüdür. Basit bir seride aritmetik ortalama serideki birimlerin toplamının birim sayısına bölümüyle elde edilir.
- **GEOMETRİK ORTALAMA (G)**
 - Geometrik ortalama da aritmetik ortalama gibi serinin bütün birimlerine tabi bir ortalama çeşididir. Bu ortalama, serideki n tane birimin çarpımının n' inci dereceden kökü alınmak suretiyle hesaplanır. Seride sıfır veya negatif değer varsa geometrik ortalama hesaplanamaz. Geometrik ortalama, geometrik dizi şeklinde artış gösteren serileri en iyi temsil eden parametrik merkezî eğilim ölçüsüdür. Geometrik dizi bir sayının katlanarak değerler alması durumunda oluşan seridir. Gözlem sonuçları arasındaki oransal (nispi) farkların mutlak farklardan önemli olduğu durumlarda geometrik ortalamaya başvurulur.
- **HARMONİK ORTALAMA (H)**
 - Harmonik ortalama serideki birimlerin çarpıma göre terslerinin aritmetik ortalamasının tersidir. Seride sıfır veya negatif birim bulunması hâlinde harmonik ortalama kullanılmaz. Harmonik ortalama parametrik merkezî eğilim ölçülerinin en küçüğü olarak elde edilir.
- **KARELİ ORTALAMA (K)**
 - Kareli ortalama, serideki birimlerin karelerinin aritmetik ortalamasının kareköküdür. Basit serilerde kareli ortalama hesaplanırken önce serideki değerlerin kareleri hesaplanır. Sonra kareli değerleri toplamı elde edilir. Daha sonra kareli değerlerin toplamı serideki değer sayısına bölündüğünde elde edilen değer karekökü alınarak kareli ortalama elde edilir.
 - Kareli ortalama parametrik merkezî eğilim ölçülerinin en büyüğü olarak elde edilir. Parametrik merkezî eğilim ölçüleri arasında $H < G < X < K$ şeklinde büyüklük ilişkisi vardır.

DEĞERLENDİRME SORULARI

1. Bir kibrit fabrikasının ürettiği kibritlerden rastgele seçilen 10 paket kibritin içindeki kibrit çöpleri sayılmış ve sırasıyla 39, 43, 35, 46, 47, 38, 41, 40, 45 ve 36 olarak bulunmuştur. Buna göre bu kibrit fabrikasının ürettiği paketlerdeki çöplerin aritmetik ortalaması kaçtır?
- 39
 - 40
 - 41
 - 42
 - 43

2. Yanda verilen sınıflandırılmış serinin aritmetik ortalaması aşağıdakilerden hangisidir?
- 5.8
 - 6.4
 - 6.8
 - 7.1
 - 7.3

X_i	f_i
3	1
5	4
7	2
9	3

3. Yanda verilen gruplandırılmış serinin aritmetik ortalaması aşağıdakilerden hangisidir?
- 3
 - 4
 - 5
 - 6
 - 7

Gruplar	f_i
3-5	4
5-7	6
7-9	3
9-11	5

4. Yanda verilen sınıflandırılmış serinin aritmetik ortalaması aşağıdakilerden hangisidir?
- 10.25
 - 10.55
 - 11.05
 - 12.15
 - 12.55

X_i	f_i
6	5
10	7
14	3
17	4

5. Dört ülkenin işsizlik oranları (%) şu şekildedir: 5, 8, 12, 20. Bu basit serinin geometrik ortalaması aşağıdakilerden hangisidir?
- 8.82
 - 9.89
 - 10.45
 - 11.34
 - 12.76

6. Yukarıdaki sınıflandırılmış serinin geometrik ortalaması aşağıdakilerden hangisidir?

- a) 4.98
b) 5.25
c) 5.68
d) 6.07
e) 6.46

X_i	f_i
3	1
5	4
7	2
9	3

7. Yanda sınıflandırılmış serinin geometrik ortalaması aşağıdakilerden hangisidir?

- a) 13.76
b) 13.10
c) 12.85
d) 11.45
e) 10.31

X_i	f_i
6	5
10	7
14	3
17	4

8. Yanda gruplandırılmış serinin geometrik ortalaması aşağıdakilerden hangisidir?

- a) 5.99
b) 6.32
c) 6.88
d) 7.51
e) 7.95

Gruplar	f_i
1-5	4
5-9	6
9-13	3

9. Bir fabrikada üretilen ilk 1000 üründe her bir ürün 3 dk'da üretilirken, ikinci 1000 üründe her bir ürün 5 dk'da üretilmektedir. Buna göre bu fabrikada ürün üretme hızının harmonik ortalaması kaçtır?

- a) 3.12
b) 3.45
c) 3.75
d) 3.98
e) 4.24

10. Yukarıdaki sınıflandırılmış serinin harmonik ortalaması aşağıdakilerden hangisidir?

- a) 9.1
b) 9.58
c) 10.24
d) 10.68
e) 11.34

X_i	f_i
6	5
10	7
14	3
17	4

Cevap Anahtarı

1.c, 2.b, 3.e, 4.c, 5.b, 6.d, 7.e, 8.a, 9.c, 10.b

YARARLANILAN KAYNAKLAR

- [1] Çakır, F. (2000). Sosyal Bilimlerde İstatistik. İstanbul: Alfa Yayıncılık.
- [2] Başar, Alaaddin, Oktay, Erkan. (2018). Uygulamalı İstatistik-I: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap ve Kirtasiye.
- [3] Tekin, A. (2012). Temel İstatistik Dersleri. Konya: Eğitim Yayınevi.
- [4] Büyüköztürk, Ş. (2018). Sosyal Bilimler için Veri Analizi El Kitabı (31. baskı). Ankara: Pegem Akademi Yayıncılık.
- [6] Gürsakal, N., Oğuzlar, A, Gürsakal, S. (2022). Betimsel İstatistik (10. Baskı). Bursa: Dora Yayıncılık.
- [7] Hayran, M. (1996). Bilgisayar İstatistik ve Tıp. Ankara: Medikomat Basım Yayın.
- [8] Köseoğlu M., Yamak, R. (2013). Uygulamalı İstatistik (4. baskı). Trabzon: Celepler Matbaacılık.
- [9] Newbold, P. (2009). Statistics for Business and Economics (7. baskı). İstanbul: Literatür Yayıncılık.
- [10] Özsoy, O. (2010). İktisatçılar ve İşletmeciler için İstatistik (3. baskı). Ankara: Siyasal Kitabevi.
- [11] Johnson, A. R., W. D. Wichern (1998). Applied Multivariate Statistical Analysis (4th Edition). London: Prentice-Hall.
- [12] Yıldız, N. T. (2008). Uygulamalı İstatistik (3. baskı). Ankara: Nobel Yayıncılık.
- [13] Yılmaz, B. (2018). İstatistik (4. baskı). Ankara: Nobel Yayıncılık.

PARAMETRİK OLMAYAN MERKEZİ EĞİLİM ÖLÇÜLERİ



İÇİNDEKİLER

- Mod
- Medyan
- Kantiller
 - Kartiller
 - Desiller
 - Pörsentiller



HEDEFLER

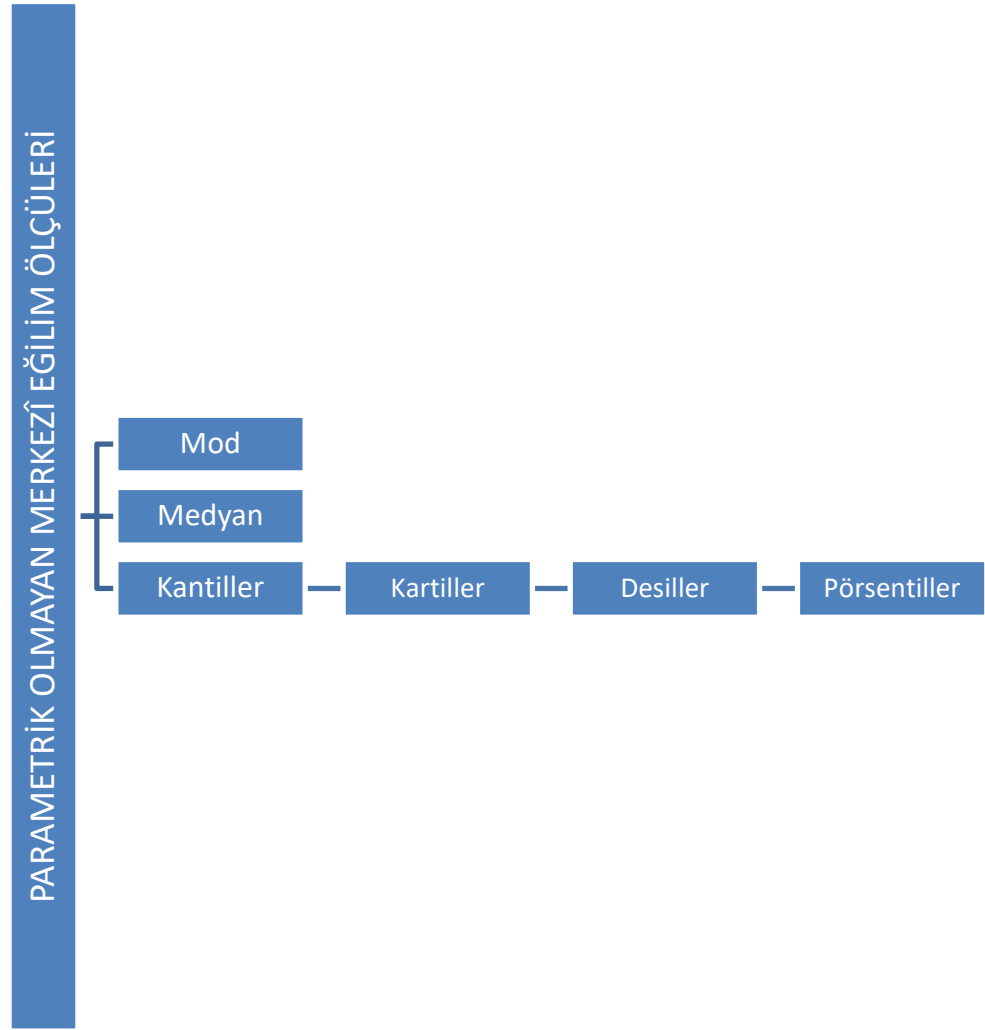
- Bu üniteyi çalıştıktan sonra;
 - Parametrik olmayan merkezî eğilim ölçüleri kavramını açıklayabilecek,
 - İstatistiksel serilerin mod, medyan ve kantillerini bulabilecek,
 - Parametrik olmayan merkezî eğilim ölçülerinin hangi serilerden elde edilebileceğini ifade edebilecek,
 - Parametrik olmayan merkezî eğilim ölçüleri hakkında bilgi sahibi olabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Dr. Emrah TALAŞ

ÜNİTE
6



GİRİŞ

Parametrik merkezî eğilim ölçüleri serideki tüm verilerden elde edilirken, parametrik olmayan merkezî eğilim ölçüleri serideki bazı verilerden elde edilmektedir. Bu tür ortalamalar serideki tek bir rakamın değişiminden ve aşırı uçlardan direkt etkilenmezler. Parametrik olmayan merkezî eğilim ölçüleri;

- Mod
- Medyan
- Kantiller

olmak üzere üç grupta sınıflandırılırlar. Ayrıca kantiller de kendi aralarında

- Kartiller
- Desiller
- Pörsentiller

olmak üzere üç grupta incelenirler.

MOD

Herkesin çok iyi bildiği “moda” kelimesi moddan türetilmiş bir kelimedir. En yaygın kullanılan, en çok beğenilen, en yaygın gözlenen anlamına gelir. *Mod*, incelenen bir seride en fazla tekrar eden ya da başka bir ifadeyle *frekansı en yüksek olan gözlem değeridir* [1].

Bir seride birden fazla mod olabileceği gibi seride mod bulunmayabilir. Bu tür serilerde modun merkezî eğilim ölçüsü olarak değeri olmaz. Mod, grafiksel olarak gösterildiğinde grafiğin tepe noktasında olduğundan, *tepe değer* olarak da ifade edilebilir.

Örnek 1: Bir iş yeri sahibi, personelinin çalışma saatlerini aşağıdaki gibi yazmıştır:

4, 4, 4, 6, 6, 6, 6, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 10, 10, 10, 10

Bu seride en çok tekrar eden değer 8 olduğu için bu serinin mod'u 8'dir. Bu iş yerinde en yaygın çalışma süresi 8 saattir.

Örnek 2: Bir sınıftaki öğrencilerin yaşlarına göre dağılımı aşağıdaki gibi

Yaşlar	Frekans
19	40
20	50
21	80
22	30
23	20

Bu sınıfta en çok gözlenen yaş 21'dir. Bu serinin modu 21'dir.

Gruplandırılmış seride içerisinde en fazla değer bulunduran sınıf, mod sınıfıdır. Mod sınıfı, içerisinde modu bulunduran sınıftır. Mod sınıfından modu bulurken aşağıdaki formül kullanılır.



Mod, incelenen bir seride en fazla tekrar eden, frekansı en yüksek olan gözlem değeridir.

$$Mod = L_1 + \frac{\Delta_1}{\Delta_2 + \Delta_1} c$$

L_1 = Mod sınıfının alt sınırındır.

Δ_1 = Mod sınıfının frekansı ile mod sınıfından bir önceki sınıfın frekansı arasındaki farktır.

Δ_2 = Serinin modunu barındıran grubun frekansı ile bir sonraki grubun frekansı arasındaki farktır.

c = Serinin modunu barındıran grubun sınıf büyüklüğüdür [2].

Örnek 3: Bir kursa katılan öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Bu kursa en çok hangi yaşta öğrenci katılmaktadır?



Bir seride verilerin hepsi birbirinden farklı ise bu seride mod yoktur.

Gruplar	Frekans
6-8	5
8-10	15
10-12	55
12-14	25

Öncelikle serinin modunun hangi sınıfta olduğu tespit edilir. Frekansı en yüksek olan 10-12 sınıfı, dağılımın modunu içinde barındıran sınıftır. Mod sınıfının alt sınırı 10'dur. Mod sınıfının frekansı ile mod sınıfından bir önceki sınıfın frekansı arasındaki fark $\Delta_1 = 40$ 'tır. Mod sınıfının frekansı ile mod sınıfından bir sonraki sınıfın frekansı arasındaki fark $\Delta_2 = 30$ 'dur. Mod sınıfının sınıf büyüklüğü üst sınıf sınırı 12 ile alt sınıf sınırı 10 arasındaki farktır. Bilinen değerler mod formülünde yerine yazıldığında gruplandırılmış serinin modu,

$$Mod = 10 + \frac{40}{40 + 30} 2 = 11.14$$

olarak elde edilir. Bahse konu olan kursa en çok yaklaşık 11 yaşındaki öğrenciler katılmıştır.

Örnek 4: Bir kursa katılan öğrencilerin yaşlara göre dağılımı aşağıdaki gibidir. Bu kursa en çok hangi yaşta öğrenci katılmaktadır?

Gruplar	Frekans
6-8	15
8-10	55
10-12	45
12-14	25

Öncelikle serinin mod'unun hangi sınıfta olduğu tespit edilir. Frekansı en yüksek olan 8-10 sınıfı, dağılımın mod'unu içinde barındıran sınıftır. Mod sınıfının alt sınırı 8'dir. Mod sınıfının frekansı ile mod sınıfından bir önceki sınıfın frekansı arasındaki fark $\Delta_1 = 40$ 'tır. Mod sınıfının frekansı ile mod sınıfından bir sonraki sınıfın frekansı arasındaki fark $\Delta_2 = 10$ 'dur. Mod sınıfının sınıf büyüklüğü, üst sınıf sınırı 10 ile alt sınıf sınırı 8 arasındaki farktır. Bilinen değerler mod formülünde yerine yazıldığında gruplandırılmış serinin modu,

$$Mod = 8 + \frac{40}{40 + 10} 2 = 9.6$$

olarak elde edilir.



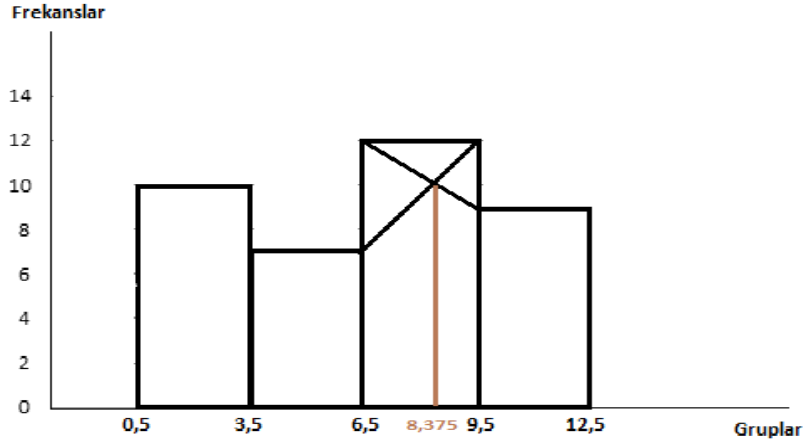
Mod, grafik ile de tespit edilebilir.

Mod, grafikte de tespit edilebilir. Verilmiş olan grubun, histogramı üzerindeki frekans eksenini en büyük olan kutucuğun, üst köşe noktaları ile komşu kutucuklara değdiği en üst noktaların çapraz bir şekilde birleştirilmesi ile oluşan kesişim noktasının yatay eksenindeki karşılığı serinin modudur [3].

Örnek 5: Bir köyde yaşayan çocukların yaşlarına göre dağılımı aşağıdaki gibidir. Bu köyde en çok hangi yaşta çocuk bulunmaktadır?

Gruplar	Frekans
1-3	10
4-6	7
7-9	12
10-12	9

Yukarıda verilmiş olan gruplandırılmış kesikli serinin modunun grafikte gösterimi Görsel 6.1'de yer almaktadır. Grafikte de görüleceği gibi serinin modu 8.375'tir.



Görsel 6.1. Çocukların yaş gruplarına göre dağılım grafiği

MEDYAN

Veriler büyükten küçüğe veya küçükten büyüğe doğru sıralandığında serinin tam ortasına karşılık gelen değere *medyan ya da ortanca* denir.

Büyüklik sırasına göre sıralanmış basit bir serideki veri sayısı n olmak üzere, n tek ise medyan $(n+1)/2$ 'inci değerdir. Örnek hacmi n çift ise medyan $n/2$ 'inci değer ile $(n/2)+1$ 'inci değer aritmetik ortalamasıdır [4].

Örnek 1: Basit bir serideki değerler **2 2 3 5 5 7 8 10 12 16 16** şeklindedir. Bu serinin ortanca değerini bulalım:

Serideki değerler küçükten büyüğe doğru sıralandığı için serinin tam ortasındaki değer medyandır. Serisinde gözlem değerleri tek olduğu için medyan



Veriler büyüklüklerine göre sıralandığında serinin tam ortasına karşılık gelen değere medyan denir.

$(n+1)/2$ 'inci değerdir. Yani medyan $(11+1)/2 = 6$ ıncı değerdir. Küçükten büyüğe doğru altı değer saydığımızda altıncı değer olan 7 değeri bu serinin medyanıdır.

Örnek 2: Basit bir serideki değerler **10 10 10 12 12 14 17 19 20 20** şeklindedir. Bu serinin medyan değerini bulalım:



Medyanı bulabilmek için serideki değerler öncelikle büyüklük sırasına konmalıdır.

Serideki değer sayısı 10'dur. Örnek hacmi çift sayıdır. Serideki değerler küçükten büyüğe doğru sıralandığı için serinin tam ortasındaki iki değer ortalaması medyandır. Serisinde gözlem değerleri çift olduğu için medyan $n/2$ inci yani beşinci değer ile $n/2+1$ yani altıncı değerin ortalaması medyandır. Bahsedilen seride beşinci değer olan 12 değeri ile altıncı değer olan 14 değerlerinin aritmetik ortalaması olan 13 değeri, bu serinin medyanıdır.

Sınıflandırılmış seriden medyan bulunabilmesi için öncelikle ...den az kümülatif frekans değerlerinin bulunması gerekir. Medyanı gösteren $n/2$ 'inci değeri ilk kez içerisinde bulunduran kümülatif frekansa sahip olan değer medyan değeridir.

Örnek 3: Bir sınıftaki öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Bu serinin medyanını bulabilmek için önce “...den az kümülatif frekans” değerleri hesaplanır.

Puanlar	Frekans	 den az kümülatif frekans
19	10	10
20	40	50
21	100	150
22	30	180
23	20	200

Bu seride toplam 200 değer vardır. Bu seride $n/2$ 'inci değer, yani $200/2 = 100$ 'üncü değer medyandır. 100'üncü değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 150'nin gösterdiği puan olan 21 değeri bu serinin medyanıdır.

Gruplandırılmış seriden de medyanın bulunabilmesi için öncelikle “...den az kümülatif frekans” değerlerinin bulunması gerekir. Kümülatif frekans serisinde $n/2$ 'inci değeri ilk kez içerisinde bulunduran sınıf medyan sınıfıdır. Medyanı içerisinde bulunduran sınıf bulunduktan sonra tam olarak medyan değerini elde edebilmek için,

$$\text{Medyan} = L_1 + \frac{\frac{n}{2} - \sum f_{m-1}}{f_m} c$$

formülü kullanılır. Bu formüldeki;

L_1 = Medyan sınıfının alt sınırı

n = Toplam frekans

$\sum f_{m-1}$ = Medyan sınıfından bir önceki sınıfın kümülatif frekansı

f_m = Medyan sınıfının frekansı

c = Medyan sınıfının sınıf büyüklüğüdür [5].

Örnek 4: Bir kursa katılan öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Bu kursa katılan öğrencilerin ortalama yaşı nedir?

Gruplar	Frekans	 den az kümülatif frekans
6-8	5	5
8-10	15	20
10-12	55	75
12-24	25	100

Öncelikle medyanın hangi sınıfta olduğu tespit edilir. Serinin $n/2 = 100/2 = 50$ ' inci değeri medyandır. 50' inci değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 75'in gösterdiği sınıf medyan sınıfıdır. Bu serinin medyanı 10 ile 12 arasındadır. Medyan sınıfının alt sınırı 10'dur. Medyan sınıfından bir önceki sınıfın kümülatif frekansı $\Sigma f_{m-1} = 20$ 'dir. Medyan sınıfının frekansı $f_m = 55$ 'tir. Medyan sınıfının sınıf büyüklüğü üst sınıf sınırı 12 ile alt sınıf sınırı 10 arasındaki farka eşittir. Yani $c = 12 - 10 = 2$ 'dir. Bilinen değerler medyan formülünde yerine yazıldığında gruplandırılmış serinin medyanı,

$$\text{Medyan} = 10 + \frac{\frac{100}{2} - 20}{55} 2 = 11.09$$

olarak elde edilir. Bahse konu olan kursa katılan öğrencilerin ortalama yaşı yaklaşık olarak 11'dir.

Örnek 5: Bir kursa katılan öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Bu kursa en çok hangi yaşta öğrenci katılmaktadır? Bu kursa katılan öğrencilerin ortalama yaşı nedir?

Gruplar	Frekans	 den az kümülatif frekans
6-8	5	5
9-11	15	20
12-14	55	75
15-17	25	100

Öncelikle medyanın hangi sınıfta olduğu tespit edilir.

Serinin $n/2 = 100/2 = 50$ ' inci değeri medyandır. 50' inci değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 75'in gösterdiği sınıf medyan sınıfıdır. Bu serinin medyanı 11.5 ile 14.5 arasındadır. Medyan sınıfının alt sınırı 11.5'tir. Medyan sınıfından bir önceki sınıfın kümülatif frekansı $\Sigma f_{m-1} = 20$ 'dir. Medyan sınıfının frekansı $f_m = 55$ 'tir. Medyan sınıfının sınıf büyüklüğü üst sınıf sınırı 14.5 ile alt sınıf sınırı 11.5 arasındaki farka eşittir. Yani $c = 14.5 - 11.5 = 3$ 'tür.

Bilinen değerler medyan formülünde yerine yazıldığında gruplandırılmış serinin medyanı,

$$Medyan = 11.5 + \frac{\frac{100}{2} - 20}{55} 3 = 13.14$$

olarak elde edilir. Bahse konu olan kursa katılan öğrencilerin ortanca yaşı yaklaşık olarak 13'tür.

KANTİLLER

Kantiller bir seriyi 4, 10 ve 100 eşit parçaya ayırarak bu serideki değerlerin dörtte, onda ve yüzde ne kadarının belirli bir değere göre yerini saptamak için kullanılır [6].

Kartiller

Büyüklik sırasına dizilmiş bir serinin dört eşit parçaya bölünmesi sonucu *üç kartil bulunur*. Küçükten büyüğe doğru sıralanan seriyi dört parçaya bölebilmek için üç bölen gerekir. Birinci kartil Q_1 , ikinci kartil Q_2 ve üçüncü kartil Q_3 ile gösterilir. Her bir kartil aralığı yaklaşık serideki rakamların %25'ini kapsar. Bir serinin ikinci kartili, medyandır. Basit serilerde;

$$Q_1 = \frac{N+1}{4}, \text{ üncü değerdir.}$$

$$Q_2 = \frac{2(N+1)}{4}, \text{ üncü değerdir.}$$

$$Q_3 = \frac{3(N+1)}{4}, \text{ üncü değerdir.}$$

Örnek 1: Basit bir serideki değerler **10 10 10 12 12 14 17 19 20 20 20 20** şeklindedir. Bu serinin kartillerini bulalım:

Serideki değer sayısı 11'dir. Serideki değerler küçükten büyüğe doğru sıralandığı $(11+1)/4 = 3$. değer olan 10 değeri bu serinin 1. kartilidir. Serinin ikinci kartili, aynı zamanda medyanı $2(11+1)/4 = 6$. değer olan 14 değeridir. Serinin üçüncü kartili, $3(11+1)/4 = 9$. değer olan 20 değeridir.

Sınıflandırılmış ve gruplandırılmış serilerde çok sayıda değer yer aldığı için yukarıdaki ifadelerde yer alan +1 eklentisi yok sayılabilir. Böylece,

$$Q_1 = \frac{N}{4}, \text{ üncü değerdir.}$$

$$Q_2 = \frac{2N}{4}, \text{ üncü değerdir.}$$

$$Q_3 = \frac{3N}{4}, \text{ üncü değerdir.}$$

Gruplandırılmış serilerde kartilleri bulabilmek için medyan formülüne benzer formüllerden yararlanılır. Medyan formülünde gerekli düzenlemeler yapıldığında kartiller,



Kantiller bir seriyi dört (kartiller), on (desiller) ve yüz (pörsentiller) eşit parçaya bölen değerlerdir.



Büyüklik sırasına dizilmiş bir serinin dört eşit parçaya bölünmesi sonucu üç kartil bulunur.

$$Q_1 = L_1 + \frac{\frac{n}{4} - \sum f_{Q_1-1}}{f_{Q_1}} c$$

$$Q_2 = L_1 + \frac{\frac{2n}{4} - \sum f_{Q_2-1}}{f_{Q_2}} c$$

$$Q_3 = L_1 + \frac{\frac{3n}{4} - \sum f_{Q_3-1}}{f_{Q_3}} c$$

formülleri ile hesaplanır. Kantilleri gösteren değerler “...den az kümülatif frekans değerleri” içerisinde aranır [7].

Örnek 2: Bir kursa katılan öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Kursa katılan öğrencilerin kartil yaşlarını bulalım:

Gruplar	Frekans	... den az kümülatif frekans
6-8	5	5
8-10	15	20
10-12	55	75
12-24	25	100

Öncelikle birinci kartilin hangi sınıfta olduğu tespit edilir. Serinin $n/4 = 100/4 = 25'$ inci değeri Q_1 'dir. 25' inci değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 75'in gösterdiği sınıf Q_1 sınıfıdır. Bu serinin Q_1 değeri 10 ile 12 arasındadır. Q_1 sınıfının alt sınırı 10'dur. Q_1 sınıfından bir önceki sınıfın kümülatif frekansı $\sum f_{Q_1-1} = 20'$ dir. Q_1 sınıfının frekansı $f_{Q_1} = 55'$ tir. Q_1 sınıfının sınıf büyüklüğü, $c = 12-10 = 2'$ dir. Bilinen değerler birinci kartil formülünde yerine yazıldığında gruplandırılmış serinin Q_1 değeri,

$$Q_1 = 10 + \frac{\frac{100}{4} - 20}{55} 2 = 10.18$$

olarak elde edilir. İkinci kartil değeri medyana eşittir. Şimdi serinin üçüncü kartili hesaplanacaktır.

Öncelikle üçüncü kartilin hangi sınıfta olduğu tespit edilir. Serinin $3n/4 = 3(100)/4 = 75'$ inci değeri Q_3 'dür. 75' inci değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 75'in gösterdiği sınıf Q_3 sınıfıdır. Bu serinin Q_3 değeri 10 ile 12 arasındadır. Q_3 sınıfının alt sınırı 10'dur. Q_3 sınıfından bir önceki sınıfın kümülatif frekansı $\sum f_{Q_3-1} = 20'$ dir. Q_3 sınıfının frekansı $f_{Q_3} = 55'$ tir. Q_3 sınıfının sınıf büyüklüğü, $c = 12-10 = 2'$ dir. Bilinen değerler üçüncü kartil formülünde yerine yazıldığında gruplandırılmış serinin Q_3 değeri,

$$Q_3 = 10 + \frac{\frac{3(100)}{4} - 20}{55} \cdot 2 = 12$$

olarak elde edilir.

Desiller

Büyüklik sırasına dizilmiş bir serinin on eşit parçaya bölünmesi için dokuz bölen gerekir. En küçük desil birinci desil, en büyük desil dokuzuncu desildir. Birinci desil D_1 , ikinci desil D_2 , ..., dokuzuncu desil D_9 ile gösterilir. Her bir desil aralığı serideki rakamların yaklaşık %10'unu kapsar. *Beşinci desil aynı zamanda serinin medyanıdır.* Basit serilerde,

$$D_1 = \frac{N+1}{10}, \text{ uncu değerdir.}$$

$$D_2 = \frac{2(N+1)}{10}, \text{ uncu değerdir.}$$

...

...

$$D_9 = \frac{9(N+1)}{10}, \text{ uncu değerdir.}$$

Sınıflandırılmış ve gruplandırılmış serilerde çok sayıda değer yer aldığı için yukarıdaki ifadelerde yer alan +1 eklentisi yok sayılabilir. Böylece,

$$D_1 = L_1 + \frac{\frac{n}{10} - \sum f_{D_1-1}}{f_{D_1}} c$$

$$D_2 = L_1 + \frac{\frac{2n}{10} - \sum f_{D_2-1}}{f_{D_2}} c$$

$$D_9 = L_1 + \frac{\frac{9n}{10} - \sum f_{D_9-1}}{f_{D_9}} c$$

formülleri ile hesaplanır. Desilleri gösteren değerler "...den az kümülatif frekans değerleri" içerisinde aranır [8].

Örnek 1: Bir kursa katılan öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Kursa katılan öğrencilerin yaşlarının birinci ve dokuzuncu desillerini bulalım.

Gruplar	Frekans	 den az kümülatif frekans
6-8	5	5
8-10	15	20
10-12	55	75
12-24	25	100



Büyüklik sırasına dizilmiş bir serinin on eşit parçaya bölünmesi için dokuz bölen gerekir. En küçük desil birinci desil, en büyük desil dokuzuncu desildir

Öncelikle birinci desilin hangi sınıfta olduğu tespit edilir. Serinin $n/10 = 100/10 = 10'$ uncu değeri D_1 'dir. $10'$ uncu değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 20 'nin gösterdiği sınıf D_1 sınıfıdır. Bu serinin D_1 değeri 8 ile 10 arasındadır. D_1 sınıfının alt sınırı 8 'dir. D_1 sınıfından bir önceki sınıfın kümülatif frekansı $\sum f_{D_1-1} = 5$ 'tir. D_1 sınıfının frekansı $f_{D_1} = 15$ 'tir. D_1 sınıfının sınıf büyüklüğü, $c = 10 - 8 = 2$ 'dir. Bilinen değerler birinci desil formülünde yerine yazıldığında gruplandırılmış serinin D_1 değeri,

$$D_1 = 8 + \frac{\frac{100}{10} - 5}{15} 2 = 8.67$$

olarak elde edilir. Bu seride en küçük $\%10$ 'u oluşturan değerlerin en büyüğü 8.67 'dir. Şimdi serinin dokuzuncu desilini bulalım:

Öncelikle dokuzuncu desilin hangi sınıfta olduğu tespit edilir. Serinin $9n/10 = 9(100)/10 = 90'$ ıncı değeri D_9 'dur. $90'$ ıncı değeri, ilk kez içerisinde bulunduran kümülatif frekans değeri olan 100 'ün gösterdiği sınıf D_9 sınıfıdır. Bu serinin D_9 değeri 12 ile 14 arasındadır. D_9 sınıfının alt sınırı 12 'dir. D_9 sınıfından bir önceki sınıfın kümülatif frekansı $\sum f_{D_9-1} = 75$ 'tir. D_9 sınıfının frekansı $f_{D_9} = 25$ 'tir. D_9 sınıfının sınıf büyüklüğü, $c = 14 - 12 = 2$ 'dir. Bilinen değerler dokuzuncu desil formülünde yerine yazıldığında gruplandırılmış serinin D_9 değeri,

$$D_9 = 12 + \frac{\frac{9(100)}{10} - 75}{25} 2 = 13.2$$

olarak elde edilir. Bu seride en büyük $\%10$ 'u oluşturan değerlerin en küçüğü 13.2 'dir.

Pörsentiller

Büyüklik sırasına dizilmiş bir serinin yüz eşit parçaya bölünmesi için 99 bölen gerekir. *En küçük pörsentil birinci pörsentil, en büyük pörsentil 99' uncu pörsentildir.* Birinci pörsentil P_1 , ikinci pörsentil P_2 , ..., 99 uncu pörsentil P_{99} ile gösterilir. Her bir pörsentil aralığı serideki rakamların yaklaşık $\%1$ 'ini kapsar. *50'inci pörsentil aynı zamanda serinin medyanıdır* [11]. Basit serilerde,

$$P_1 = \frac{N+1}{100}, \text{ üncü değerdir.}$$

$$P_2 = \frac{2(N+1)}{100}, \text{ üncü değerdir.}$$

.

.

.

$$P_{99} = \frac{99(N+1)}{100}, \text{ üncü değerdir.}$$



Büyüklik sırasına dizilmiş bir serinin yüz eşit parçaya bölünmesi için 99 bölen gerekir. En küçük pörsentil birinci pörsentil, en büyük pörsentil 99' uncu pörsentildir.

Sınıflandırılmış ve gruplandırılmış serilerde çok sayıda değer yer aldığı için yukarıdaki ifadelerde yer alan +1 eklentisi yok sayılabilir. Böylece,

$$P_1 = L_1 + \frac{\frac{n}{100} - \sum f_{P_1-1}}{f_{P_1}} c$$

$$P_2 = L_1 + \frac{\frac{2n}{100} - \sum f_{P_2-1}}{f_{P_2}} c$$

·
·
·

$$P_{99} = L_1 + \frac{\frac{99n}{100} - \sum f_{P_{99}-1}}{f_{P_{99}}} c$$

formülleri ile hesaplanır. Pörsentilleri gösteren değerler "...den az kümülatif frekans değerleri" içerisinde aranır [12].

Örnek 2: Bir kursa katılan öğrencilerin yaşlarına göre dağılımı aşağıdaki gibidir. Kursa katılan öğrencilerin yaşlarının beşinci ve doksan beşinci pörsentillerini bulalım;

Gruplar	Frekans	 den az kümülatif frekans
6-8	5	5
8-10	15	20
10-12	55	75
12-24	25	100

Öncelikle beşinci pörsentilin hangi sınıfta olduğu tespit edilir.

Serinin $5n/100 = 5(100)/100 = 5'$ inci değeri P_5' tir. Beşinci değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 5'in gösterdiği sınıf P_5 sınıfıdır. Bu serinin P_5 değeri 6 ile 8 arasındadır. P_5 sınıfının alt sınırı 6'dır. P_5 sınıfından bir önceki sınıfın kümülatif frekansı $\sum f_{P_5-1} = 0$ 'dır. P_5 sınıfının frekansı $f_{P_5} = 5$ 'tir. P_5 sınıfının sınıf büyüklüğü, $c = 8-6 = 2$ 'dir. Bilinen değerler beşinci pörsentil formülünde yerine yazıldığında gruplandırılmış serinin P_5 değeri,

$$P_5 = 6 + \frac{\frac{5(100)}{100} - 0}{5} 2 = 8$$

olarak elde edilir. Bu seride en küçük %5'i oluşturan değerlerin en büyüğü 8'dir. Şimdi serinin 95' inci pörsentilini bulalım.

Öncelikle 95' inci pörsentilin hangi sınıfta olduğu tespit edilir. Serinin $95n/100 = 95(100)/100 = 95'$ inci değeri P_{95}' tir. Doksan beşinci değeri ilk kez içerisinde bulunduran kümülatif frekans değeri olan 95'in gösterdiği sınıf P_{95} sınıfıdır. Bu serinin P_{95} değeri 12 ile 14 arasındadır. P_{95} sınıfının alt sınırı 12'dir. P_{95}

sınıfından bir önceki sınıfın kümülatif frekansı $\sum f_{P_{95}-1} = 75$ 'tir. P_{95} sınıfının frekansı $f_{P_{95}} = 25$ 'tir. P_{95} sınıfının sınıf büyüklüğü, $c = 14 - 12 = 2$ 'dir. Bilinen değerler 957'nci pörsentil formülünde yerine yazıldığında gruplandırılmış serinin P_{95} değeri,

$$P_{95} = 12 + \frac{\frac{95(100)}{100} - 75}{25} 2 = 13.6$$

olarak elde edilir. Bu seride en büyük %5'i oluşturan değerlerin en küçüğü 13.6'dır.



Bireysel Etkinlik

- Basit bir seride mod nasıl bulunur?
- 4, 4, 5, 8, 8, 9, 9, 9, 9, 10 sayılarından oluşan serinin modu kaçtır?
- 15, 16, 18, 40, 50, 50, 50, 60, 80, 90 sayılarından oluşan seri için mod, medyan ve kartil hesaplaması yapınız.



Özet

- Parametrik merkezî eğilim ölçüleri serideki tüm verilerden elde edilirken, parametrik olmayan merkezî eğilim ölçüleri serideki bazı verilerden elde edilmektedir. Bu tür ortalamalar serideki tek bir rakamın değişiminden ve aşırı uçlardan direkt etkilenmezler. Parametrik olmayan merkezî eğilim ölçüleri; mod, medyan ve kantiller olmak üzere üç grupta sınıflandırılırlar. Ayrıca kantiller de kendi aralarında kartiller, desiller ve pörsentiller olmak üzere üç grupta incelenirler.
- Mod, en yaygın kullanılan, en çok beğenilen, en yaygın gözlenen anlamına gelir. Mod, incelenen bir seride en fazla tekrar eden ya da başka bir ifadeyle frekansı en yüksek olan gözlem değeridir.
- Bir seride birden fazla mod olabileceği gibi seride mod bulunmayabilir. Bu tür serilerde modun merkezî eğilim ölçüsü olarak değeri olmaz.
- Gruplandırılmış seride içerisinde en fazla değer bulunduran sınıf, mod sınıfıdır. Mod sınıfı, içerisinde modu bulunduran sınıftır. Mod grafikte de tespit edilebilir. Verilmiş olan grubun, histogramı üzerindeki frekans eksenine en büyük olan kutucuğun, üst köşe noktaları ile komşu kutucuklara değdiği en üst noktaların çapraz bir şekilde birleştirilmesi ile oluşan kesişim noktasının yatay eksenindeki karşılığı serinin modudur.
- Veriler büyüktür küçüğe veya küçüktür büyüğe doğru sıralandığında serinin tam ortasına karşılık gelen değere medyan ya da ortanca denir.
- Sınıflandırılmış seriden medyan bulunabilmesi için öncelikle ...den az kümülatif frekans değerlerinin bulunması gerekir. Medyanı gösteren $n/2$ inci değeri ilk kez içerisinde bulunduran kümülatif frekansa sahip olan değer medyan değeridir.
- Gruplandırılmış seriden de medyanın bulunabilmesi için öncelikle ...den az kümülatif frekans değerlerinin bulunması gerekir. Kümülatif frekans serisinde $n/2$ 'inci değeri ilk kez içerisinde bulunduran sınıf medyan sınıfıdır. Medyanı içerisinde bulunduran sınıf bulunduktan sonra tam olarak medyan değerini elde edebilmek için belirli bir formül kullanılır.
- Kantiller bir seriyi 4, 10 ve 100 eşit parçaya ayırarak bu serideki değerlerin dörtte, onda ve yüzde ne kadarının belirli bir değere göre yerini saptamak için kullanılır.
- Büyüklük sırasına dizilmiş bir serinin dört eşit parçaya bölünmesi sonucu üç kartil bulunur. Küçükten büyüğe doğru sıralanan seriyi dört parçaya bölebilmek için üç bölen gerekir. Birinci kartil Q1, ikinci kartil Q2 ve üçüncü kartil Q3 ile gösterilir. Her bir kartil aralığı yaklaşık serideki rakamların %25'ini kapsar. Bir serinin ikinci kartili medyanıdır.
- Gruplandırılmış serilerde kartilleri bulabilmek için medyan formülüne benzer formüllerden yararlanılır. Medyan formülünde gerekli düzenlemeler yapılarak kartiller hesaplanabilir.
- Büyüklük sırasına dizilmiş bir serinin on eşit parçaya bölünmesi için dokuz bölen gerekir. En küçük desil birinci desil, en büyük desil dokuzuncu desildir. Her bir desil aralığı serideki rakamların yaklaşık %10'unu kapsar. Beşinci desil aynı zamanda serinin medyanıdır.
- Büyüklük sırasına dizilmiş bir serinin yüz eşit parçaya bölünmesi için 99 bölen gerekir. En küçük pörsentil birinci pörsentil, en büyük pörsentil 99' uncu pörsentildir. Birinci pörsentil P1, ikinci pörsentil P2, ..., 99 uncu pörsentil P99 ile gösterilir. Her bir pörsentil aralığı serideki rakamların yaklaşık %1'ini kapsar. 50'inci pörsentil aynı zamanda serinin medyanıdır.

DEĞERLENDİRME SORULARI

1. Parametrik olmayan merkezî eğilim ölçüleri arasında aşağıdakilerden hangisi bulunmaz?
 - a) Mod
 - b) Medyan
 - c) Kareli ortalama
 - d) Kartiller
 - e) Pörsentiller

2. “3 5 7 2 4 8 5 7 9 7 8 2 7 6 7 3 4 6 7” serisinin modu aşağıdakilerden hangisidir?
 - a) 2
 - b) 3
 - c) 4
 - d) 5
 - e) 7

3. Yanda gruplandırılmış serisinin modu yaklaşık olarak aşağıdakilerden hangisidir?

Gruplar	Frekans
3-5	5
6-8	7
9-11	3
12-14	4

- a) 6
- b) 6.5
- c) 7
- d) 8
- e) 8.5

4. “3 5 7 2 4 8 5 7 9 7 8 2 7 6 7 3 4 6 7” serisinin medyanı kaçtır?
 - a) 2
 - b) 4
 - c) 5
 - d) 6
 - e) 7

5. “3 5 7 2 4 8 5 7 9 7 8 2 7 6 7 3 4 6 7 7” serisinin medyanı kaçtır?
 - a) 5
 - b) 5.5
 - c) 5
 - d) 6.5
 - e) 7

6. Yanda gruplandırılmış serinin medyanı kaçtır?

- a) 6.08
- b) 6.23
- c) 7.43
- d) 7.87
- e) 8.33

Gruplar	Frekans
3-5	5
6-8	7
9-11	3
12-14	4

7. Yanda gruplandırılmış serinin medyanı aşağıdakilerden hangisidir?

- a) 5.78
- b) 6.05
- c) 6.19
- d) 6.43
- e) 6.85

Gruplar	Frekans
3-5	5
5-7	7
7-9	3
9-11	5

8. Yanda gruplandırılmış serinin 1. kartili aşağıdakilerden hangisidir?

- a) 1.78
- b) 1.95
- c) 2.12
- d) 2.67
- e) 2.81

Gruplar	Frekans
1-3	6
3-5	3
5-7	7
7-9	4

9. Yukarıdaki gruplandırılmış serinin 3. kartili aşağıdakilerden hangisidir?

- a) 5.48
- b) 5.95
- c) 6.12
- d) 6.39
- e) 6.71

Gruplar	Frekans
1-3	6
3-5	3
5-7	7
7-9	4

10. Yanda gruplandırılmış serinin 9. desili aşağıdakilerden hangisidir?

- a) 7
- b) 7.2
- c) 7.7
- d) 8
- e) 8.3

Gruplar	Frekans
1-3	6
3-5	3
5-7	7
7-9	4

Cevap Anahtarı

1.c, 2.e, 3.b, 4.d, 5.d, 6.c, 7.d, 8.d, 9.e, 10.d

YARARLANILAN KAYNAKLAR

- [1] Çakır, F. (2000). Sosyal Bilimlerde İstatistik, İstanbul: Alfa Yayıncılık.
- [2] Başar, Alaaddin, Oktay, Erkan. (2018). Uygulamalı İstatistik-I: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap ve Kirtasiye.
- [3] Bülbül, E. Ş. (2007). Çözümsel İstatistik. İstanbul: Alfa yayıncılık.
- [4] Tekin, A. (2012). Temel İstatistik Dersleri. Konya: Eğitim Yayınevi.
- [5] Gürsakal, N., Oğuzlar, A, Gürsakal, S. (2022). Betimsel İstatistik (10. Baskı). Bursa: Dora Yayıncılık.
- [6] Köseoğlu M., Yamak, R. (2009). Uygulamalı İstatistik (4. baskı). Trabzon: Celepler Matbaacılık.
- [7] Newbold, P. (2009). Statistics for Business and Economics (7. baskı). İstanbul: Literatür Yayıncılık.
- [8] Özsoy, O. (2010). İktisatçılar ve İşletmeciler için İstatistik (3. baskı). Ankara: Siyasal Kitabevi.
- [9] Serper, Ö. (2014). Uygulamalı İstatistik-I (6.baskı). Bursa: Ezgi Kitabevi.
- [10] Johnson, A. R., Wichern, W. D. (1998). Applied Multivariate Statistical Analysis (4th edition). London: Prentice-Hall.
- [11] Yıldız, N. T. (2008). Uygulamalı İstatistik (3. baskı). Ankara: Nobel Yayıncılık.
- [12] Yılmaz, B. (2010). İstatistik (3. baskı). Ankara: Nobel Yayıncılık.

PARAMETRİK DEĞİŞKENLİK ÖLÇÜLERİ



İÇİNDEKİLER

- Ortalama Sapma
- Varyans
- Standart Sapma
- Değişim Katsayısı



HEDEFLER

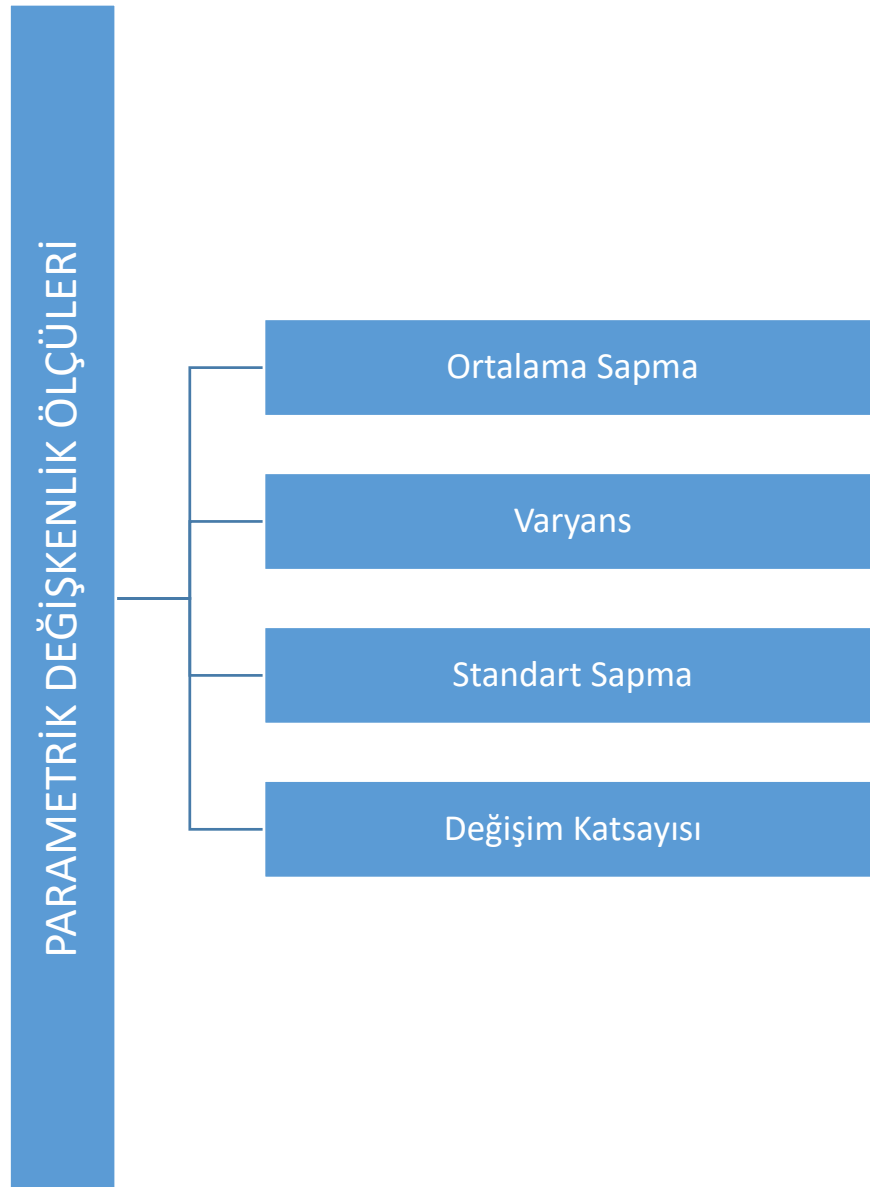
- Bu üniteyi çalıştıktan sonra;
 - Değişkenlik kavramını açıklayabilecek,
 - Serilerdeki değişkenliği ölçebilecek,
 - Aritmetik ortalamadan mutlak sapmaları ölçen değişkenlik ölçüsünü açıklayabilecek,
 - Aritmetik ortalamadan kareleri sapmaları ölçen değişkenlik ölçülerini açıklayabilecek,
 - Serilerden elde edilebilecek parametrik değişkenlik ölçüleri hakkında bilgi sahibi olabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Prof. Dr. Mehmet
Suphi ÖZÇOMAK

ÜNİTE
7





Merkezî eğilim etrafındaki dağılma durumunu ortaya koymak amacıyla değişkenlik ölçülerinden yararlanılır.

GİRİŞ

Merkezî eğilim ölçüleri, istatistiksel serilerin incelenmesi ve karşılaştırılmasında tek başlarına her zaman yeterli olmayabilirler. Bu nedenle istatistiksel serilerin karşılaştırılmasında birimlerin merkezî eğilim etrafındaki dağılma durumunu da ortaya koymak gerekmektedir. *Merkezî eğilim etrafındaki dağılma durumunu ortaya koymak amacıyla değişkenlik ölçülerinden yararlanılır* [1]. Bazı durumlarda istatistiksel serilerin ortalamaları birbirine eşit olsa da serilerin dağılımları birbirinden farklı olabilir. X ve Y iki farklı basit seriler olsun.

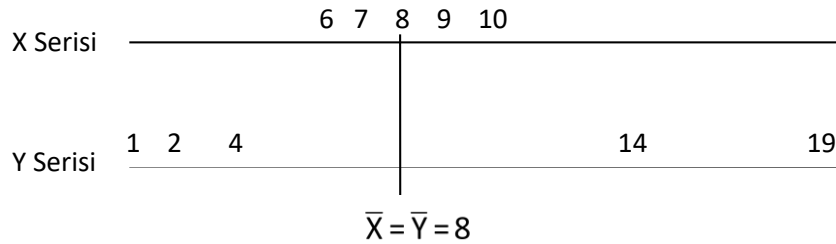
X	Y
6	1
7	2
8	4
9	14
10	19

Bu serilerin aritmetik ortalamaları sırasıyla;

$$\bar{X} = \frac{\sum X_i}{n} = \frac{40}{5} = 8$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{40}{5} = 8$$

şeklinde hesaplandığında görüldüğü üzere her iki serinin aritmetik ortalamaları birbirine eşit çıkmıştır. Ancak seriler incelendiğinde ortalamaları aynı olmasına rağmen farklı özellikler taşıdığı görülmektedir. X serisinde gözlemler birbirine yakın değerler alırken, Y serisinde gözlemler birbirlerinden uzak değerler almışlardır. Başka bir ifade ile X serisi homojen bir yapıya sahip yani gözlemler birbirine benzerken, Y serisi heterojen bir yapıya sahip olup gözlemler birbirine benzememektedir. Bu durum bir ölçek üzerinde gösterildiğinde konu daha iyi anlaşılacaktır:



Görsel 7.1. X ve Y serilerinin dağılımı

Görsel 7.1’de görüldüğü üzere X serisinde gözlem değerleri birbirine yakın iken, Y serisinde ise gözlem değerleri birbirlerinden uzaktır. Oysa her iki serinin ortalaması eşit çıkmıştır.

Yukarıda verilen örnekte serilerin gözlem değerleri yerine sadece ortalamaları verilerek karşılaştırma yapılması istenildiğinde serilerin birbirine benzer oldukları ya



İstatistiksel bir seriyi oluşturan gözlem değerlerinin değer itibarıyla birbirlerinden ya da herhangi bir ortalamadan uzaklıkları esas alınarak hesaplanan ölçülere “değişkenlik ölçüleri” denir.



Değişkenlik ölçüsü küçük olan seriler, değişkenlik ölçüsü büyük olan serilere göre ortalama etrafında daha sık bir dağılım gösterirler.

da aynı gözlem değerlerine sahip oldukları düşünülebilir. Bu durumda istatistiksel serileri karşılaştırmak için sadece merkezî eğilim ölçülerini hesaplamak yeterli olmamaktadır. Merkezî eğilim ölçülerinin yanında serideki birimlerin nasıl dağıldıklarını gösteren bir ölçüye de ihtiyaç duyulur. İstatistiksel serilerdeki gözlemlerin birbirine benzeyip benzemediklerini, birbirlerinden uzak olup olmadıklarını, başka bir ifade ile serilerin homojen mi yoksa heterojen mi olduklarını tespit etmek amacıyla değişkenlik ölçüleri kullanılmaktadır.

İstatistiksel bir seriyi oluşturan gözlem değerlerinin değer itibarıyla birbirlerinden ya da herhangi bir ortalamadan uzaklıkları esas alınarak hesaplanan ölçülere “değişkenlik ölçüleri” denir. Değişkenlik ölçüleri serideki gözlem değerlerinin dağılımlarının bir ölçüsüdür. Bu ölçüler serideki gözlem değerlerinin ortalama etrafında ne kadar sık dağıldıklarını belirtirler.

Değişkenlik ölçüleri tek bir seri için hesaplandığında tek başlarına bir anlam ifade etmezler. Bu ölçüler ancak birbirleri ile karşılaştırıldıklarında hesaplandıkları serilerin hangisinde birimlerin ortalama etrafında daha sık, hangisinde daha seyrek biçimde dağıldığını belirtirler. Diğer bir ifade ile *değişkenlik ölçüsü küçük olan seriler, değişkenlik ölçüsü büyük olan serilere göre ortalama etrafında daha sık bir dağılım gösterirler.* Bu nedenle değişkenlik ölçüleri, serideki gözlem değerlerinin homojenliği ile ilgilidirler. Değişkenlik ölçüsü küçük olan seriler karşılaştırma yapılan diğer serilere göre daha homojendirler. Gözlem değerleri aritmetik ortalama etrafında daha sık dağılan serilerde aritmetik ortalama seriyi daha iyi temsil etmektedir. Homojen serilerde ortalamanın seriyi temsil etme gücü daha yüksektir [2].

Bilindiği üzere anakütledeki bütün birimler kullanılarak hesaplanan ölçülere “parametre” adı verilmektedir. Bu nedenle bu ölçülere “parametrik değişkenlik ölçüleri” adı verilmiştir. Parametrik değişkenlik ölçüleri hesaplanırken serideki bütün birimler kullanılmaktadır.

Bu bölümde parametrik değişkenlik ölçülerinden ortalama sapma, varyans, standart sapma ve değişim katsayısı ele alınmıştır.

ORTALAMA SAPMA

Değişkenliğin hesaplanmasında kullanılan ölçülerden biri de *ortalama sapmadır. Ortalama sapma, değişkenliğin ölçülmesinde serideki gözlem değerlerinin aritmetik ortalamadan ne kadar uzak olduklarını belirlemeye çalışan değişkenlik ölçüsüdür.* Bu amaçla ortalama sapmanın hesaplanması için gözlem değerlerinin aritmetik ortalamadan farklarının toplamı elde edilir. Bu fark değerlerinin toplamı,

$$\sum_{i=1}^n (x_i - \bar{x})$$

şeklinde ifade edilir. Bilindiği üzere bir serideki rakamların bu serinin aritmetik ortalamasından farklarının toplamı her zaman sıfır değerine eşittir. Yukarıdaki toplamda, ortalamadan küçük gözlem değerleri negatif, büyük olan gözlem değerleri ise pozitif sonuç vereceğinden dolayı bu değerlerin toplamı sıfıra eşit çıkacaktır. Bu sorun, söz konusu farkların mutlak değerleri alınarak çözülebilir.

$$\sum_{i=1}^n |x_i - \bar{x}|$$

Yukarıdaki ifade bir seride yer alan değerlerin aritmetik ortalamadan toplam sapma miktarını göstermektedir. Bu toplam, gözlem sayısına bölünerek ortalama sapma değeri elde edilir. Basit serilerde ortalama sapma değeri,

$$O.S. = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

eşitliği yardımıyla hesaplanır [3]. Basit serilerde ortalama sapma değerinin nasıl hesaplanacağı aşağıdaki örnekte açıklanmıştır.

Örnek 1: Aşağıdaki basit serinin ortalama sapma değerini hesaplayınız.

x_i
7
8
10
11
14

Bu serinin aritmetik ortalaması, $\bar{x} = 10$ 'dur. Gözlem değerlerinin aritmetik ortalamadan farklarının mutlak değerleri aşağıdaki gibi elde edilerek toplanır:

x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $
7	-3	3
8	-2	2
10	0	0
11	1	1
14	4	4

Böylece ortalama sapma değeri,

$$O.S. = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} = \frac{10}{5} = 2$$

olarak elde edilir. Sınıflandırılmış ve gruplandırılmış serilerde ortalama sapma,

$$O.S. = \frac{\sum_{i=1}^n f_i |x_i - \bar{x}|}{\sum f_i}$$

formülüyle hesaplanır. Sınıflandırılmış ve gruplandırılmış serilerde ortalama sapma değerinin nasıl hesaplanacağı aşağıda verilen örneklerde açıklanmıştır.

Örnek 2: Aşağıdaki sınıflandırılmış serinin ortalama sapma değerini hesaplayınız.

X	f
10	4
20	6
30	9
40	10
50	8
60	8

Ortalama sapma değerinin hesaplanabilmesi için ilk olarak aritmetik ortalama hesaplanır. Bu serinin aritmetik ortalaması, $\bar{X} = 38$ 'dir.

X_i	f_i	$X_i - \bar{X}$	$ X_i - \bar{X} $	$f_i X_i - \bar{X} $
10	4	-28	28	112
20	6	-18	18	108
30	9	-8	8	72
40	10	2	2	20
50	8	12	12	96
60	8	22	22	176

Gözlem değerlerinin aritmetik ortalamadan farklarının mutlak değerleri elde edilip karşılarındaki frekans değerleri ile çarpılarak toplanır. Elde edilen toplam, toplam frekansa bölüldüğünde serinin ortalama sapma değeri,

$$O.S. = \frac{\sum_{i=1}^n f_i |X_i - \bar{X}|}{\sum f_i} = \frac{584}{45} = 12.98$$

olarak elde edilir.

Örnek 3: Aşağıdaki gruplandırılmış serinin ortalama sapma değerini hesaplayınız.

Gruplar	f
1 - 3	1
3 - 5	2
5 - 7	4
7 - 9	3

Ortalama sapma değerinin hesaplanabilmesi için ilk önce sınıf sınırları kullanılarak sınıf değerleri (X_i) elde edilir. Daha sonra yapılacak işlemler sınıflandırılmış seri için ortalama sapma değeri hesaplanırken yapılan işlemlerin aynısıdır. Bu serinin aritmetik ortalaması, $\bar{X} = 5.8$ 'dir.

X_i	f_i	$f_i X_i$	$X_i - \bar{X}$	$ X_i - \bar{X} $	$f_i X_i - \bar{X} $
2	1	2	-3.8	3.8	3.8
4	2	8	-1.8	1.8	3.6
6	4	24	0.2	0.2	0.8
8	3	24	2.2	2.2	6.6

Bu gruplandırılmış serinin ortalama sapma değeri,

$$O.S. = \frac{\sum_{i=1}^n f_i |X_i - \bar{X}|}{\sum f_i} = \frac{14.8}{10} = 1.48$$

olarak hesaplanır.

VARYANS

Ortalama sapma hesaplanırken, aritmetik ortalamanın $\sum_{i=1}^n (X_i - \bar{X}) = 0$ özelliğinden yararlanılmıştır. Fark değerlerinin mutlak değerleri toplamı, $\sum_{i=1}^n |X_i - \bar{X}|$ bulunarak bu değerlerin ortalaması bulunmuştur. Gözlem değerlerinin aritmetik ortalamadan farklarının toplamının sıfır çıkmaması için gözlem değerlerinin aritmetik ortalamadan farklarının kareleri de hesaplanabilir.



Seride bulunan gözlem değerlerinin aritmetik ortalamadan farklarının kareleri alınarak toplanırsa $\sum_{i=1}^n (X_i - \bar{X})^2$ fark kare toplamı elde edilir, hesaplanır.

Seride bulunan gözlem değerlerinin aritmetik ortalamadan farklarının kareleri alınarak toplanırsa $\sum_{i=1}^n (X_i - \bar{X})^2$ fark kare toplamı elde edilir fark kare toplamı serbestlik derecesine bölünerek varyans değeri hesaplanır. Serbestlik derecesi toplam gözlem sayısının bir eksiğidir.

Varyans değeri hesaplanırken ortalamadan farkların karesi alınarak fark toplamının sıfıra eşit çıkma sorunu giderilmiştir. Böylece ortalama sapmanın matematik işlemlere elverişli olmama dezavantajı da bu değişkenlik ölçüsünde ortadan kaldırılmıştır. Örnek değerleri kullanılarak hesaplanan varyans değeri s^2 ile ifade edilirken, anakütledeki tüm değerler kullanılarak hesaplanan varyans değeri ise σ^2 ile ifade edilir. Basit serilerde varyans değeri,

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

formülü ile hesaplanır. Görüldüğü üzere fark kare toplamı (n-1) serbestlik derecesine bölünerek varyans değeri elde edilmektedir [4]. Basit serilerde varyans değerinin nasıl hesaplanacağı aşağıdaki örnekte açıklanmıştır.

Örnek 4: Aşağıdaki basit serinin varyans değerini hesaplayınız.

X_i
20
22
28
30

Varyans değerinin hesaplanması için ilk önce aritmetik ortalama hesaplanır. Bu serinin aritmetik ortalaması, $\bar{X} = 25$ 'tir.

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
20	-5	25
22	-3	9
28	3	9
30	5	25

Aritmetik ortalama hesaplandıktan sonra gözlem değerlerinin aritmetik ortalamadan farklarının kareleri elde edilerek toplanır. Elde edilen toplam serbestlik derecesi olan $4 - 1 = 3$ 'e bölüldüğünde varyans,

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{68}{4-1} = 22.67$$

olarak elde edilir.

Sınıflandırılmış ve gruplandırılmış serilerde varyans değeri aşağıdaki formül yardımıyla hesaplanır:

$$s^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum f_i}$$

bu formül düzenlendiğinde,

$$s^2 = \frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2$$

formülü elde edilir. Sınıflandırılmış ve gruplandırılmış serilerde varyans değerinin nasıl hesaplanacağı aşağıdaki örneklerle açıklanmıştır.

Örnek 5: Aşağıdaki sınıflandırılmış serinin varyansını hesaplayınız.

X	f
1	3
3	5
5	8
7	2

Bu serinin aritmetik ortalaması, $\bar{X} = 4$ 'tür. Gözlem değerlerinin aritmetik ortalamadan farklarının kareleri elde edildikten sonra karşılardaki frekanslar ile çarpılır.

X	f	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
1	3	-3	9	27
3	5	-1	1	5
5	8	1	1	8
7	2	3	9	18

$\sum_{i=1}^n f_i (X_i - \bar{X})^2 = 58$ değeri, toplam frekans değerine ($\sum f_i = 18$) bölünerek varyans değeri,

$$s^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum f_i} = \frac{58}{18} = 3.22$$

olarak elde edilir. Varyans değerini veren diğer formül kullanılarak aşağıdaki gibi hesaplanabilir.

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
1	3	3	1	3
3	5	15	9	45
5	8	40	25	200
7	2	14	49	98

Varyans değeri,

$$s^2 = \frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2 = \frac{346}{18} - \left(\frac{72}{18} \right)^2 = 3.22$$

olarak elde edilir. Görüldüğü üzere elde edilen sonuç bir önceki varyans formülünde bulunan değer ile aynıdır.

Örnek 6: Aşağıdaki gruplandırılmış serinin varyans değerini hesaplayınız.

Gruplar	f
1 - 3	2
3 - 5	3
5 - 7	5
7 - 9	7

Gruplandırılmış seride varyans değerinin hesaplanabilmesi için gerekli hesaplamalar aşağıda verilmiştir. İlk olarak sınıf sınırları kullanılarak sınıf değerleri (X_i) elde edilir. Daha sonra yapılan işlemler sınıflandırılmış serideki varyans hesabı ile aynıdır. Bu serinin aritmetik ortalaması, $\bar{X} = 6$ 'dır. Varyans değerini elde etmek için gerekli işlemler aşağıdaki gibidir.

X_i	f_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
2	2	-4	16	32
4	3	-2	4	12
6	5	0	0	0
8	7	2	4	28

Böylece varyans değeri ,

$$s^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum f_i} = \frac{72}{17} = 4.24$$

olarak elde edilir. Varyans değeri diğer formül yardımıyla aşağıdaki gibi hesaplanır:



Standart sapma, serideki gözlem değerlerinin aritmetik ortalamadan sapmalarını bir başka ifade ile ortalamadan uzaklıklarını ifade eden değişkenlik ölçüsüdür.

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
2	2	4	4	8
4	3	12	16	48
6	5	30	36	180
8	7	56	64	448

Varyans değeri,

$$s^2 = \frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2 = \frac{684}{17} - \left(\frac{102}{17} \right)^2 = 4.24$$

olarak elde edilir. Görüldüğü üzere elde edilen sonuç bir önceki varyans formülünde bulunan değer ile aynıdır.

STANDART SAPMA

Serideki gözlem değerlerinin aritmetik ortalamadan farklarının kareli ortalamasına standart sapma denilir. Standart sapma, serideki gözlem değerlerinin aritmetik ortalamadan sapmalarını bir başka ifade ile ortalamadan uzaklıklarını ifade eden değişkenlik ölçüsüdür.



Bir serideki gözlem değerleri için hesaplanan standart sapma değeri küçük olduğunda gözlem değerlerinin aritmetik ortalamaya daha yakın olduklarını, aksi durumda ise uzak olduklarını ifade etmektedir.

Bir serideki gözlem değerleri için hesaplanan varyans değerinin karekökü alındığında standart sapma elde edilir. Standart sapma, istatistiki uygulamalarda en çok kullanılan değişkenlik ölçüsüdür. Standart sapma ölçüm biriminden bağımsız değildir. Yani anakütle veya örnekteki gözlem değerleri hangi ölçekle ölçülmüşse standart sapma da o ölçekle ölçülür. Örneğin, anakütledeki gözlem değerleri cm ile ifade edilmiş ise anakütlenin standart sapması da cm ile ifade edilir. Uygulamada genellikle örnek standart sapması “s”, anakütle standart sapması ise “σ” ile gösterilir.

Bir serideki gözlem değerleri için hesaplanan standart sapma değeri küçük olduğunda gözlem değerlerinin aritmetik ortalamaya daha yakın olduklarını aksi durumda ise uzak olduklarını ifade etmektedir. Bu durum en az iki istatistiksel seri karşılaştırıldığında daha iyi anlaşılabilir. Örneğin; X serisinin standart sapma değerinin 1.79, Y serisinin standart sapma değerinin ise 2.06 olarak elde edildiğini farz edelim. Bu durumda iki seri değişkenlik bakımından karşılaştırılmak istenildiğinde Y serisindeki değişkenliğin X serisine göre daha fazla olduğu söylenir.



Bir serinin varyans değeri biliniyorsa, bu değer karekökü alınarak standart sapma değeri elde edilir.

Standart sapma hesaplanırken ortalamadan sapmaların karesi alınarak fark toplamalarının sıfıra eşit çıkma sorunu giderilir. *Standart sapma değeri varyans değerinin kareköküne eşittir. Dolayısıyla bir serinin varyans değeri biliniyorsa, bu değer karekökü alınarak standart sapma değeri elde edilir. Benzer şekilde bir serinin standart sapma değeri biliniyorsa bu değer karesi alınarak, varyans değeri elde edilir.* Basit serilerde standart sapma,



Bir serinin standart sapma değeri biliniyorsa bu değerin karesi alınarak, varyans değeri elde edilir.

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

formülü ile hesaplanır [5]. Basit serilerde standart sapmanın nasıl hesaplanacağı aşağıdaki örnekte açıklanmıştır.

Örnek 7: Aşağıdaki basit serinin standart sapma değerini hesaplayınız.

X_i
2
4
8
10

Standart sapma değerinin hesaplanması için ilk önce aritmetik ortalama hesaplanır. Bu serinin aritmetik ortalaması, $\bar{X} = 6$ 'dır.

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
2	-4	16
4	-2	4
8	2	4
10	4	16

Aritmetik ortalama hesaplandıktan sonra gözlem değerlerinin aritmetik ortalamadan farklarının kareleri elde edilerek toplanır. Elde edilen toplam serbestlik derecesi olan $4 - 1 = 3$ 'e bölünüp karekökü alındığında standart sapma,

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{40}{4-1}} = \sqrt{13.33} = 3.65$$

olarak elde edilir. Sınıflandırılmış ve gruplandırılmış serilerde standart sapma değeri aşağıdaki formül yardımıyla hesaplanır:

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum f_i}}$$

bu formül düzenlendiğinde,

$$s = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2}$$

formülü elde edilir. Sınıflandırılmış ve gruplandırılmış serilerde standart sapma değerinin nasıl hesaplanacağı aşağıdaki örnekte açıklanmıştır.

Örnek 8: Aşağıdaki sınıflandırılmış serinin standart sapma değerini hesaplayınız.

X	f
2	3
4	5
6	8
8	2

Bu serinin aritmetik ortalaması, $\bar{X} = 5$ 'tir. Gözlem değerlerinin aritmetik ortalamadan farklarının kareleri elde edildikten sonra karşılardaki frekanslar ile çarpılır.

X	f	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^2$
2	3	-3	9	27
4	5	-1	1	5
6	8	1	1	8
8	2	3	9	18

$\sum_{i=1}^n f_i(X_i - \bar{X})^2 = 58$ değeri, toplam frekans değerine ($\sum f_i = 18$) bölünüp karekökü alındığında standart sapma değeri,

$$s = \sqrt{\frac{\sum_{i=1}^n f_i(X_i - \bar{X})^2}{\sum f_i}} = \sqrt{\frac{58}{18}} = \sqrt{3.22} = 1.79$$

olarak elde edilir. Standart sapma, diğer formül kullanılarak aşağıdaki gibi hesaplanabilir.

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
2	3	6	4	12
4	5	20	16	80
6	8	48	36	288
8	2	16	64	128

Standart sapma değeri,

$$s = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2} = \sqrt{\frac{508}{18} - \left(\frac{90}{18} \right)^2} = \sqrt{28.22 - 25} = \sqrt{3.22} = 1.79$$

olarak elde edilir. Bu formülle hesaplanan standart sapma değeri ile diğer formülle elde edilen değer aynı sonucu vermektedir

Örnek 9: Örnek 7.6 da verilen aşağıdaki gruplandırılmış serinin standart sapmasını hesaplayınız.

Gruplar	f
1 - 3	2
3 - 5	3
5 - 7	5
7 - 9	7

Bu örnekteki standart sapma değeri hesaplandığında varyans ile standart sapma arasındaki ilişki net olarak görülerek bir kıyaslama imkanı verecektir. Gruplandırılmış seride standart sapma değerinin hesaplanabilmesi için gerekli hesaplamalar aşağıda verilmiştir. İlk olarak sınıf sınırları kullanılarak sınıf değerleri (X_i) elde edilir. Daha sonra yapılan işlemler sınıflandırılmış serideki standart sapma hesabı ile aynıdır. Bu serinin aritmetik ortalaması, $\bar{X} = 6$ 'dır. Standart sapmayı elde etmek için gerekli işlemler aşağıdaki gibidir.

X_i	f_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
2	2	-4	16	32
4	3	-2	4	12
6	5	0	0	0
8	7	2	4	28

Böylece standart sapma değeri ,

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum f_i}} = \sqrt{\frac{72}{17}} = \sqrt{4.24} = 2.06$$

olarak elde edilir. Varyans bölümünde Örnek 7.6 da verilen aynı değerler için hesaplanan varyans değeri 4.24 olarak elde edilmişti. Bu değer karekökü alındığında standart sapma değeri 2.06 olarak elde edilir. Böylece bir serinin varyans değeri biliniyorsa, bu değer karekökü alınarak standart sapma değerinin nasıl elde edildiği bu örnekte net bir şekilde görülmektedir.

Standart sapma değeri diğer formül yardımıyla aşağıdaki gibi hesaplanır:

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
2	2	4	4	8
4	3	12	16	48
6	5	30	36	180
8	7	56	64	448

Standart sapma değeri,

$$s = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2} = \sqrt{\frac{684}{17} - \left(\frac{102}{17} \right)^2} = \sqrt{4.24} = 2.06$$

olarak elde edilir. Görüldüğü üzere bir önceki standart sapma formülünde bulunan değerle aynı sonuç elde edilmiştir.

Örnek 10: Bir serinin varyansı 16.24 olarak elde edilmiştir. Bu serinin standart sapma değerini hesaplayınız.

Standart sapma değeri varyans değerinin kareköküne eşittir. Dolayısıyla standart sapma değeri

$$s = \sqrt{s^2} = \sqrt{16.24} = 4.03$$

olarak elde edilir.

DEĞİŞİM KATSAYISI

Gerek standart sapma gerekse diğer değişkenlik ölçüleri, ölçü biriminden bağımsız değildirler. Bundan dolayı aynı seri farklı ölçü birimleriyle (mesela, kilogram yerine gram ile) ifade edildiğinde değişik standart sapma değerleri elde edileceği gibi, farklı ölçü birimleriyle ifade edilmiş iki ayrı serinin karşılaştırılması da yanıltıcı sonuçlar verecektir.

Örneğin, aşağıdaki X ve Y serilerini değişkenlik bakımından karşılaştıralım:

X	Y
3	3000
5	5000
7	7000
9	9000

Her iki serinin aritmetik ortalama ve standart sapmaları hesaplandığında,

$$\begin{aligned}\bar{X} &= 6 & \bar{Y} &= 6000 \\ s_x &= 2.582 & s_y &= 2582\end{aligned}$$

olarak elde edilir. Her iki seri karşılaştırıldığında Y serisinin standart sapma değeri X serisinin standart sapma değerinden büyük olduğu için Y serisinde değişkenliğin X serisinin değişkenliğinden daha fazla olduğu ifade edilir. Oysa X serisi kilogram cinsinden ve Y serisi ise gram cinsinden ifade edilirse gerçekte her iki serinin değişkenliğinin aynı olduğu görülecektir. İşte diğer değişkenlik ölçülerinin bu dezavantajını gidermek amacıyla değişim katsayısı geliştirilmiştir.

Değişim katsayısı hesaplanırken mutlak dağılma yerine nispi dağılma esas alınmıştır. Değişim katsayısı yüzde olarak ifade edildiğinden dolayı ölçü biriminden bağımsızdır. Yani değişim katsayısında, standart sapma aritmetik ortalamanın yüzdesi olarak ifade edilir. Değişim katsayısı,

$$D.K. = \frac{s}{\bar{X}} 100$$

formülü ile hesaplanır [1]. Yukarıda verilen örneğin değişim katsayılarını hesaplayarak tekrar karşılaştıralım. X serisi için değişim katsayısı,



Gerek standart sapma gerekse diğer değişkenlik ölçüleri, ölçü biriminden bağımsız değildirler.



Değişim katsayısı yüzde olarak ifade edildiğinden dolayı ölçü biriminden bağımsızdır.

$$D.K. = \frac{s}{\bar{X}} 100 = \frac{2.582}{6} 100 = \%43.03$$

Y serisi için değişim katsayısı,

$$D.K. = \frac{s}{\bar{X}} 100 = \frac{2582}{6000} 100 = \%43.03$$

olarak elde edilir. Görüldüğü gibi her iki serinin değişim katsayıları aynı değere sahiptir. Değişim katsayısının nasıl hesaplanacağı aşağıdaki örnekle açıklanmıştır.

Örnek 11: Aşağıdaki sınıflandırılmış serinin değişim katsayısını hesaplayınız.

X	f
20	3
40	5
60	8
80	2

Bu serinin aritmetik ortalaması, $\bar{X} = 50$ 'dir. Serinin standart sapma değeri ise;

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
20	3	60	400	1200
40	5	200	1600	8000
60	8	480	3600	28800
80	2	160	6400	12800

$$s = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2} = \sqrt{\frac{50800}{18} - \left(\frac{900}{18} \right)^2} = \sqrt{322.22} = 17.95$$

olarak elde edilir. Aritmetik ortalama ve standart sapma hesaplandıktan sonra değişim katsayısı,

$$D.K. = \frac{s}{\bar{X}} 100 = \frac{17.95}{50} 100 = \%35.9$$

olarak elde edilir.

Örnek 12: Aritmetik ortalaması 8 ve standart sapması 3.09 olan serinin değişim katsayısını hesaplayınız. Değişim katsayısı,

$$D.K. = \frac{3.09}{8} 100 = \%38.6$$

olarak elde edilir.

Örnek 13: Aşağıdaki X ve Y serilerinin değişim katsayılarını hesaplayarak değişkenlik bakımından karşılaştırınız.

$$\bar{X} = 24$$

$$\bar{Y} = 24$$

$$s_x = 3.32$$

$$s_y = 1.28$$

X serisi için değişim katsayısı,

$$D.K. = \frac{s}{\bar{X}} 100 = \frac{3.32}{24} 100 = \%13.83$$

Y serisi için değişim katsayısı,

$$D.K. = \frac{s}{\bar{X}} 100 = \frac{1.28}{24} 100 = \%5.33$$

olarak elde edilir. Her iki seri değişim katsayıları bakımından karşılaştırıldığında X serisinin değişim katsayısı, Y serisinin değişim katsayısından daha büyük olduğu için X serisinin değişkenliğinin Y serisinin değişkenliğinden daha fazla olduğuna karar verilir.

**Bireysel Etkinlik**

- Seriler arasında değişkenlik karşılaştırmaları yapmada başvurulabilecek en iyi değişkenlik ölçüsü hangisidir? Tartışınız.
- Birer örnek yardımıyla basit, sınıflandırılmış ve gruplandırılmış serilerden elde edilebilecek parametrik değişkenlik ölçülerini hesaplayınız.



Özet

- Merkezî eğilim ölçüleri, istatistiksel serilerin incelenmesi ve karşılaştırılmasında tek başlarına her zaman yeterli olmayabilirler. Bu nedenle istatistiksel serilerin karşılaştırılmasında birimlerin merkezî eğilim etrafındaki dağılıma durumunu da ortaya koymak gerekmektedir. Merkezî eğilim etrafındaki dağılıma durumunu ortaya koymak amacıyla değişkenlik ölçülerinden yararlanılır.
- İstatistiksel serilerdeki gözlemlerin birbirine benzeyip benzemediklerini, birbirlerinden uzak olup olmadıklarını başka bir ifade ile serilerin homojen mi yoksa heterojen mi olduklarını tespit etmek amacıyla değişkenlik ölçüleri kullanılmaktadır.
- İstatistiksel bir seriyi oluşturan gözlem değerlerinin değer itibarıyla birbirlerinden ya da herhangi bir ortalama uzaklıkları esas alınarak hesaplanan ölçülere “değişkenlik ölçüleri” denir. Değişkenlik ölçüleri serideki gözlem değerlerinin dağılımlarının bir ölçüsüdür. Bu ölçüler serideki gözlem değerlerinin ortalama etrafında ne kadar sık dağıldıklarını belirtirler.
- Değişkenlik ölçüleri tek bir seri için hesaplandığında tek başlarına bir anlam ifade etmezler. Bu ölçüler ancak birbirleri ile karşılaştırıldıklarında hesaplandıkları serilerin hangisinde birimlerin ortalama etrafında daha sık, hangisinde daha seyrek biçimde dağıldığını belirtirler. Diğer bir ifade ile değişkenlik ölçüsü küçük olan seriler, değişkenlik ölçüsü büyük olan serilere göre ortalama etrafında daha sık bir dağılım gösterirler. Bu nedenle değişkenlik ölçüleri serideki gözlem değerlerinin homojenliği ile ilgilidirler. Değişkenlik ölçüsü küçük olan seriler karşılaştırma yapılan diğer serilere göre daha homojendirler. Gözlem değerleri ortalama etrafında daha sık dağılan bir başka bir ifade ile homojen olan serilerde, ortalamanın seriyi temsil etme gücü daha yüksektir.
- Bilindiği üzere anakütledeki bütün birimler kullanılarak hesaplanan ölçülere “parametre” adı verilmektedir. Bu nedenle bu ölçülere “parametrik değişkenlik ölçüleri” adı verilmiştir. Parametrik değişkenlik ölçüleri hesaplanırken serideki bütün birimler kullanılmaktadır.
- Değişkenliğin hesaplanmasında kullanılan ölçülerden biri de ortalama sapmadır. Ortalama sapma, değişkenliğin ölçülmesinde serideki gözlem değerlerinin aritmetik ortalama ne kadar uzak olduklarını belirlemeye çalışan değişkenlik ölçüsüdür. Bu amaçla ortalama sapmanın hesaplanması için gözlem değerlerinin aritmetik ortalama farklarının toplamı elde edilir.
- Bilindiği üzere bir serideki rakamların bu serinin aritmetik ortalamasından farklarının toplamı her zaman sıfır değerine eşittir. Fark değerlerinin toplamında, ortalama küçük gözlem değerleri negatif, büyük olan gözlem değerleri ise pozitif sonuç vereceğinden dolayı bu değerlerin toplamı sıfıra eşit çıkacaktır. Bu sorun, söz konusu farkların mutlak değerleri alınarak çözülebilir.
- Ortalama sapma hesaplanırken, aritmetik ortalamanın $\sum_{i=1}^n (X_i - \bar{X}) = 0$ özelliğinden yararlanılmıştır. Fark değerlerinin mutlak değerleri toplamı, $\sum_{i=1}^n |X_i - \bar{X}|$ bulunarak bu değerlerin ortalaması bulunmuştur. Gözlem değerlerinin aritmetik ortalama farklarının toplamının sıfır çıkmaması için gözlem değerlerinin aritmetik ortalama farklarının kareleri de hesaplanabilir.
- Seride bulunan gözlem değerlerinin aritmetik ortalama farklarının kareleri alınarak toplanırsa $\sum_{i=1}^n (X_i - \bar{X})^2$ en küçük fark kare toplamı elde edilir. En küçük fark kare toplamı serbestlik derecesine bölünerek varyans değeri hesaplanır. Serbestlik derecesi toplam gözlem sayısının bir eksiğidir.



Özet

- Örnek değerleri kullanılarak hesaplanan varyans değeri s^2 ile ifade edilirken anakütledeki tüm değerler kullanılarak hesaplanan varyans değeri ise σ^2 ile ifade edilir.
- Serideki gözlem değerlerinin aritmetik ortalamadan farklarının kareli ortalamasına standart sapma denilir. Standart sapma serideki gözlem değerlerinin aritmetik ortalamadan sapmalarını bir başka ifade ile ortalamadan uzaklıklarını ifade eden değişkenlik ölçüsüdür.
- Bir serideki gözlem değerleri için hesaplanan varyans değerinin karekökü alındığında standart sapma elde edilir. Standart sapma, istatistiki uygulamalarda en çok kullanılan değişkenlik ölçüsüdür. Uygulamada genellikle örnek standart sapması " s ", anakütle standart sapması ise " σ " ile gösterilir.
- Standart sapma hesaplanırken ortalamadan sapmaların karesi alınarak fark toplamalarının sıfıra eşit çıkma sorunu giderilir. Standart sapma değeri varyans değerinin kareköküne eşittir. Dolayısıyla bir serinin varyans değeri biliniyorsa, bu değerin karekökü alınarak standart sapma değeri elde edilir. Benzer şekilde bir serinin standart sapma değeri biliniyorsa, bu değerin karesi alınarak varyans değeri elde edilir.
- Gerek standart sapma gerekse diğer değişkenlik ölçüleri, ölçü biriminden bağımsız değildirler. Bundan dolayı aynı seri farklı ölçü birimleriyle (mesela 4 kg yerine gram ile) ifade edildiğinde değişik standart sapma değerleri elde edileceği gibi, farklı ölçü birimleriyle ifade edilmiş iki ayrı serinin karşılaştırılması da yanıltıcı sonuçlar verecektir. Diğer değişkenlik ölçülerinin dezavantajını gidermek amacıyla değişim katsayısı geliştirilmiştir.
- Değişim katsayısı hesaplanırken mutlak dağılma yerine nispi dağılma esas alınmıştır. Değişim katsayısı yüzde olarak ifade edildiğinden dolayı ölçü biriminden bağımsızdır. Yani değişim katsayısında, standart sapma aritmetik ortalamanın yüzdesi olarak ifade edilir.

DEĞERLENDİRME SORULARI

1. Parametrik değişkenlik ölçüleri hesaplanırken hangi merkezî eğilim ölçüsü kullanılır?
 - a) Harmonik ortalama
 - b) Aritmetik ortalama
 - c) Geometrik ortalama
 - d) Kareli ortalama
 - e) Medyan
2. Beş değere sahip basit seride ortalama sapma 8.25 olarak bulunmuştur. Buna göre bu serinin aritmetik ortalamadan mutlak sapmaların toplamı kaçtır?
 - a) 33.37
 - b) 35.12
 - c) 38.11
 - d) 41.25
 - e) 44.35
3. Beş değere sahip basit seride aritmetik ortalamadan farklarının kareleri toplamı 11.14 olarak bulunmuştur. Bu serinin varyansı kaçtır?
 - a) 2.79
 - b) 2.99
 - c) 3.22
 - d) 3.45
 - e) 3.88
4. Beş değere sahip basit seride aritmetik ortalamadan farklarının kareleri toplamı 11.14 olarak bulunmuştur. Bu serinin standart sapması kaçtır?
 - a) 1.03
 - b) 1.33
 - c) 1.67
 - d) 1.89
 - e) 2.22
5. Bir seride aritmetik ortalama 15.4 ve varyans 5.3 olarak elde edilmiştir. Bu serinin değişim katsayısı kaçtır?
 - a) 14.94
 - b) 20.25
 - c) 22.04
 - d) 25.55
 - e) 28.28

6. Bir seride aritmetik ortalama 5.4 ve standart sapma 1.3 olarak elde edilmiştir. Bu serinin değişim katsayısı kaçtır?
- 16.11
 - 18.88
 - 21.23
 - 24.07
 - 27.84
7. Bir basit seride: 5, 6, 8, 9, 12 değerleri yer almaktadır. Serinin ortalama sapması kaçtır?
- 2
 - 4
 - 8
 - 9
 - 10
8. Beş değerden oluşan basit bir seride $\sum |X_i - \bar{X}| = 4.24$ olarak hesaplandığına göre ortalama sapma değeri kaçtır?
- 0.85
 - 1.96
 - 1.89
 - 2.54
 - 3.22
9. Bir basit seride 2, 3, 5, 7, 8 değerleri yer almaktadır. Serinin varyansı kaçtır?
- 2.9
 - 3.4
 - 4.8
 - 5.2
 - 6.5
10. Bir sınıfta okuyan 6 öğrencinin not ortalaması $\bar{X} = 83$ ve $\sum (X_i - \bar{X})^2 = 318.4$ olarak elde edildiğine göre varyans değeri kaçtır?
- 33.24
 - 43.55
 - 55.28
 - 66.21
 - 63.68

Cevap Anahtarı

1.b, 2.d, 3.a, 4.c, 5.a, 6.d, 7.a, 8.a, 9.e, 10.e

YARARLANILAN KAYNAKLAR

- [1] Başar, Alaaddin, Oktay, Erkan. (2018). Uygulamalı İstatistik – I: Kısa Teorik Bilgiler ve Çözülmüş Problemler, EKEV Yayınları, 12. Baskı, Erzurum.
- [2] Gürtan, Kenan (1982). İstatistik ve Araştırma Metodları, İstanbul Üniversitesi Yayınları, İstanbul.
- [3] Serper, Ö. (2014). Uygulamalı İstatistik I, (7. Baskı), Filiz Kitabevi, İstanbul.
- [4] Daniel, Wayne W., James C. Terrell (1995). Business Statistics: For Management and Economics (7. Baskı), Houghton Mifflin Company, Boston.
- [5] Armutlulu, İ. H.(2018). İşletmelerde Uygulamalı İstatistik, 3. Baskı, Alfa Yayınları, İstanbul.

PARAMETRİK OLMAYAN DEĞİŞKENLİK ÖLÇÜLERİ



İÇİNDEKİLER

- Değişim Aralığı
- Kartil Aralığı
- Desil Aralığı
- Pörsentil Aralığı



HEDEFLER

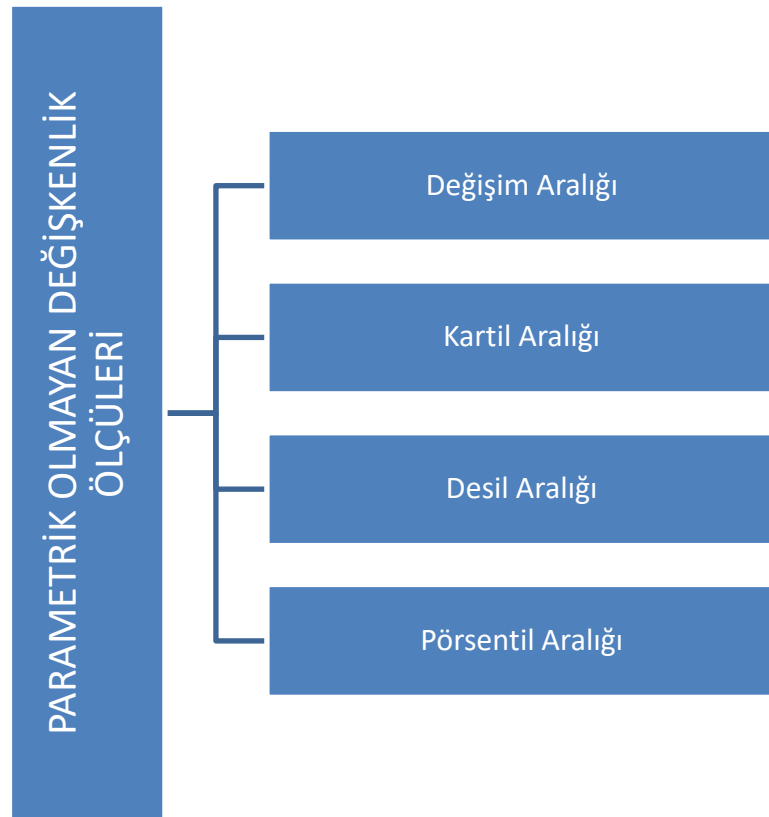
- Bu üniteyi çalıştıktan sonra;
- Parametrik olmayan değişkenlik kavramını açıklayabilecek,
- Parametrik olmayan değişkenlik ölçülerinin avantaj ve dezavantajlarını açıklayabilecek,
- Parametrik olmayan değişkenlik ölçülerini hesaplayıp yorumlayabilecek,
- Parametrik olmayan değişkenlik ölçülerini birbirleriyle karşılaştırabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Prof. Dr. Mehmet
Suphi ÖZÇOMAK

ÜNİTE
8



GİRİŞ

İstatistiki araştırmaya konu olan veriler nitel değişkenlerden elde edilen veriler bir başka ifade ile kategorik verilerle çalıştığımızda verilerin dağılımındaki değişkenliği nasıl izleyeceğiz? Bu konudaki sorularımıza cevap aramak için parametrik olmayan değişkenlik ölçüleri incelenebilir.

Kategorik verilerin değişkenliğini ölçmede parametrik değişkenlik ölçüleri bazen mantıklı olmayan sonuçlar doğurabilir. Parametrik değişkenlik ölçülerinin kullanılmadığı durumlarda parametrik olmayan değişkenlik ölçülerinden yararlanılır. Parametrik olmayan değişkenlik ölçülerinin elde edilmesi kolaydır. Serideki tüm değerlere tabi olmadığı için hesaplanmaları pratiktir. Bununla birlikte parametrik değişkenlik ölçülerine kıyasla daha az bilgi yer alır. Bu nedenle temsil yetenekleri zayıftır [1].

Anakütledeki bütün birimler kullanılarak hesaplanan ölçülere “parametre” adı verilmektedir. Söz konusu değişkenlik ölçüleri hesaplanırken serideki bütün birimler kullanılmadığı için bu ölçülere “parametrik olmayan değişkenlik ölçüleri” adı verilmiştir.

Bu ünite parametrik olmayan değişkenlik ölçülerinden değişim aralığı (ranj), kartil aralığı, desil aralığı ve pörsentil aralığı tanıtılacaktır. Basit, sınıflandırılmış ve gruplandırılmış seriler için parametrik olmayan değişkenlik ölçülerinin nasıl hesaplanacağı izah edilecektir.

DEĞİŞİM ARALIĞI

Değişim aralığı, serilerin değişkenliği hakkında yorum yapabilmek için kullanılabilecek en basit ve hesaplanması için uzun matematiksel işlemler gerektirmeyen bir ölçüdür. Aşırı uç değerlere sahip olmayan ve simetrik dağılımlarda değerlerin dağılım aralığını göstermesi bakımından kullanışlı bir ölçüdür. *Değişim aralığı, serinin değişkenliği hakkında zaman kaybetmeden genel bir bilgi sağlaması açısından bir avantaj sağlamaktadır. Ancak değişim aralığının en büyük dezavantajı; hesaplamalarda serideki bütün birimler hesaplamaya dâhil edilmeyip, sadece iki değer kullanılarak hesaplanmasıdır.* Bu yüzden, değişim aralığı aşırı değerlerin direkt etkisi altındadır. Bu durum araştırmacılar için yanıltıcı sonuçlar doğurabilir.

Basit ve sınıflandırılmış serilerde değişim aralığı, serideki en büyük değerden en küçük değerin çıkarılmasıyla elde edilir [2].

$$D. A. = X_{\max} - X_{\min}$$

Basit ve sınıflandırılmış serilerde değişim aralığının nasıl hesaplanacağı aşağıdaki örneklerle açıklanmıştır.



Değişim aralığı, serideki en büyük değerden en küçük değerin çıkarılmasıyla elde edilir.

Örnek 8.1. Aşağıdaki basit serinin değişim aralığını hesaplayınız.

X
20
22
26
30

Bu serideki en büyük değer 30 ve en küçük değer 20'dir. Böylece serinin değişim aralığı,

$$D. A. = 30 - 20 = 10$$

olarak elde edilir.

Örnek 8.2. Aşağıdaki sınıflandırılmış serinin değişim aralığını hesaplayınız.

X	f
14	5
15	8
17	4
20	6

Bu serideki en büyük değer 20 ve en küçük değer 14'tür. Böylece serinin değişim aralığı,

$$D. A. = 20 - 14 = 6$$

olarak elde edilir.

Gruplandırılmış serilerde değişim aralığı serideki son sınıfın üst sınırından ($X_{\max}$) ilk sınıfın alt sınırı ($X_{\min}$) çıkarılarak elde edilir. Gruplandırılmış serilerde değişim aralığının nasıl hesaplanacağı aşağıdaki örneklerle açıklanmıştır.

Örnek 8.3. Aşağıdaki gruplandırılmış serinin değişim aralığını hesaplayınız.

Gruplar	f
10 - 20	10
20 - 40	22
40 - 50	45
50 - 60	8

Bu gruplandırılmış serideki son sınıfın üst sınırı 60 ve ilk sınıfın alt sınırı 10'dur. Bu serideki en büyük değer olan 60'tan en küçük değer olan 10'un çıkarılması ile serinin değişim aralığı,

$$D. A. = 60 - 10 = 50$$

olarak elde edilir.

Eğer seri kesikli gruplandırılmış bir seri ise serinin sınıf sınırları elde edildikten sonra değişim aralığı hesaplanır. Bunu aşağıdaki örnekle anlatalım.

Örnek 8.4. Aşağıdaki gruplandırılmış serinin değişim aralığını hesaplayınız.

Gruplar	f
10 - 20	10
30 - 40	22
50 - 60	45
70 - 80	8

Buradaki değerler sınıf uçları olduğu için bu değerlerin öncelikle sınıf sınırlarına dönüştürülmesi gerekir. Bunun nedenle ilk önce sınıf sınırları elde edilir. Birinci sınıfın üst ucu 20 ile ikinci sınıfın alt ucu 30 arasındaki mesafe 10 birimdir. Bu mesafeyi ikiye bölüp bütün sınıf alt uçlarından çıkarır, üst sınıf uçlarına eklediğimizde sınıf uçları sınıf sınırlarına dönüşür. Yukarıdaki kesikli gruplandırılmış seri aşağıdaki şekline dönüşür.

Gruplar	f
5 - 25	10
25 - 45	22
45 - 65	45
65 - 85	8

Bu gruplandırılmış serideki son sınıfın üst sınırı 85 ve ilk sınıfın alt sınırı 5'tir. Böylece serinin değişim aralığı,

$$D. A. = 85 - 5 = 80$$

olarak elde edilir.

KARTİL ARALIĞI

Değişim aralığının hesaplanmasında sadece iki değer kullanılması nedeniyle, değişim aralığının aşırı uç değerlerin direkt etkisi altında olduğu daha önce ifade edilmişti. Değişim aralığının bu dezavantajını gidermek amacıyla kullanılan bir başka değişkenlik ölçüsü kartil aralığıdır [3].

Kartil aralığı, üçüncü kartilden birinci kartilin çıkarılmasıyla elde edilir.

Böylece en küçük ve en büyük %25'i oluşturan rakamlar dikkate alınmadığı için kartil aralığı değişim aralığına nispeten uç değerlerden daha az etkilenmektedir. Bununla birlikte hesaplamaya bütün birimlerin katılmaması bir dezavantaj olarak karşımıza çıkmaktadır. Dolayısıyla aşırı uç değerlerin olduğu serilerin değişkenlik ölçüsü hesaplanırken değişim aralığı yerine kartil aralığının kullanılması, aksi durumda ise değişim aralığının kullanılması daha uygun olacaktır.

Basit ve sınıflandırılmış serilerde kartil aralığı, serinin üçüncü kartil değerinden birinci kartil değeri çıkarılarak elde edilir. Böylece kartil aralığı,

$$K.A. = Q_3 - Q_1$$

şeklinde hesaplanır. Formülde Q_3 serinin üçüncü kartil değerini ifade ederken, Q_1 birinci kartil değerini ifade etmektedir. Basit ve sınıflandırılmış serilerde kartil aralığının nasıl hesaplanacağını anlatmadan önce basit ve sınıflandırılmış serilerde



Kartil aralığı, üçüncü kartilden birinci kartilin çıkarılmasıyla elde edilir.

birinci ve üçüncü kartillerin formüllerinin hatırlatılmasında fayda vardır. Birinci kartil değeri,

$$Q_1 = \frac{N+1}{4}$$

formülü ile hesaplanırken; üçüncü kartil değeri,

$$Q_3 = \frac{3(N+1)}{4}$$

formülü ile hesaplanmaktadır [4]. Basit ve sınıflandırılmış serilerde kartil aralığının nasıl hesaplanacağı aşağıdaki örneklerle açıklanmıştır.

Örnek 8.5. 3, 5, 8, 9, 12, 15, 20 değerlerinden oluşan basit serinin kartil aralığını hesaplayınız.

Yukarıdaki serinin birinci kartil değeri,

$$Q_1 = \frac{N+1}{4} = \frac{7+1}{4} = 2' \text{ inci değer olan } 5 \text{ değeridir.}$$

Yine söz konusu serinin üçüncü kartil değeri,

$$Q_3 = \frac{3(N+1)}{4} = \frac{3(7+1)}{4} = 6' \text{ ıncı değer olan } 15 \text{ değeridir.}$$

Bu durumda serinin kartil aralığı,

$$K.A. = Q_3 - Q_1 = 15 - 5 = 10$$

olarak elde edilir.

Örnek 8.6. Aşağıdaki sınıflandırılmış serinin kartil aralığını hesaplayınız.

X	f	k.f.
14	5	5
15	8	13
17	4	17
20	6	23

Serinin birinci kartil değeri,

$$Q_1 = \frac{N+1}{4} = \frac{23+1}{4} = 6' \text{ ıncı değer olan } 15 \text{ değeridir.}$$

Yine aynı seride üçüncü kartil değeri,

$$Q_3 = \frac{3(N+1)}{4} = \frac{3(23+1)}{4} = 18' \text{ inci değer olan } 20 \text{ değeridir.}$$

Bu durumda serinin kartil aralığı,

$$K.A. = Q_3 - Q_1 = 20 - 15 = 5$$

olarak elde edilir.



Gruplandırılmış serilerde kartil aralığı, serinin üçüncü kartil değerinden birinci kartil değeri çıkarılarak elde edilir.

Gruplandırılmış serilerde kartil aralığı, basit ve sınıflandırılmış serilerde olduğu gibi serinin üçüncü kartil değerinden birinci kartil değeri çıkarılarak elde edilir. Gruplandırılmış serilerde kartil aralığının nasıl hesaplanacağını anlatmadan önce gruptandırılmış serilerde birinci ve üçüncü kartillerin formüllerinin hatırlatılmasında fayda vardır. Birinci kartil değeri,

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum f_{Q_1-1}}{f_{Q_1}} c$$

formülü ile hesaplanırken; üçüncü kartil değeri,

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum f_{Q_3-1}}{f_{Q_3}} c$$

formülü ile hesaplanmaktadır [3]. Gruplandırılmış serilerde kartil aralığının nasıl hesaplanacağı aşağıdaki örnekle açıklanmıştır.

Örnek 8.7. Aşağıdaki gruptandırılmış serinin kartil aralığını hesaplayınız.

Gruplar	f	k.f.
1 - 3	1	1
3 - 5	2	3
5 - 7	4	7
7 - 9	3	10

Yukarıdaki gruptandırılmış seride, $N/4 = 10/4 = 2.5$ ' inci değerin olduğu sınıf birinci kartil sınıfıdır. Bu durumda birinci kartil değeri,

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum f_{Q_1-1}}{f_{Q_1}} c = 3 + \frac{2.5-1}{2} 2 = 4.5$$

olarak hesaplanır. Seride, $3N/4 = 3(10)/4 = 7.5$ ' inci değerin olduğu sınıf üçüncü kartil sınıfıdır. Bu durumda üçüncü kartil değeri,

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum f_{Q_3-1}}{f_{Q_3}} c = 7 + \frac{7.5-7}{3} 2 = 7.33$$

olarak elde edilir. Böylece yukarıdaki gruptandırılmış serinin kartil aralığı,

$$K.A. = Q_3 - Q_1 = 7.33 - 4.5 = 2.83$$

olarak elde edilir.

DESİL ARALIĞI

Değişim aralığının hesaplanmasında sadece iki değerin kullanılması nedeniyle, değişim aralığının aşırı uç değerlerin etkisi altında olduğu durumlarda değişim aralığının bu dezavantajını gidermek amacıyla kartil aralığı

kullanılmaktadır. Ancak Kartil aralığının hesaplanmasında en büyük %25'i ve en küçük %25'i oluşturan değerlerin büyüklükleri dikkate alınmamaktadır. Bu durumda serideki değerlerin %50'si değerlendirmeye alınmamaktadır. *Kartil aralığının bu dezavantajını ortadan kaldırmak için bir başka değişkenlik ölçüsü olan desil aralığı kullanılabilir.*



Desil aralığı, dokuzuncu desilden birinci desilin çıkarılmasıyla elde edilir.

Desil aralığı, en büyük desil olan dokuzuncu desilden en küçük desil olan birinci desilin çıkarılmasıyla elde edilir. Böylece en küçük ve en büyük %10'u oluşturan rakamlar dikkate alınmaz. Bu sebeple desil aralığı, değişim aralığına nispeten uç değerlerden etkilenmemektedir. Dahası desil aralığının hesaplanmasında her iki uçta yer alan %20'yi oluşturan değerler dikkate alınmaz. Bu oran kartil aralığına göre daha düşük bir orandır [5].

Basit ve sınıflandırılmış serilerde desil aralığı hesaplanırken dokuzuncu desil değerinden birinci desil değeri çıkarılır. Böylece desil aralığı,

$$D.A. = D_9 - D_1$$

şeklinde hesaplanır. Formülde D_9 serinin dokuzuncu desil değerini ifade ederken, D_1 birinci desil değerini ifade etmektedir. Basit ve sınıflandırılmış serilerde desil aralığının nasıl hesaplanacağını anlatmadan önce basit ve sınıflandırılmış serilerde birinci ve dokuzuncu desillerin formüllerinin hatırlatılmasında fayda vardır. Birinci desil değeri,

$$D_1 = \frac{N+1}{10}$$

formülü ile hesaplanırken; dokuzuncu desil değeri,

$$D_9 = \frac{9(N+1)}{10}$$

formülü ile hesaplanmaktadır [3]. Basit ve sınıflandırılmış serilerde desil aralığının nasıl hesaplanacağı aşağıdaki örneklerle açıklanmıştır.

Örnek 8.8. 3, 5, 8, 9, 12, 15, 20 değerlerinden oluşan basit serinin desil aralığını hesaplayınız.

Yukarıdaki serinin birinci desil değeri,

$$D_1 = \frac{N+1}{10} = \frac{7+1}{10} = 0.8 \text{ 'inci değer (birinci değer) olan 3 değeridir.}$$

Yine söz konusu serinin dokuzuncu desil değeri,

$$D_9 = \frac{9(N+1)}{10} = \frac{9(7+1)}{10} = 7.2 \text{ 'inci değer (yedinci değer) olan 20 değeridir. Bu}$$

durumda serinin desil aralığı,

$$D.A. = D_9 - D_1 = 20 - 3 = 17$$

olarak elde edilir.

Örnek 8.9. Aşağıdaki sınıflandırılmış serinin desil aralığını hesaplayınız.

X	f	k.f.
14	5	5
15	8	13
17	4	17
20	6	23

Yukarıdaki sınıflandırılmış seride,

$$D_1 = \frac{N+1}{10} = \frac{23+1}{10} = 2.4 \text{ 'üncü (ikinci değer) değer}$$

olan 14 değeri birinci desil değerini vermektedir. Yine aynı seride,

$$D_9 = \frac{9(N+1)}{10} = \frac{9(23+1)}{10} = 21.6 \text{ 'ıncı (yirmi ikinci değer)}$$

değer olan 20 değeri dokuzuncu desil değerini vermektedir. Bu durumda desil aralığı,

$$D. A. = D_9 - D_1 = 20 - 14 = 6$$

olarak elde edilir.

Gruplandırılmış serilerde desil aralığı, basit ve sınıflandırılmış serilerde olduğu gibi serinin dokuzuncu desil değerinden birinci desil değeri çıkarılarak elde edilir. Gruplandırılmış serilerde desil aralığının nasıl hesaplanacağını anlatmadan önce gruplandırılmış serilerde birinci ve dokuzuncu desillerin formüllerinin hatırlatılmasında fayda vardır. Birinci desil değeri,

$$D_1 = L_1 + \frac{\frac{N}{10} - \sum f_{D_1-1}}{f_{D_1}} c$$

formülü ile hesaplanırken; dokuzuncu desil değeri,

$$D_9 = L_1 + \frac{\frac{9N}{10} - \sum f_{D_9-1}}{f_{D_9}} c$$

formülü ile hesaplanmaktadır [3]. Gruplandırılmış serilerde desil aralığının nasıl hesaplanacağı aşağıdaki örnekle açıklanmıştır.

Örnek 8.10. Aşağıdaki gruplandırılmış serinin desil aralığını hesaplayınız.

Gruplar	f	k.f.
1 - 3	1	1
3 - 5	2	3
5 - 7	4	7
7 - 9	3	10

Yukarıdaki gruplandırılmış seride $N/10 = 10/10 = 1$ 'inci değerin olduğu sınıf, birinci desil sınıfıdır.

Bu durumda birinci desil değeri,

$$D_1 = L_1 + \frac{\frac{N}{10} - \sum f_{D_1-1}}{f_{D_1}} c = 1 + \frac{1-0}{1} 2 = 3$$

olarak hesaplanır. Serinin dokuzuncu desil sınıfı ise $9N/10 = 9(10)/10 = 9'$ uncu değerin olduğu sınıftır. Bu durumda dokuzuncu desil değeri,

$$D_9 = L_1 + \frac{\frac{9N}{10} - \sum f_{D_9-1}}{f_{D_9}} c = 7 + \frac{9-7}{3} 2 = 8.33$$

olarak elde edilir. Böylece desil aralığı,

$$D. A. = D_9 - D_1 = 8.33 - 3 = 5.33$$

olarak elde edilir.

PÖRSENTİL ARALIĞI



Pörsentil aralığı, doksan dokuzuncu pörsentilden birinci pörsentilin çıkarılması ile elde edilir.

Serideki değişkenliği ölçmede kullanılabilecek bir başka ölçü pörsentil aralığıdır. *Pörsentil aralığı, en büyük pörsentil olan doksan dokuzuncu pörsentilden en küçük pörsentil olan birinci pörsentilin çıkarılması ile elde edilir. Pörsentil aralığının hesaplanmasında en büyük %1 ve en küçük %1'e isabet eden değerler dikkate alınmaz.*

Ölçüm, tartım veya kayıt hatalarından kaynaklanan problemler nedeniyle seride aşırı küçük veya aşırı büyük değerler yer alabilir. Bu değerlerin dışlanması bakımından seride aşırı büyük veya aşırı küçük %2'yi oluşturan değerlerin dikkate alınmaması iyi bir yol olabilir.

Basit ve sınıflandırılmış serilerde pörsentil aralığı hesaplanırken, serinin doksan dokuzuncu değerinden birinci pörsentil değeri çıkarılır. Böylece pörsentil aralığı,

$$P.A. = P_{99} - P_1$$

şeklinde hesaplanır. Formülde; P_{99} serinin doksan dokuzuncu pörsentil değerini ifade ederken, P_1 birinci pörsentil değerini ifade etmektedir. Basit ve sınıflandırılmış serilerde pörsentil aralığının nasıl hesaplanacağını anlatmadan önce basit ve sınıflandırılmış serilerde birinci ve doksan dokuzuncu pörsentillerin formüllerinin hatırlatılmasında fayda vardır. Birinci pörsentil değeri,

$$P_1 = \frac{N+1}{100}$$

formülü ile hesaplanırken; doksan dokuzuncu pörsentil değeri,

$$P_{99} = \frac{99(N+1)}{100}$$

formülü ile hesaplanmaktadır [2].

Basit ve sınıflandırılmış serilerde pörsentil aralığının nasıl hesaplanacağı aşağıdaki örneklerle açıklanmıştır.

Örnek 8.11. 6, 7, 8, 15, 25, 30, 35, 40, 40, 42, 44, 44, 50, 60, 65, 80, 85 değerlerinden oluşan basit serinin pörsentil aralığını hesaplayınız.

Yukarıdaki serinin birinci pörsentil değeri,

$$P_1 = \frac{N+1}{100} = \frac{17+1}{100} = 0.18 \text{ 'inci değeri (birinci değeri) olan 6 değeridir.}$$

Yine söz konusu serinin doksan dokuzuncu pörsentil değeri,

$$P_{99} = \frac{99(N+1)}{100} = \frac{99(17+1)}{100} = 17.82 \text{ 'inci değeri (son değeri alınır) olan 85 değeridir.}$$

Bu durumda serinin pörsentil aralığı,

$$P.A. = P_{99} - P_1 = 85 - 6 = 79$$

olarak elde edilir.

Örnek 8.12. Aşağıdaki sınıflandırılmış serinin pörsentil aralığını hesaplayınız.

X	f	k.f.
10	5	5
15	30	35
20	45	80
40	25	105

Yukarıdaki serinin birinci pörsentil değeri,

$$P_1 = \frac{N+1}{100} = \frac{105+1}{100} = 1.06 \text{ 'inci değeri (birinci değeri) olan 10 değeridir.}$$

Yine söz konusu serinin doksan dokuzuncu pörsentil değeri,

$$P_{99} = \frac{99(N+1)}{100} = \frac{99(105+1)}{100} = 104.94 \text{ 'inci değeri (son değeri) olan 40 değeridir. Bu}$$

durumda serinin pörsentil aralığı,

$$P.A. = P_{99} - P_1 = 40 - 10 = 30$$

olarak elde edilir.



Gruplandırılmış serilerde pörsentil aralığı, serinin doksan dokuzuncu pörsentil değerinden birinci pörsentil değeri çıkarılarak elde edilir.

Gruplandırılmış serilerde pörsentil aralığı, basit ve sınıflandırılmış serilerde olduğu gibi serinin doksan dokuzuncu pörsentil değerinden birinci pörsentil değeri çıkarılarak elde edilir. Gruplandırılmış serilerde pörsentil aralığının nasıl hesaplanacağını anlatmadan önce gruplandırılmış serilerde birinci ve doksan dokuzuncu pörsentillerin formüllerinin hatırlatılmasında fayda vardır.

Birinci pörsentil değeri,

$$P_1 = L_1 + \frac{\frac{N}{100} - \sum f_{p_1-1}}{f_{p_1}} \cdot c$$

formülü ile hesaplanırken; doksan dokuzuncu pörsentil değeri,

$$P_{99} = L_1 + \frac{\frac{99N}{100} - \sum f_{p_{99}-1}}{f_{p_{99}}} c$$

formülü ile hesaplanmaktadır [3]. Gruplandırılmış serilerde pörsentil aralığının nasıl hesaplanacağı aşağıdaki örnekle açıklanmıştır.

Örnek 8.13. Aşağıdaki gruplandırılmış serinin pörsentil aralığını hesaplayınız.

Gruplar	f	k.f.
20 -40	10	10
40 - 60	25	35
60 - 80	42	77
80 - 100	38	115

Yukarıdaki gruplandırılmış seride $N/100 = 115/100 = 1.15$ ' inci değerin olduğu sınıf birinci pörsentil sınıfıdır. Bu durumda serinin birinci pörsentil değeri,

$$P_1 = L_1 + \frac{\frac{N}{100} - \sum f_{p_1-1}}{f_{p_1}} c = 20 + \frac{1.15 - 0}{10} 20 = 22.3$$

olarak hesaplanır. Seride $99N/100 = 99(115)/100 = 113.85$ 'inci değerin olduğu sınıf doksan dokuzuncu pörsentil sınıfıdır. Bu durumda serinin doksan dokuzuncu pörsentil değeri,

$$P_{99} = L_1 + \frac{\frac{99N}{100} - \sum f_{p_{99}-1}}{f_{p_{99}}} c = 80 + \frac{113.85 - 77}{38} 20 = 99.39$$

olarak elde edilir. Böylece yukarıdaki gruplandırılmış serinin pörsentil aralığı,

$$P.A. = P_{99} - P_1 = 99.39 - 22.3 = 77.09$$

olarak elde edilir.



Bireysel Etkinlik

- Seriler arasında değişkenlik karşılaştırmaları yapmada başvurulabilecek en iyi parametrik olmayan değişkenlik ölçüsü hangisidir? Nedenini tartışınız.
- Bir gruplandırılmış seri yardımıyla elde edilebilecek parametrik olmayan değişkenlik ölçülerini hesaplayınız.



Özet

- Zayıf ölçekle ölçülmüş verilerin değişkenliğini ölçmede parametrik değişkenlik ölçüleri bazen mantıklı olmayan sonuçlar doğurabilir. Parametrik değişkenlik ölçülerinin kullanılmadığı durumlarda parametrik olmayan değişkenlik ölçülerinden yararlanılır. Parametrik olmayan değişkenlik ölçülerinin elde edilmesi kolaydır. Serideki tüm değerlere tabi olmadığı için hesaplanmaları pratiktir. Bununla birlikte parametrik değişkenlik ölçülerine kıyasla içlerinde daha az bilgi yer alır.
- Değişim aralığı, serilerin değişkenliği hakkında yorum yapabilmek için kullanılabilecek en basit ve hesaplanması için uzun matematiksel işlemler gerektirmeyen bir ölçüdür. Aşırı uç değerlere sahip olmayan ve simetrik dağılımlarda değerlerin dağılım aralığını göstermesi bakımından kullanışlı bir ölçüdür. Değişim aralığı, serinin değişkenliği hakkında zaman kaybetmeden genel bir bilgi sağlaması açısından bir avantaj sağlamaktadır. Ancak değişim aralığının en büyük dezavantajı hesaplamalarda serideki bütün birimler hesaplamaya dâhil edilmeyip, sadece iki değerle hesaplanmasıdır. Bu yüzden, değişim aralığı aşırı değerlerin direkt etkisi altındadır. Değişim aralığı, serideki en büyük değerden en küçük değerin çıkarılmasıyla elde edilir.
- Kartil aralığı üçüncü kartilden birinci kartilin çıkarılmasıyla elde edilir. Böylece en küçük ve en büyük %25'i oluşturan rakamlar dikkate alınmadığı için kartil aralığı değişim aralığına nispeten uç değerlerden daha az etkilenmektedir. Bununla birlikte hesaplamaya bütün birimlerin katılmaması bir dezavantaj olarak karşımıza çıkmaktadır. Dolayısıyla aşırı uç değerlerin olduğu serilerin değişkenlik ölçüsü hesaplanırken değişim aralığı yerine kartil aralığının kullanılması, aksi durumda ise değişim aralığının kullanılması daha uygun olacaktır.
- Desil aralığı, en büyük desil olan dokuzuncu desilden en küçük desil olan birinci desilin çıkarılmasıyla elde edilir. Böylece en küçük ve en büyük %10'u oluşturan rakamlar dikkate alınmaz. Bu sebeple desil aralığı, değişim aralığına nispeten uç değerlerden etkilenmemektedir. Dahası desil aralığının hesaplanmasında her iki uçta yer alan %20'yi oluşturan değerler dikkate alınmaz. Bu oran kartil aralığına göre daha düşük bir orandır.
- Pörsentil aralığı, en büyük pörsentil olan doksan dokuzuncu pörsentilden en küçük pörsentil olan birinci pörsentilin çıkarılması ile elde edilir. Pörsentil aralığının hesaplanmasında en büyük %1 ve en küçük %1'e isabet eden değerler dikkate alınmaz.
- Ölçüm, tartım veya kayıt hatalarından kaynaklanan problemler nedeniyle seride aşırı küçük veya aşırı büyük değerler yer alabilir. Bu değerlerin dışlanması bakımından seride aşırı büyük veya aşırı küçük %2'yi oluşturan değerlerin dikkate alınmaması iyi bir yol olabilir.

DEĞERLENDİRME SORULARI

1. Yandaki serinin değişim aralığı kaçtır?

- a) 10
b) 60
c) 70
d) 75
e) 80

Gruplar	f
10 - 20	10
30 - 40	22
50 - 60	45
70 - 80	8

2. Bir serideki en küçük değer $X_{\min} = 5$ ve değişim aralığı D. A.=28 olarak tespit edildiğine göre bu serinin en büyük değeri kaçtır?

- a) 20
b) 23
c) 28
d) 33
e) 48

3. Bir basit seride; 19, 28, 33, 45, 79, 85, 94 değerleri yer almaktadır. Serinin kartil aralığı kaçtır?

- a) 75
b) 38
c) 57
d) 77
e) 85

4. Yandaki serinin birinci kartili $Q_1 = 12.92$ olarak elde edildiğine göre serinin kartil aralığı kaçtır?

- a) 10.16
b) 17.55
c) 20.25
d) 30.29
e) 32.04

Gruplar	f
2.5 – 7.5	6
7.5 – 12.5	15
12.5 – 17.5	18
17.5 – 22.5	27
22.5 – 27.5	13
27.5 – 32.5	11

5. Bir serinin üçüncü kartili $Q_3 = 32.88$ ve kartil aralığı K.A.=19.23 olarak hesaplandığına göre bu serinin birinci kartil değeri kaçtır?

- a) 13.65
b) 19.23
c) 24.84
d) 32.88
e) 52.11

6. Bir basit seride; 5, 6, 8, 9, 12, 16, 19, 24, 24, değerleri yer almaktadır. Serinin desil aralığı kaçtır?

a) 5
b) 9
c) 19
d) 21
e) 24

7. Yukarıdaki serinin desil aralığı kaçtır?

a) 5
b) 10
c) 15
d) 25
e) 45

X	f
5	2
10	7
15	13
20	12
25	15

8. Yukarıdaki serinin dokuzuncu desili $D_9 = 9.75$ olarak hesaplanmıştır. Serinin desil aralığı kaçtır?

a) 2.25
b) 3.04
c) 5.25
d) 7.25
e) 9.25

Gruplar	f
2 - 4	2
5 - 7	13
8 - 10	4
10 - 13	1

9. Bir serinin birinci pörsentili $P_1 = 1.25$ ve pörsentil aralığı P.A. = 56.84 olarak hesaplandığına göre bu serinin doksan dokuzuncu pörsentil değeri kaç olur?

a) 14.42
b) 25.85
c) 52.45
d) 55.59
e) 58.09

10. Yukarıdaki serinin doksan dokuzuncu pörsentili $P_{99} = 89.13$ olarak hesaplanmıştır. Serinin pörsentil aralığı kaçtır?

a) 27.19
b) 49.27
c) 77.13
d) 80.13
e) 89.13

Gruplar	f
10 - 30	10
30 - 50	22
50 - 70	45
70 - 90	23

Cevap Anahtarı

1.c, 2.d, 3.c, 4.a, 5.a, 6.c, 7.c, 8.c, 9.e, 10.e

YARARLANILAN KAYNAKLAR

- [1] Daniel, Wayne W., James C. Terrell (1995). Business Statistics: For Management and Economics (7. Baskı), Houghton Mifflin Company, Boston.
- [2] Armutlulu, İ. H.(2018). İşletmelerde Uygulamalı İstatistik, 3. Baskı, Alfa Yayınları, İstanbul.
- [3] Başar, A., Oktay, E. (2018). Uygulamalı İstatistik – I: Kısa Teorik Bilgiler ve Çözülmüş Problemler, Kültür ve Eğitim Kitap ve Kirtasiye, 12. Baskı, Erzurum.
- [4] Serper, Ö. (2014). Uygulamalı İstatistik I, (7. Baskı), Filiz Kitabevi, İstanbul.
- [5] Gürtan, K. (1982). İstatistik ve Araştırma Metodları, İstanbul Üniversitesi Yayınları, İstanbul.

ŞEKİL ÖLÇÜLERİ



İÇİNDEKİLER

- Çarpıklık Ölçüleri
 - Pearson Çarpıklık Ölçüleri
 - Kartillere Dayalı Çarpıklık Ölçüsü
 - Momentlere Dayalı Çarpıklık Ölçüsü
- Basıklık Ölçüsü



HEDEFLER

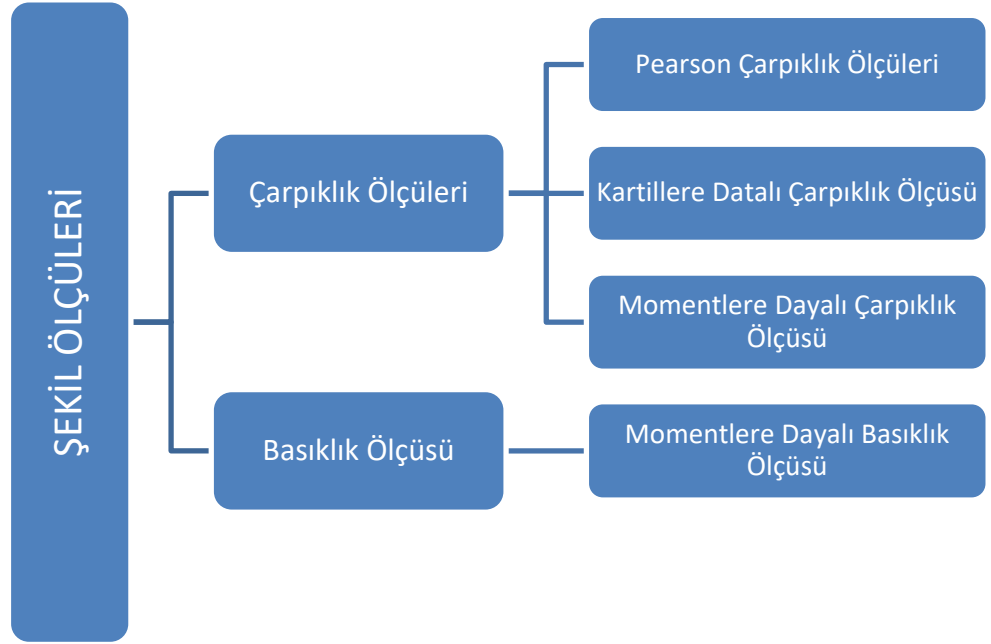
- Bu üniteyi çalıştıktan sonra;
 - Dağılım şekillerini öğrenebilecek,
 - İyi tahminler yapabilmek için gerekli şekil özelliklerini tanıyabilecek,
 - Serideki bir kısım değerler üzerinden tespit edilecek çarpıklık ölçülerini kavrayabilecek,
 - Serideki tüm değerler üzerinden çarpıklık ölçüsünü hesaplayabilecek,
 - Serilerden basıklık ölçüsünü belirleyebilecek,
 - Hesaplanan çarpıklık ve basıklık ölçüleri hakkında bilgi sahibi olabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Prof. Dr. Mehmet
Suphi ÖZÇOMAK

ÜNİTE
9



GİRİŞ

Değişkenlik ölçüleri istatistiki serilerin değişkenliğini ölçebilmesine rağmen serilerin dağılma şekilleri hakkında bilgi vermemektedir. Merkezî eğilim ölçüleri ve değişkenlik ölçüleri birbirine eşit veya yakın olan serilerin dağılımları birbirinden farklı olabilir. Örneğin, aşağıda verilen X ve Y serilerini karşılaştıralım. Bu serileri karşılaştırmak amacıyla serilerin aritmetik ortalama ve standart sapmaları aşağıdaki gibi elde edilmiştir:

X_i	f_i	Y_i	f_i
3	5	3	8
4	19	4	11
5	12	5	17
7	6	6	7
8	6	8	6
9	3	9	4
$\bar{X} = 5.25$		$\bar{Y} = 5.26$	
$s_x = 1.78$		$s_y = 1.76$	

Yukarıdaki seriler incelendiğinde, serilerin aritmetik ortalamalarının ve standart sapmalarının birbirine çok yakın olduğu görülmektedir. Oysa serilerin dağılımları incelendiğinde dağılımların birbirinden farklı olduğu görülmektedir. Bu nedenle seriler karşılaştırılmak istenildiğinde değişkenlik ölçüleri tek başlarına yeterli olmamaktadır. Değişkenlik ölçüleri ile birlikte dağılma şekillerinin de tespit edilmesi gerekir.

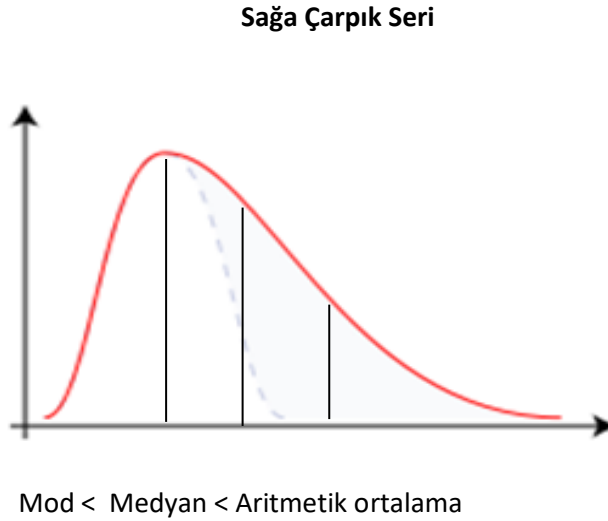
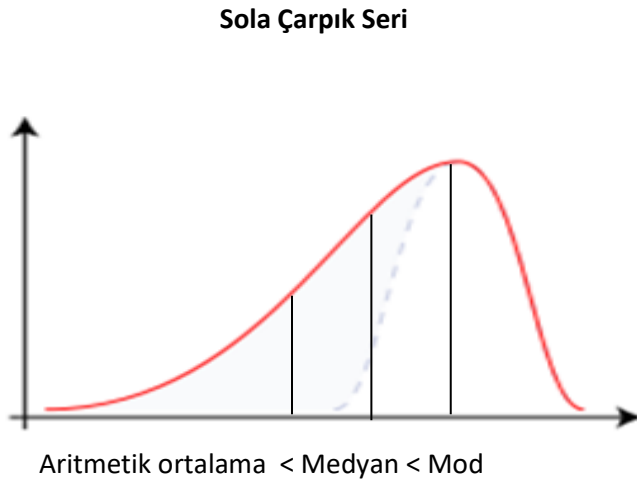
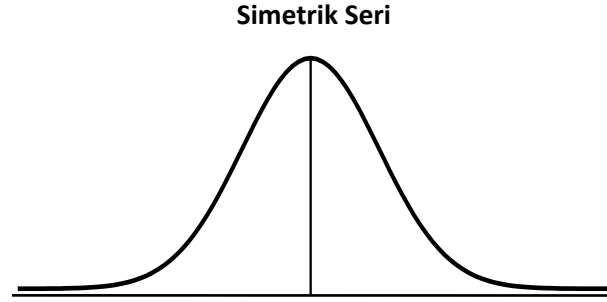
ÇARPIKLIK ÖLÇÜLERİ

Serilerin dağılma şekillerini belirlemek için çarpıklık ölçüleri adı verilen ölçüler hesaplanmaktadır. *Çarpıklık ölçüleri yardımı ile serideki gözlem değerlerinin merkezî eğilim ölçüsü etrafında simetrik dağılıp dağılmadığı tespit edilebilir.* Çarpıklık ölçüleri ile dağılma şekilleri belirlenecek serilerin bir kısmı, belli bir değere göre simetrik dağılım göstermektedir. Bu tür serilere simetrik seriler adı verilmektedir. Dağılımı simetrik olmayıp belirli değerlerde yoğunlaşan serilere ise asimetrik seriler adı verilmektedir. Asimetrik seriler dağılımın yoğunlaştığı değere göre sağa veya sola çarpık seriler olabilmektedir.

Çarpıklık kavramı, bir serinin simetriden ayrılmasıdır. Simetrik dağılım gösteren serilerde merkezî eğilim ölçüleri, dağılımın tam ortasında yer almaktadır. Serideki gözlem değerlerinin dağılımı her zaman simetrik olmaz. Bazı durumlarda sağa veya sola çarpık serilerle de karşılaşmaktadır. Serideki asimetri, hesaplanan çarpıklık ölçülerine bakılarak tespit edilebilir. *Çarpıklık ölçüsü sıfıra eşit olduğunda serinin simetrik bir seri olduğu ifade edilir. Çarpıklık ölçüsü negatif olduğunda seri sola çarpık, pozitif olduğunda ise sağa çarpıktır.* Sağa ve sola çarpık dağılım ve simetrik dağılım grafikleri Şekil 9.1’de gösterilmiştir.



Çarpıklık ölçüsü sıfıra eşit olduğunda serinin simetrik bir seri olduğu ifade edilir. Çarpıklık ölçüsü negatif olduğunda seri sola çarpık, pozitif olduğunda ise sağa çarpıktır.



Şekil 9.1. Serilerin Çarpıklık Durumları

Şekil 9.1 incelendiğinde, simetrik dağılım gösteren serilerde merkezî eğilim ölçüleri dağılımın tam ortasında yer almaktadır. Bir başka ifade ile serideki gözlem değerlerinin %50'si merkezî eğilim ölçüsünden küçük ve %50'si merkezî eğilim ölçüsünden büyüktür. Sola çarpık serilerde gözlem değerlerinin çoğunluğu sağ tarafta dağılırken, sağa çarpık serilerde gözlem değerlerinin çoğunluğu sol tarafta

dağılmaktadır. Serilerde sağa veya sola çarpıklık kavramı, eğrinin kuyruk kısmının sağa veya sola çarpık olması ile ilgilidir.

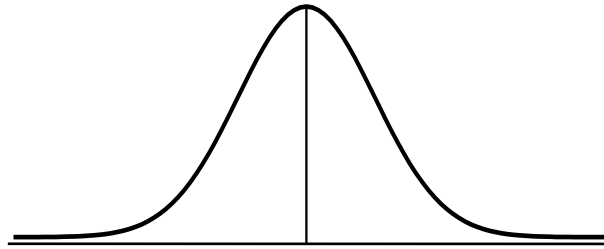
Eğrinin kuyruk kısmının sol tarafa doğru çarpık olması durumunda serinin sola çarpık, eğrinin kuyruk kısmının sağ tarafa doğru çarpık olması durumunda ise serinin sağa çarpık bir dağılım gösterdiği ifade edilir [1].

Şekil 9.1’de verilen simetrik eğri normal dağılım eğrisini göstermektedir. Normal dağılım eğrisinin ne olduğu ve ne anlam ifade ettiğini anlamak için öncelikle normal dağılımın bilinmesi gerekmektedir. Bu nedenle bu bölümde normal dağılım kısaca açıklanmıştır.

Günlük hayatta yapılan araştırmalarda çoğu zaman; boy uzunluğu, gelir, ağırlık, akülerin dayanma süresi, araçların belli bir zamanda aldıkları mesafeler, herhangi bir koşuya katılan sporcuların koşuyu bitirme süreleri örneklerinde olduğu gibi, hem tam sayı hem de kesirli sayı olarak değer alabilen sürekli değişkenler kullanılmaktadır. Sürekli değişkenlere ilişkin sürekli olasılık dağılımı veya olasılık yoğunluk fonksiyonu bir eğri yardımıyla belirlenmektedir. Söz konusu eğrinin şekli araştırılan sürekli değişkene göre değişmektedir. Örneğin; akülerin dayanma süresi için başka, herhangi bir koşuya katılan sporcuların koşuyu bitirme süreleri için başka eğri elde edilebilmektedir. Ancak birçok uygulamalı bilim dalında gözlemlenen sürekli değişkenlerin büyük bir çoğunluğu “çan eğrisi” görünümünde bir grafiğe sahiptir. Bu grafik normal olasılık yoğunluk fonksiyonu adı verilen bir fonksiyonla ifade edilmektedir.

Normal olasılık yoğunluk fonksiyonu gerek teoride gerekse uygulamada en çok kullanılan bir matematiksel model durumundadır. Nitekim yapılan birçok araştırmada; öğrenci seçme sınavlarının puanlarının, zekâ testi sonuçlarının, erkek veya kadın öğrencilerin boy uzunluklarının ve ağırlıklarının, aynı cinsten ağaç gövdelerinin çaplarının normal dağılım gösterdikleri saptanmıştır. Verilen bu örnekler uygulamada birçok dağılımın normal dağılım gösterdiğini kanıtlamaktadır.

Normal dağılım; ortalaması, modu ve medyanı birbirine eşit olan simetrik bir dağılımdır ve bu dağılım ne dik ne de basıktır. Şekil 9.2’de örnek bir normal dağılım eğrisi verilmiştir.



$$\bar{X} = \text{Medyan} = \text{Mod}$$

Şekil 9.2. Örnek Bir Normal Dağılım Eğrisi



Normal dağılımın çarpıklık katsayısı 0 ve basıklık katsayısı 3 olduğu için bir serinin çarpıklık veya basıklığı hakkında karar verilirken bu değerler referans olarak alınmaktadır.

Şekil 9.2’de görüldüğü gibi eğri çana benzediği için eğriye “çan eğrisi” adı da verilmektedir. Gerçek hayatta sürekli değişkenlere ilişkin olarak toplanan verilerin dağılımlarının her zaman normal dağılım olması beklenemez. Normal dağılıma, bir referans noktası görevi gören bir dağılım olarak bakmak daha doğru olacaktır. O yüzden çarpıklık ve basıklık ölçüleri ile ilgili karar verilirken normal dağılım referans olarak alınır. *Normal dağılımın çarpıklık katsayısı 0 ve basıklık katsayısı 3 olduğu için bir serinin çarpıklık veya basıklığı hakkında karar verilirken bu değerler referans olarak alınmaktadır* [2].

Bir serinin çarpıklığını tespit etmek için çeşitli asimetri ölçüleri geliştirilmiştir. Bu ölçüler sırasıyla Pearson çarpıklık ölçüleri, kartillere dayalı çarpıklık ölçüsü ve momentlere dayalı çarpıklık ölçüsüdür. Bu bölümde söz konusu asimetri ölçüleri anlatılmıştır.

Pearson Çarpıklık Ölçüleri

İlk kez Karl Pearson (1895) tarafından ortaya konulan bu asimetri ölçüleri, merkezî eğilim ölçüleri arasındaki ilişkiye dayanmaktadır. *Merkezî eğilim ölçüleri arasındaki ilişki,*

$$\bar{X} - \text{Mod} \cong 3(\bar{X} - \text{Medyan})$$

şeklinde ifade edilir. Bu ifadeden yola çıkan Pearson, eşitliğin her iki tarafındaki değerleri standart sapma değerine bölerek bu ifadeleri ölçü biriminden bağımsız hâle getirmiş ve daha sonra kendi adıyla anılan iki asimetri ölçüsü geliştirmiştir. Bu asimetri ölçüleri moda ve medyana dayalı Pearson çarpıklık ölçüleridir. Moda dayalı Pearson çarpıklık ölçüsü,

$$\zeta_{\text{Mod}} = \frac{\bar{X} - \text{Mod}}{s}$$

formülü ile elde edilirken; medyana dayalı Pearson çarpıklık ölçüsü ise,

$$\zeta_{\text{Medyan}} = \frac{3(\bar{X} - \text{Medyan})}{s}$$

formülü ile elde edilir.

Simetrik serilerde, aritmetik ortalama, mod ve medyan değerleri birbirine eşittir. Bu sebeple dağılım simetrik olduğunda Pearson çarpıklık ölçüleri sıfıra eşit olur. Sağa çarpık serilerde $\bar{X} > \text{Medyan} > \text{Mod}$ ve sola çarpık serilerde ise $\bar{X} < \text{Medyan} < \text{Mod}$ olur. Bu sebeple Pearson çarpıklık ölçüleri sıfırdan büyük yani pozitif olduğunda dağılımın sağa çarpık ve negatif olduğunda ise dağılımın sola çarpık olduğu ifade edilir [3].

Aşağıda basit, sınıflandırılmış ve gruplandırılmış serilerde Pearson çarpıklık ölçülerinin nasıl hesaplanacağı örneklerle açıklanmıştır.

Örnek 9.1. 1, 4, 5, 6, 7, 7, 7, 8, 9 değerlerinden oluşan basit serinin moda ve medyana dayalı Pearson çarpıklık ölçülerini hesaplayarak, yorumlayınız.



Simetrik serilerde, aritmetik ortalama, mod ve medyan değerleri birbirine eşittir.



Sağa çarpık serilerde $\bar{X} > \text{Medyan} > \text{Mod}$ ve sola çarpık serilerde ise $\bar{X} < \text{Medyan} < \text{Mod}$ olur.

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	-5	25
4	-2	4
5	-1	1
6	0	0
7	1	1
7	1	1
7	1	1
8	2	4
9	3	9

Söz konusu serinin Pearson çarpıklık ölçülerini hesaplamak için öncelikle serinin; aritmetik ortalama, mod, medyan ve standart sapma değerlerinin hesaplanması gerekir. Bu değerler aşağıdaki gibi hesaplanır:

$$\bar{X} = \frac{54}{9} = 6$$

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{46}{9-1}} = \sqrt{5.75} = 2.40$$

$$\text{Mod} = 7$$

$$\text{Medyan} = \frac{N+1}{2} = \frac{9+1}{2} = 5 \text{ 'inci}$$

değer olan 7 değeridir. Böylece moda ve medyana dayalı Pearson çarpıklık ölçüleri,

$$C_{\text{Mod}} = \frac{\bar{X} - \text{Mod}}{s} = \frac{6-7}{2.40} = -0.42$$

$$C_{\text{Medyan}} = \frac{3(\bar{X} - \text{Medyan})}{s} = \frac{3(6-7)}{2.40} = -1.25$$

olarak hesaplanır. Her iki çarpıklık katsayısına göre de söz konusu basit serinin dağılımının sola çarpık olduğu ifade edilir. Böylece serideki gözlem değerlerinin çoğunluğunun aritmetik ortalamanın üzerinde dağıldığı, ayrıca serinin dağılımı sola çarpık olduğundan $\bar{X} < \text{Medyan} < \text{Mod}$ olduğu da söylenebilir. Burada mod ve medyan değerleri birbirine eşit çıkmıştır.

Örnek 9.2. Aşağıdaki sınıflandırılmış serinin Pearson çarpıklık ölçülerini hesaplayarak, yorumlayınız.

X	f
10	8
20	25
30	10
40	5

50 2

Söz konusu serinin Pearson çarpıklık ölçülerini hesaplamak için öncelikle serinin; aritmetik ortalama, mod, medyan ve standart sapma değerlerinin hesaplanması gerekir. Bu değerler aşağıdaki gibi hesaplanır:

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
10	8	80	100	800
20	25	500	400	10000
30	10	300	900	9000
40	5	200	1600	8000
50	2	100	2500	5000

$$\bar{X} = \frac{1180}{50} = 23.6$$

$$s = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2} = \sqrt{\frac{32800}{50} - \left(\frac{1180}{50} \right)^2} = \sqrt{99.04} = 9.95$$

Mod = 20

$$\text{Medyan} = \frac{N+1}{2} = \frac{50+1}{2} = 25.5' \text{ inci}$$

değer olan 20 değeridir. Böylece moda ve medyana dayalı Pearson çarpıklık ölçüleri,

$$C_{\text{Mod}} = \frac{\bar{X} - \text{Mod}}{s} = \frac{23.6 - 20}{9.95} = 0.36$$

$$C_{\text{Medyan}} = \frac{3(\bar{X} - \text{Medyan})}{s} = \frac{3(23.6 - 20)}{9.95} = 1.09$$

olarak hesaplanır. Her iki çarpıklık ölçüsüne göre söz konusu sınıflandırılmış serinin dağılımının sağa çarpık olduğu ifade edilir. Serinin dağılımı sağa çarpık olduğundan dolayı gözlem değerlerinin çoğunluğunun aritmetik ortalamanın altında olduğu, ayrıca seride $\bar{X} > \text{Medyan} > \text{Mod}$ olduğu tespit edilir. Burada mod ve medyan değerleri birbirine eşit çıkmıştır.

Örnek 9.3. Aşağıdaki gruplandırılmış serinin Pearson çarpıklık ölçülerini hesaplayarak, yorumlayınız.

Gruplar	f
1 - 3	2
4 - 6	3
7 - 9	5
10 - 12	7

Söz konusu serinin Pearson çarpıklık ölçülerini hesaplamak için öncelikle serinin; aritmetik ortalama, mod, medyan ve standart sapma değerlerinin hesaplanması gerekir. Bu değerler aşağıdaki gibi hesaplanır:

X_i	f_i	$f_i X_i$	X_i^2	$f_i X_i^2$
2	2	4	4	8
5	3	15	25	75
8	5	40	64	320
11	7	77	121	847

$$\bar{X} = \frac{136}{17} = 8$$

$$s = \sqrt{\frac{\sum_{i=1}^n f_i X_i^2}{\sum f_i} - \left(\frac{\sum_{i=1}^n f_i X_i}{\sum f_i} \right)^2} = \sqrt{\frac{1250}{17} - \left(\frac{136}{17} \right)^2} = \sqrt{9.53} = 3.09$$

$$\text{Mod} = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} c = 9.5 + \frac{2}{2+7} 3 = 10.17$$

$$\text{Medyan} = L_1 + \frac{\frac{N}{2} - \sum f_{M-1}}{f_M} c = 6.5 + \frac{\frac{17}{2} - 5}{5} 3 = 8.6$$

Böylece moda ve medyana dayalı Pearson çarpıklık ölçüleri,

$$C_{\text{Mod}} = \frac{\bar{X} - \text{Mod}}{s} = \frac{8 - 10.17}{3.09} = -0.70$$

$$C_{\text{Medyan}} = \frac{3(\bar{X} - \text{Medyan})}{s} = \frac{3(8 - 8.6)}{3.09} = -0.58$$

olarak hesaplanır. Pearson çarpıklık ölçüleri negatif olduğundan dağılımın sola çarpık olduğu ifade edilir. Gruplandırılmış serinin dağılımı sola çarpık olduğundan dolayı gözlem değerlerinin çoğunluğunun aritmetik ortalamanın üstünde dağıldığı, ayrıca seride $\bar{X} < \text{Medyan} < \text{Mod}$ olduğu tespit edilir.

Kartillere Dayalı Çarpıklık Ölçüsü

Kartiller, küçükten büyüğe doğru sıralanan değerleri dört eşit parçaya bölen değerlerdir. Bir serinin çarpıklığının hesaplanmasında kullanılan bir başka ölçü de kartillere dayalı çarpıklık ölçüsüdür. *Bilindiği üzere kartiller arasında şu şekilde ilişkiler bulunmaktadır:*

Simetrik serilerde: $(Q_3 - Q_2) = (Q_2 - Q_1)$

Sağa çarpık serilerde: $(Q_3 - Q_2) > (Q_2 - Q_1)$

Sola çarpık serilerde: $(Q_3 - Q_2) < (Q_2 - Q_1)$

Bu yaklaşımdan hareketle kartillere dayalı çarpıklık ölçüsü geliştirilmiştir. Söz konusu çarpıklık ölçüsü Bowley çarpıklık ölçüsü olarak da adlandırılmaktadır. Kartillere dayalı çarpıklık ölçüsü,

$$\zeta_Q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$



Kartillere dayalı çarpıklık ölçüsü, Pearson çarpıklık ölçüleri gibi yorumlanır.

formülü yardımıyla hesaplanmaktadır. Formülde Q_1 birinci kartil değerini, Q_2 ikinci kartil değerini ve Q_3 ise üçüncü kartil değerini ifade etmektedir. Kartillere dayalı çarpıklık ölçüsü, Pearson çarpıklık ölçüleri gibi yorumlanır. Yani çarpıklık ölçüsü sıfıra eşit olduğunda dağılımın simetrik olduğu, sıfırdan büyük yani pozitif olduğunda dağılımın sağa çarpık ve negatif olduğunda ise dağılımın sola çarpık olduğu ifade edilir [2].

Kartil değerleri hesaplanırken serideki bütün gözlem değerleri kullanılmadığından dolayı kartillere dayalı çarpıklık ölçüsü çarpıklık hakkında yaklaşık sonuçlar vermektedir. Birer örnek yardımıyla basit, sınıflandırılmış ve gruplandırılmış serilerde kartil çarpıklık ölçüsünün elde edilişi anlatılmıştır.

Örnek 9.4. 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 değerlerinden oluşan basit serinin kartillere dayalı çarpıklık ölçüsünü hesaplayarak, yorumlayınız.

İlk olarak serinin kartil değerleri aşağıdaki hesaplanır:

$$Q_1 = \frac{N+1}{4} = \frac{11+1}{4} = 3 \text{ 'üncü değer olan 16 değeridir.}$$

$$Q_2 = \frac{N+1}{2} = \frac{11+1}{2} = 6 \text{ 'ıncı değer olan 22 değeridir.}$$

$$Q_3 = \frac{3(N+1)}{4} = \frac{3(11+1)}{4} = 9 \text{ 'uncu değer olan 28 değeridir.}$$

Buna göre kartillere dayalı çarpıklık ölçüsü,

$$\zeta_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{28 - 2(22) + 16}{28 - 16} = 0$$

olarak hesaplanır. Kartillere dayalı çarpıklık ölçüsü sıfıra eşit olduğundan, serinin dağılımının simetrik olduğu ifade edilir.

Örnek 9.5. Aşağıdaki sınıflandırılmış serinin kartillere dayalı çarpıklık ölçüsünü hesaplayarak, yorumlayınız.

X_i	f_i
6	15
9	35
11	55
13	20
15	10

İlk olarak serinin kartil değerleri aşağıdaki gibi hesaplanır:

$$Q_1 = \frac{N+1}{4} = \frac{135+1}{4} = 34 \text{ 'üncü değer olan 9 değeridir.}$$

$$Q_2 = \frac{N+1}{2} = \frac{135+1}{2} = 68 \text{ 'inci değer olan 11 değeridir.}$$

$$Q_3 = \frac{3(N+1)}{4} = \frac{3(135+1)}{4} = 102 \text{ 'inci değer olan 11 değeridir.}$$

Buna göre kartillere dayalı çarpıklık ölçüsü,

$$\zeta_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{11 - 2(11) + 9}{11 - 9} = -1$$

olarak hesaplanır. Çarpıklık ölçüsü negatif olduğundan dağılımın sola çarpık olduğu ifade edilir.

Örnek 9.6. Aşağıdaki gruplandırılmış serinin kartillere dayalı çarpıklık ölçüsünü hesaplayarak, yorumlayınız.

Gruplar	f
1 - 3	1
3 - 5	2
5 - 7	4
7 - 9	3

İlk olarak serinin kartil değerleri aşağıdaki gibi hesaplanır. Yukarıdaki gruplandırılmış seride, $N/4 = 10/4 = 2.5$ 'inci değer olduğu sınıf birinci kartil sınıfıdır. Bu durumda birinci kartil değeri,

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum f_{Q_1-1}}{f_{Q_1}} c = 3 + \frac{2.5-1}{2} 2 = 4.5$$

olarak hesaplanır.

Seride, $N/2 = 10/2 = 5$ 'inci değer olduğu sınıf ikinci kartil sınıfıdır. Bu durumda ikinci kartil değeri,

$$Q_2 = L_1 + \frac{\frac{N}{2} - \sum f_{Q_2-1}}{f_{Q_2}} c = 5 + \frac{5-3}{4} 2 = 6$$

olarak hesaplanır.

Son olarak seride, $3N/4 = 3(10)/4 = 7.5$ 'inci değer olduğu sınıf üçüncü kartil sınıfıdır.

Bu durumda üçüncü kartil değeri,

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum f_{Q_3-1}}{f_{Q_3}} c = 7 + \frac{7.5 - 7}{3} 2 = 7.33$$

olarak hesaplanır. Bu değerler kullanılarak seriye ait kartillere dayalı çarpıklık ölçüsü,

$$\zeta_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{7.33 - 2(6) + 4.5}{7.33 - 4.5} = -0.06$$

olarak hesaplanır. Çarpıklık ölçüsü negatif olduğundan serinin dağılımının sola çarpık olduğu ifade edilir.

Momentlere Dayalı Çarpıklık Ölçüsü

Gerek Pearson çarpıklık ölçüleri gerekse kartillere dayalı çarpıklık ölçüsü bir serinin çarpıklık durumu hakkında yaklaşık bir fikir vermektedir. Çarpıklığı daha duyarlı bir şekilde ölçebilmek amacıyla momentlere dayalı çarpıklık ölçüsü kullanılabilir. Momentlere dayalı çarpıklık ölçüsü, ortalama etrafındaki üçüncü momentin standart sapmanın üçüncü kuvvetine bölünmesi ile elde edilmektedir. Momentlere dayalı çarpıklık ölçüsü α_3 ile ifade edilir ve

$$\alpha_3 = \frac{\mu_3}{(s)^3}$$

formülü yardımıyla hesaplanır.

Momentlere dayalı çarpıklık ölçüsü daha önce anlatılan çarpıklık ölçüleri gibi yorumlanmaktadır. Yani çarpıklık ölçüsü sıfıra eşit olduğunda dağılımın simetrik olduğu, sıfırdan büyük yani pozitif olduğunda dağılımın sağa çarpık ve negatif olduğunda ise dağılımın sola çarpık olduğu ifade edilir [4]. Söz konusu ölçünün temelini momentler oluşturmaktadır. Bu nedenle öncelikle momentler hakkında bilgi verilmiştir.

Serideki gözlem değerlerinin sıfırdan veya aritmetik ortalamadan farklarının çeşitli kuvvetlerinin aritmetik ortalamalarının tamamına moment adı verilmektedir. Momentler, sıfır etrafındaki momentler ve aritmetik ortalama etrafındaki momentler olmak üzere iki gruba ayrılmaktadır. Her iki moment çeşidinde sıfırdan veya aritmetik ortalamadan farkların derecesi momentin derecesini belirlemektedir [5].

Basit seride X_i değişkeninin sıfır etrafındaki r 'inci momenti,

$$M_r = \frac{\sum_{i=1}^n x_i^r}{n}$$

formülüyle hesaplanırken; sınıflandırılmış ve gruplandırılmış serilerde sıfır etrafındaki r 'inci moment,



Momentlere dayalı çarpıklık ölçüsü, ortalama etrafındaki üçüncü momentin standart sapmanın üçüncü kuvvetine bölünmesi ile elde edilmektedir.



Momentlere dayalı çarpıklık ölçüsü, Pearson çarpıklık ölçüleri gibi yorumlanır.

$$M_r = \frac{\sum_{i=1}^n f_i X_i^r}{\sum_{i=1}^n f_i}$$

formülü ile hesaplanır. Sıfır etrafındaki birinci moment için $r = 1$, ikinci moment için $r = 2$ ve sıfır etrafındaki n 'inci moment için $r = n$ olarak alınır. Basit seride X değişkeninin aritmetik ortalama etrafındaki r 'inci momenti,

$$\mu_r = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n}$$

formülüyle hesaplanırken; sınıflandırılmış veya gruplandırılmış bir serideki X değişkeninin aritmetik ortalama etrafındaki r 'inci momenti,

$$\mu_r = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^r}{\sum_{i=1}^n f_i}$$

formülü ile hesaplanır. Bir seride gözlem değerlerinin aritmetik ortalamadan farklarının toplamı sıfır olduğundan, ortalama etrafındaki birinci moment $\mu_1 = 0$ 'dır. Ortalama etrafındaki ikinci moment ise varyans değerini vermektedir [6]. Şimdi basit, sınıflandırılmış ve gruplandırılmış serilerde momentlere dayalı çarpıklık ölçüsünün nasıl hesaplanacağı birer örnekle açıklanmıştır.

Örnek 9.7. Aşağıdaki basit serinin momentlere dayalı çarpıklık ölçüsünü hesaplayarak, yorumlayınız.

X
10
12
18
20

Bu serinin aritmetik ortalaması, $\bar{X} = 15$ 'tir. Momentlere dayalı çarpıklık ölçüsünü hesaplamak için ilk önce aritmetik ortalama etrafındaki üçüncü moment ile standart sapma değerleri hesaplanır. Bu hesaplamalar aşağıda verilmiştir:

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^3$
10	-5	25	-125
12	-3	9	-27
18	3	9	27
20	5	25	125

$$\mu_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n} = \frac{0}{4} = 0$$

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{68}{3}} = 4.76$$

Aritmetik ortalama etrafındaki üçüncü moment ile standart sapma değerleri hesaplandıktan sonra momentlere dayalı çarpıklık ölçüsü,

$$\alpha_3 = \frac{\mu_3}{(s)^3} = \frac{0}{(4.76)^3} = 0$$

olarak elde edilir. Momentlere dayalı çarpıklık ölçüsü sıfıra eşit olduğundan dolayı serinin dağılımının simetrik olduğuna karar verilir. Dolayısıyla serideki gözlem değerlerinin %50'si aritmetik ortalamadan küçük ve %50'sinin aritmetik ortalamadan büyük olduğu ifade edilir.

Örnek 9.8. Aşağıdaki sınıflandırılmış serinin momentlere dayalı asimetri ölçüsünü hesaplayarak, yorumlayınız.

X_i	f_i
2	1
4	2
6	4
8	3

Bu serinin aritmetik ortalaması, $\bar{X} = 5.8$ 'dir. Sınıflandırılmış serilerde momentlere dayalı çarpıklık ölçüsünü hesaplamak amacıyla ilk olarak aritmetik ortalama etrafındaki üçüncü moment ile standart sapma hesaplanır.

X_i	f_i	$f_i X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$	$(X_i - \bar{X})^3$	$f_i (X_i - \bar{X})^3$
2	1	2	-3.8	14.44	14.44	-54.87	-54.87
4	2	8	-1.8	3.24	6.48	-5.83	-11.66
6	4	24	0.2	0.04	0.16	0.008	0.03
8	3	24	2.2	4.84	14.52	10.65	31.95
	10	58			35.60		-34.55

$$\mu_3 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^3}{\sum_{i=1}^n f_i} = \frac{-34.55}{10} = -3.46$$

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum_{i=1}^n f_i}} = \sqrt{\frac{35.60}{10}} = 1.89$$

Aritmetik ortalama etrafındaki üçüncü moment ile standart sapma değerleri hesaplandıktan sonra momentlere dayalı çarpıklık ölçüsü,

$$\alpha_3 = \frac{\mu_3}{(s)^3} = \frac{-3.46}{(1.89)^3} = -0.51$$

olarak elde edilir. Momentlere dayalı çarpıklık ölçüsü negatif olduğundan dolayı serinin dağılımının sola çarpık olduğuna karar verilir.

Örnek 9.9. Aşağıdaki gruplandırılmış serinin momentlere dayalı çarpıklık ölçüsünü hesaplayarak, yorumlayınız.

Gruplar	f_i
0 - 3	5
3 - 6	30
6 - 9	15
9 - 12	10

Bu serinin aritmetik ortalaması, $\bar{X} = 6$ 'dır. İlk olarak aritmetik ortalama etrafındaki üçüncü moment ile standart sapma değerleri, hesaplanır.

Gruplar	f_i	X_i	$f_i X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$	$(X_i - \bar{X})^3$	$f_i (X_i - \bar{X})^3$
0 - 3	5	1.5	7.5	-4.5	20.25	101.25	-91.13	-455.65
3 - 6	30	4.5	135.0	-1.5	2.25	67.50	-3.38	-101.40
6 - 9	15	7.5	112.5	1.5	2.25	33.75	3.38	50.70
9 - 12	10	10.5	105.0	4.5	20.25	202.50	91.13	911.30
	60		360			405		404.95

$$\mu_3 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^3}{\sum_{i=1}^n f_i} = \frac{404.95}{60} = 6.75$$

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum_{i=1}^n f_i}} = \sqrt{\frac{405}{60}} = 2.6$$

olarak elde edilir. Böylece momentlere dayalı çarpıklık ölçüsü,

$$\alpha_3 = \frac{\mu_3}{(s)^3} = \frac{6.75}{(2.6)^3} = 0.38$$

olarak bulunur. Momentlere dayalı çarpıklık ölçüsü pozitif olduğundan dolayı serinin dağılımının sağa çarpık olduğuna karar verilir.

BASIKLIK ÖLÇÜSÜ



Momentlere dayalı basıklık ölçüsü, normal dağılımın basıklık ölçüsü olan 3 değeri ile karşılaştırılır.

İstatistiki bir seride gözlem değerleri simetrik dağılmadıkları durumlarda merkezî eğilim ölçüsü etrafında bazen toplu hâlde bazen de dağılmış olabilir. Serinin çarpıklık ölçüsünden farklı olarak basıklık kavramında, serideki gözlem değerlerinin belli bir aralıkta yoğunluğu veya seyrekliği ile ilgilenilir. Basıklık kavramı bir dağılımın diklik derecesinin ölçüsüdür. Bu konuda kullanılan en yaygın ölçü, momentlere dayalı basıklık ölçüsüdür.

Momentlere dayalı basıklık ölçüsü, ortalama etrafındaki dördüncü momentin standart sapmanın dördüncü kuvvetine bölünmesiyle elde edilir. Böylece momentlere dayalı basıklık ölçüsü,

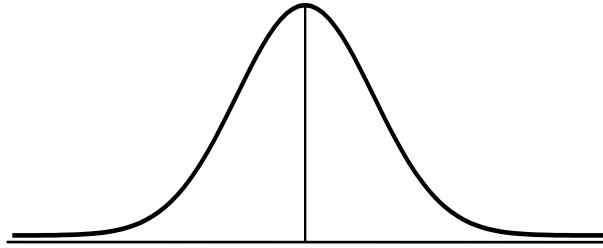
$$\alpha_4 = \frac{\mu_4}{(s)^4}$$

formülü ile hesaplanır. Momentlere dayalı basıklık ölçüsü, normal dağılımın basıklık ölçüsü olan 3 değeri ile karşılaştırılır. Şayet $\alpha_4 = 3$ ise yüksekliğin normal olduğu, $\alpha_4 > 3$ olduğunda dağılımın normale göre dik ve $\alpha_4 < 3$ olduğunda ise dağılımın normale göre basık olduğu ifade edilir [3]. Basıklığa ilişkin görsel gösterimler şekil 9.3'de yer almaktadır.

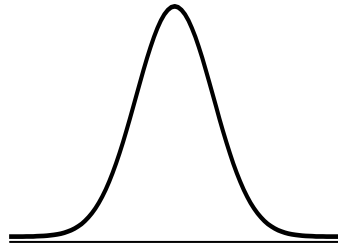


$\alpha_4 = 3$ ise yüksekliğin normal olduğu, $\alpha_4 > 3$ olduğunda dağılımın normale göre dik ve $\alpha_4 < 3$ olduğunda ise dağılımın normale göre basık olduğu ifade edilir.

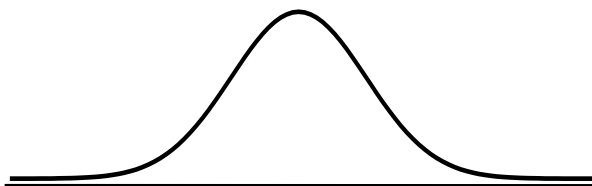
Dağılımın Yüksekliği Normal



Dağılım Normale Göre Dik



Dağılım Normale Göre Basık



Şekil 9.3. Serilerin Basıklık Dağılımları

Şekil 9.3'te görüldüğü gibi yüksekliğin normal olması dağılımın simetrik olduğunu ifade etmektedir. Dağılımın normale göre daha basık olması, dağılımın değişkenliğinin fazla olduğunu gösterirken; dağılımın normale göre daha dik olması, serideki gözlem değerlerinin merkezî eğilim ölçüsü etrafında daha sık dağıldığını göstermektedir [7].

Aşağıdaki örnekler yardımıyla basit, sınıflandırılmış ve gruplandırılmış serilerde momentlere dayalı basıklık ölçüsünün nasıl hesaplanacağı açıklanmıştır.

Örnek 9.10. Aşağıdaki basit serinin momentlere dayalı basıklık ölçüsünü hesaplayarak, yorumlayınız.

X
10
12
18
20

Bu serinin aritmetik ortalaması, $\bar{X} = 15$ 'tir. Momentlere dayalı basıklık ölçüsünü hesaplamak için ilk önce aritmetik ortalama etrafındaki dördüncü moment ile standart sapma değerleri aşağıdaki gibi hesaplanır:

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^4$
10	-5	25	625
12	-3	9	81
18	3	9	81
20	5	25	625
60		68	1412

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{68}{4-1}} = 4.76$$

$$\mu_4 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n} = \frac{1412}{4} = 353$$

Momentlere dayalı basıklık ölçüsü,

$$\alpha_4 = \frac{\mu_4}{(s)^4} = \frac{353}{(4.76)^4} = \frac{353}{513.37} = 0.69$$

olarak elde edilir. Basıklık ölçüsü $\alpha_4 = 0.69$ değeri 3 değerinden küçük olduğu için dağılımın normale göre basık olduğu ifade edilir. Böylece dağılımın değişkenliğinin fazla olduğuna karar verilir.

Örnek 9.11. Aşağıdaki sınıflandırılmış serinin momentlere dayalı basıklık ölçüsünü hesaplayarak, yorumlayınız.

X_i	f_i
2	15
6	44
10	30
14	18
18	13

Bu serinin aritmetik ortalaması, $\bar{X} = 9$ 'dur. İlk olarak standart sapma ile aritmetik ortalama etrafındaki dördüncü moment değerleri hesaplanmıştır.

X_i	f_i	$f_i X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$	$(X_i - \bar{X})^4$	$f_i (X_i - \bar{X})^4$
2	15	30	-7	49	735	2401	36015
6	44	264	-3	9	396	81	3584
10	30	300	1	1	30	1	30
14	18	252	5	25	450	625	11250
18	13	234	9	81	1053	6561	85293
	120	1080			2664		136152

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum_{i=1}^n f_i}} = \sqrt{\frac{2664}{120}} = 4.71$$

$$\mu_4 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^4}{\sum_{i=1}^n f_i} = \frac{136152}{120} = 1134.60$$

Buna göre, söz konusu gruplandırılmış serinin momentlere dayalı basıklık ölçüsü,

$$\alpha_4 = \frac{\mu_4}{(s)^4} = \frac{1134.60}{(4.71)^4} = 2.31$$

olarak hesaplanır. $\alpha_4 = 2.31 < 3$ olduğundan söz konusu serinin normale göre basık olduğu tespit edilir. Bir başka deyişle, söz konusu dağılımın değişkenliğinin fazla olduğuna karar verilir.

Gruplandırılmış serilerde momentlere dayalı basıklık ölçüsünün nasıl elde edileceği aşağıdaki örnekle açıklanmıştır.

Örnek 9.12. Aşağıdaki gruptandırılmış serinin momentlere dayalı basıklık ölçüsünü hesaplayarak, yorumlayınız.

Gruplar	f_i
0 - 3	5
3 - 6	30
6 - 9	15
9 - 12	10

Bu serinin aritmetik ortalaması, $\bar{X} = 6$ 'dır. İlk olarak standart sapma ile aritmetik ortalama etrafındaki dördüncü moment değerleri aşağıdaki gibi hesaplanır:

Gruplar	f_i	X_i	$f_i X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$	$(X_i - \bar{X})^4$	$f_i (X_i - \bar{X})^4$
0 - 3	5	1.5	7.5	-4.5	20.25	101.25	410.06	2050.3
3 - 6	30	4.5	135.0	-1.5	2.25	67.50	5.06	151.8
6 - 9	15	7.5	112.5	1.5	2.25	33.75	5.06	75.9
9 - 12	10	10.5	105.0	4.5	20.25	202.50	410.06	4100.6
	60		360			405		6378.6

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{\sum_{i=1}^n f_i}} = \sqrt{\frac{405}{60}} = 2.6$$

$$\mu_4 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^4}{\sum_{i=1}^n f_i} = \frac{6378.6}{60} = 106.31$$

Buna göre, söz konusu gruptandırılmış serinin momentlere dayalı basıklık ölçüsü,

$$\alpha_4 = \frac{\mu_4}{(s)^4} = \frac{106.31}{(2.6)^4} = 2.33$$

olarak hesaplanır. Elde edilen momentlere dayalı basıklık ölçüsü $\alpha_4 = 2.33 < 3$ olduğundan söz konusu seri normale göre basıktır. Bir başka deyişle, serideki değişkenliğin fazla olduğuna karar verilir.



Bireysel Etkinlik

- İyi tahminler yapabilmek için dağılımın şeklini bilmek niçin önemlidir? Tartışınız.
- Birer örnek yardımıyla basit, sınıflandırılmış ve gruplandırılmış serilerden elde edilebilecek şekil ölçülerini hesaplayıp yorumlayınız.



Özet

- Değişkenlik ölçüleri istatistiki serilerin değişkenliğini ölçebilmesine rağmen serilerin dağılma şekilleri hakkında bilgi vermemektedir.
- Serilerin dağılma şekillerini belirlemek için çarpıklık ölçüleri adı verilen ölçüler hesaplanmaktadır. Çarpıklık ölçüleri yardımı ile serideki gözlem değerlerinin merkezî eğilim ölçüsü etrafında simetrik dağılıp dağılmadığı tespit edilebilir. Çarpıklık ölçüleri ile dağılma şekilleri belirlenecek serilerin bir kısmı, belli bir değere göre simetrik dağılım göstermektedir. Bu tür serilere simetrik seriler adı verilmektedir. Dağılımı simetrik olmayıp belirli değerlerde yoğunlaşan serilere ise asimetrik seriler adı verilmektedir. Asimetrik seriler dağılımın yoğunlaştığı değere göre sağa veya sola çarpık seriler olabilmektedir.
- Çarpıklık kavramı, bir serinin simetriden ayrılmasıdır. Simetrik dağılım gösteren serilerde merkezî eğilim ölçüleri, dağılımın tam ortasında yer almaktadır. Serideki gözlem değerlerinin dağılımı her zaman simetrik olmaz. Uygulamada sağa veya sola çarpık serilerle de karşılaşmaktadır. Serideki asimetri, hesaplanan çarpıklık ölçülerine bakılarak tespit edilebilir. Çarpıklık ölçüsü sıfıra eşit olduğunda serinin simetrik bir seri olduğu ifade edilir. Çarpıklık ölçüsü negatif olduğunda seri sola çarpık, pozitif olduğunda ise sağa çarpıktır.
- Simetrik dağılım gösteren serilerde merkezî eğilim ölçüleri dağılımın tam ortasında yer almaktadır. Sola çarpık serilerde gözlem değerlerinin yoğunluğu sağ tarafta dağılırken, sağa çarpık serilerde gözlem değerlerinin yoğunluğu sol tarafta dağılmaktadır. Serilerde sağa veya sola çarpıklık kavramı, normal dağılım eğrisinin kuyruk kısmının sağa veya sola çarpık olması ile ilgilidir.
- Simetrik serilerde; aritmetik ortalama, mod ve medyan değerleri birbirine eşittir. Bu sebeple dağılım simetrik olduğunda Pearson çarpıklık ölçüleri sıfıra eşit olur. Sağa çarpık serilerde Aritmetik Ortalama > Medyan > Mod ve sola çarpık serilerde ise Aritmetik Ortalama < Medyan < Mod olur. Bu sebeple Pearson çarpıklık ölçüleri sıfırdan büyük yani pozitif olduğunda dağılımın sağa çarpık ve negatif olduğunda ise dağılımın sola çarpık olduğu ifade edilir.
- Kartillere dayalı çarpıklık ölçüsü Bowley çarpıklık ölçüsü olarak da adlandırılmaktadır. Kartillere dayalı çarpıklık ölçüsü, Pearson çarpıklık ölçüleri gibi yorumlanır.
- Çarpıklığı daha duyarlı bir şekilde ölçebilmek amacıyla momentlere dayalı çarpıklık ölçüsü kullanılabilir. Momentlere dayalı çarpıklık ölçüsü ortalama etrafındaki üçüncü momentin standart sapmanın üçüncü kuvvetine bölünmesi ile elde edilmektedir. Momentlere dayalı çarpıklık ölçüsü daha önce anlatılan çarpıklık ölçüleri gibi yorumlanmaktadır.
- Serideki gözlem değerlerinin sıfırdan veya aritmetik ortalamadan farklarının çeşitli kuvvetlerinin aritmetik ortalamalarının tamamına moment adı verilmektedir. Momentler, sıfır etrafındaki momentler ve aritmetik ortalama etrafındaki momentler olmak üzere iki gruba ayrılmaktadır. Her iki moment çeşidinde sıfırdan veya aritmetik ortalamadan farkların derecesi momentin derecesini belirlemektedir.
- Basıklık kavramı bir dağılımın diklik derecesinin ölçüsüdür. Bu konuda kullanılan en yaygın ölçü, momentlere dayalı basıklık ölçüsüdür. Momentlere dayalı basıklık ölçüsü, ortalama etrafındaki dördüncü momentin standart sapmanın dördüncü kuvvetine bölünmesiyle elde edilir.

DEĞERLENDİRME SORULARI

1. Bir serinin mod değeri $Mod=6.15$ ve medyan değeri $Medyan=6.35$ olarak hesaplandığına göre bu serinin aritmetik ortalaması yaklaşık olarak kaç olur?
 - a) 2.33
 - b) 3.89
 - c) 5.12
 - d) 6.13
 - e) 6.45
2. 3, 5, 6, 9, 9, 10 değerlerinden oluşan basit serinin standart sapması $s=2.76$ olarak hesaplandığına göre medyana dayalı Pearson çarpıklık ölçüsü kaçtır?
 - a) -2.76
 - b) -0.54
 - c) -0.18
 - d) 0.54
 - e) 1.12
3. İkinci soruda verilen serinin moda dayalı Pearson çarpıklık ölçüsü kaçtır?
 - a) -2.54
 - b) -2.48
 - c) -0.72
 - d) 1.12
 - e) 3.72
4. Yandaki serinin varyansı $s^2=3.56$ olarak hesaplanmıştır. Serinin moda dayalı Pearson çarpıklık ölçüsü kaçtır?

X	f
2	1
4	2
6	4
8	3

 - a) -1.34
 - b) -0.89
 - c) -0.11
 - d) 0.20
 - e) 3.56
5. 1, 1, 2, 3, 4, 4, 6, 7, 9, 11 değerlerinden oluşan basit serinin birinci kartil değeri $Q_1=2$ olduğuna göre serinin kartillere dayalı çarpıklık ölçüsü kaç olur?
 - a) -3.72
 - b) -0.24
 - c) 0.20
 - d) 0.43
 - e) 1.57

6. Yandaki serinin birinci ve ikinci kartilleri sırasıyla $Q_1=4.27$ ve $Q_2=5.45$ olduğuna göre serinin kartillere dayalı çarpıklık ölçüsü kaç olur?

Gruplar	f
2 - 4	10
4 - 6	22
6 - 8	12
8 - 10	8

- a) 2.456
b) 0.328
c) 0.186
d) 3.335
e) 0.489
7. 20, 25, 40, 50, 65 değerlerinden oluşan basit serinin standart sapma değeri $s = 18.37$ olarak hesaplanmıştır. Serinin momentlere dayalı çarpıklık ölçüsünün değeri kaç olur?
a) 0.17
b) 0.88
c) 1.23
d) 2.36
e) 3.15
8. Aritmetik ortalaması $\bar{X} = 9$, varyans değeri $s^2 = 22.2$ ve aritmetik ortalama etrafındaki üçüncü moment değeri $\mu_3 = 45.2$ olan bir serinin momentlere dayalı çarpıklık ölçüsünün değeri kaç olur?
a) -4.37
b) -2.64
c) 0.43
d) 1.26
e) 3.45
9. Momentlere dayalı çarpıklık ölçüsü $\alpha_3 = 0.18$ ve varyans değeri $s^2 = 2.29$ olan serinin aritmetik ortalama etrafındaki üçüncü moment değeri kaç olur?
a) 0.62
b) 2.11
c) 3.24
d) 5.89
e) 4.37
10. 3, 5, 7, 9 değerlerinden oluşan basit serinin momentlere dayalı basıklık ölçüsü kaçtır?
a) 0.02
b) 0.47
c) 0.93
d) 3.49
e) 4.56

Cevap Anahtarı

1.e, 2.b, 3.c, 4.c, 5.c, 6.c, 7.a, 8.c, 9.a, 10.c

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İ. H. (2018), İşletme İstatistiğine Giriş, 3.baskı, Alfa Yayınları, İstanbul.
- [2] Serper, Ö. (2014), Uygulamalı İstatistik-I, 7. Baskı, Filiz Kitabevi, İstanbul.
- [3] Başar, A., E. Oktay (2018), Uygulamalı İstatistik – I: Kısa Teorik Bilgiler ve Çözülmüş Problemler, 6. Baskı, Kültür ve Eğitim Kitap ve Kirtasiye, Erzurum.
- [4] Turanlı, M., S. Güriş (2018), Temel İstatistik, 6.Baskı Der Yayınları, İstanbul.
- [5] Güler, F. (2012), İstatistik Metotları ve Uygulamaları, 3. Baskı, Beta Yayınları, İstanbul.
- [6] Köksal, B. A. (2000), İstatistik Analiz Metodları, 6. Baskı, Çağlayan Kitabevi, İstanbul.
- [7] Berenson, M. L., D. M. Levine (1996), Basic Business Statistics: Concepts and Applications, 8. Baskı, Prentice Hall, Inc., New Jersey.

İNDEKSLER



İÇİNDEKİLER

- Basit İndeksler
- Bileşik İndeksler
 - Tartısız İndeksler
 - Tartılı İndeksler
- Mekân İndeksleri



HEDEFLER

- Bu üniteyi çalıştıktan sonra;
 - İktisadi göstergeleri daha iyi yorumlayıp karşılaştırılabilecek,
 - Çeşitli göstergelerin zaman içerisindeki seyrini daha iyi izleyebilecek,
 - Çeşitli göstergelerin mekânlar itibarıyla büyüklüklerini daha iyi karşılaştırılabilecek,
 - Büyük rakamlar arasında daha kolay karşılaştırmalar yapabileceksiniz.

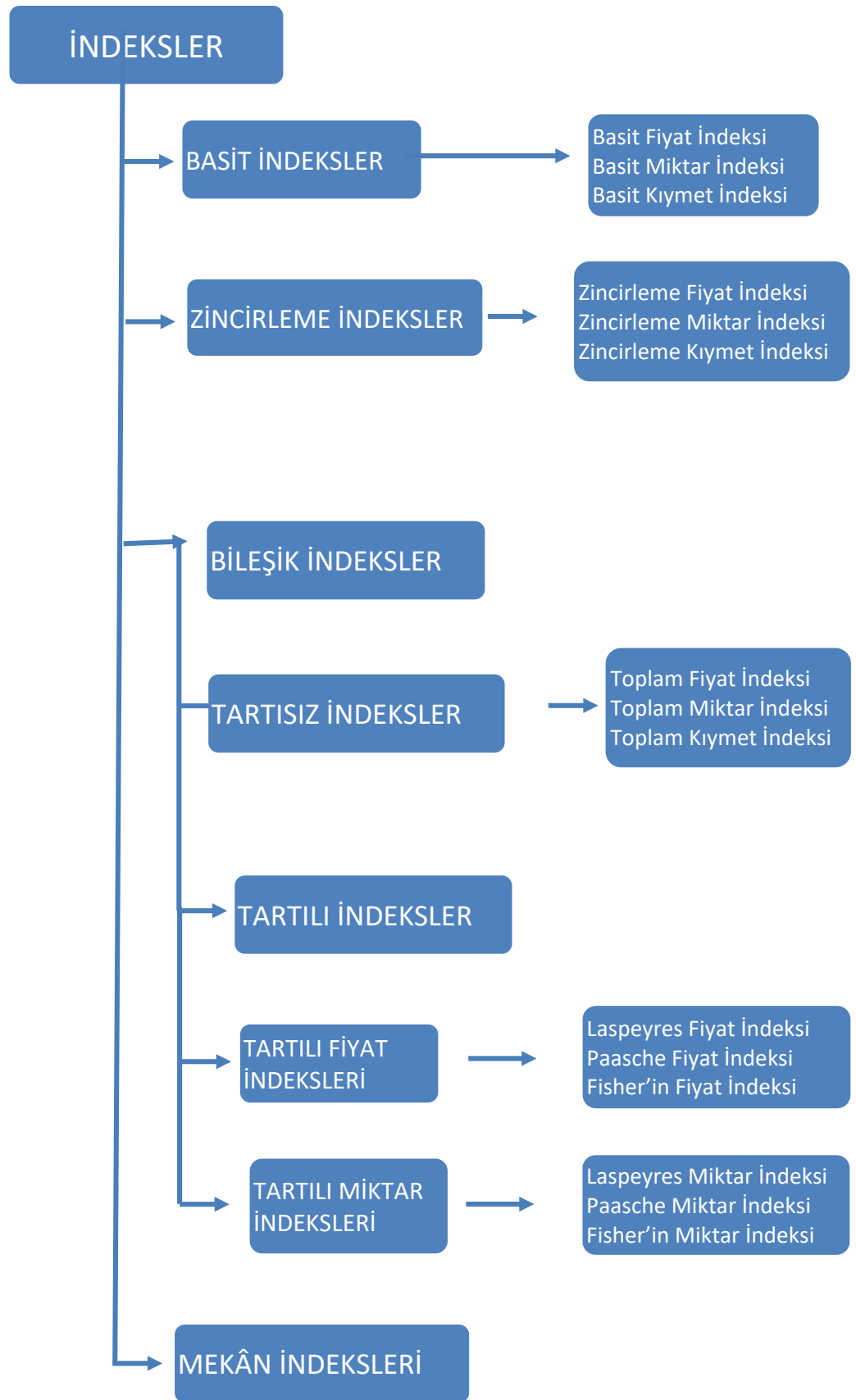


Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ

Prof. Dr. Erkan OKTAY

ÜNİTE 10



GİRİŞ

İndeksler, basit veya bileşik bir iktisadi olayın, zaman veya mekân itibarıyla gösterdiği değişimleri, bir nispet hâlinde ifade ederler. Değişikliklerin mutlak rakamlar yerine, yüzdeler hâlinde ifade edilmesi daha rahat anlaşılır [1]. Meselâ, “A malının üretimi 2023 yılında 256500 ton iken, 2024’te 275700 ton olmuştur.” demek yerine, 2023 yılı üretimini 100 kabul ederek 2024 yılındaki üretimi 107,5 olarak tespit ederiz. Bu şekilde 2024 yılı üretiminin 2023’e göre %7,5’lik bir artış gösterdiği hemen anlaşılır.



İndeksler, basit veya bileşik bir iktisadi olayın, zaman veya mekân itibarıyla gösterdiği yüzde değişimi ifade ederler.

Hem zaman serilerinden hem de mekân serilerinden indeksler hesaplanabilir. İndeksler temel yılın indeks puanı 100 olacak şekilde hesaplanır. Zaman serilerinden hesaplanan indeksler temel yıla göre değerlendirilir. Diğer yılların indeks puanları 100’ün üzerinde de olabilir altında da olabilir. 100’ün üzerindeki indeks sayıları temel yıla göre artışları ifade eder. Mesela indeks sayısı 135 olduğunda bu sayı söz konusu yılın değerinin temel yılın değerine göre %35 artış olduğunu gösterir. İndeks sayısı 100’ün altında olduğunda bu durum temel yıla göre söz konusu yılın değerinde azalma olduğunu gösterir. Mesela bir yılın indeks değeri 95 ise bu yılın değeri, temel yıl değerinden %5 daha az olduğunu ifade eder.

Mekân serilerinden de indeks sayıları elde edilebilir. Bu durumda mekân serisindeki değerlerin aritmetik ortalaması temel puan olarak kabul edilir. Mekân serisindeki bütün değerler temel puana (aritmetik ortalamaya) bölünüp 100 ile çarpılarak mekân serisindeki mekânlara ait indeksler elde edilir. 100’ün üzerindeki indeks sayıları ortalamaya göre artışları ifade eder. Mesela indeks sayısı 120 olduğunda bu sayı söz konusu mekânın değerinin ortalamadan %20 daha fazla olduğunu gösterir. İndeks sayısı 100’ün altında olduğunda bu durum, söz konusu mekânın değerinde ortalamaya göre azalma olduğunu gösterir. Mesela bir mekânın indeks sayısı 90 ise bu mekânın değeri mekânlar ortalamasının %10 altında olduğunu göstermektedir.

Zaman serisinden elde edilen indeksler herhangi bir mal veya hizmetin fiyat, miktar veya kıymetinde meydana gelen değişimi izlemede kullanılabileceği gibi, herhangi bir iktisadi göstergenin değerinde meydana gelen değişimi izlemede de kullanılabilir. Hatta birden fazla mal veya hizmetin fiyat, miktar veya kıymetinde meydana gelen değişimi izlemede de kullanılabilir.

BASİT İNDEKSLER

Basit indeksler; bir mal veya hizmetin fiyat, miktar veya kıymetindeki artış veya azalışı ifade eder.

Fiyat İndeksi: İndeksi hesaplanacak yılın fiyatı temel yılın fiyatına bölünerek 100 ile çarpılırsa basit fiyat indeksi elde edilir. İndeksi hesaplanacak dönemdeki fiyat p_n ve temel dönemdeki fiyat p_0 olmak üzere basit fiyat indeksi,

$$I = \left(\frac{p_n}{p_0} \right) 100$$



Basit indeksler, bir mal veya hizmetin fiyat, miktar veya kıymetindeki yüzde artış veya azalışı ifade eder.

formülüyle hesaplanır. [2]

Örnek: Türkiye’de 2021 yılından 2024 yılına kadar olan dönemde yıl sonları itibarıyla 24 ayar altının gram fiyatları aşağıdaki gibidir:

Yıllar	2021	2022	2023	2024
Altın Fiyatı (gr/TL)	772	1096	1954	2961

2021 yılı fiyatı esas alınır, zaman serisindeki bütün fiyatlar 2021 yılı altın fiyatına bölünerek 100 ile çarpılır. 2021 yılı temel alındığında fiyat indeksleri,

Yıllar	2021	2022	2023	2024
Altın Fiyat İndeksi	100	142	253	384

şeklinde hesaplanır. Temel yıl indeksi 100 olarak kabul edildiği için 100’ün üzerindeki indeks sayıları artışları, 100’ün altındaki indeks sayıları ise azalışları ifade eder. Buna göre 2021 yılı temel yıl kabul edilirse Türkiye’de altının gram fiyatı 2022 yılında %42 artmıştır. Temel yıl 2021 olduğunda Türkiye’de altının gram fiyatı 2023 yılında %153 artmıştır. Temel yıl 2021 olduğunda altının gram fiyatı %284 artmıştır. %100’den fazla artışları temel yıl fiyatlarının katlarıyla ifade etmek de mümkündür. Mesela 2024 yılı gram altın fiyatları 2021 yılına göre yaklaşık 3,8 kat artmıştır denilebilir.

Miktar İndeksi: İndeksi hesaplanacak yıldaki miktar temel yıldaki miktara bölünerek 100 ile çarpılırsa basit miktar indeksi hesaplanır. İndeksi hesaplanacak dönemdeki miktar q_n ve temel dönemdeki miktar q_0 olmak üzere basit miktar indeksi,

$$I = \left(\frac{q_n}{q_0} \right) 100$$

formülü ile hesaplanır [3].

Örnek: Bir otomobil fabrikasının son 4 yılda ürettiği araba sayısı aşağıdaki gibidir:

Yıllar	2021	2022	2023	2024
Oto Adedi	2450	2550	3500	4000

2021 yılındaki üretim miktarını esas aldığımızda diğer yıllara ait miktar indeksleri aşağıdaki gibi olur:

Yıllar	2021	2022	2023	2024
Otomobil Miktar İndeksi	100	104	143	163

2021 yılında üretilen araba miktarına göre 2022’de %4, 2023’de %43 ve 2024’te %63 oranında üretim artışı olduğunu söyleyebiliriz.



Zaman serisindeki fiyat verilerinden fiyat indeksi, miktar verilerinden miktar indeksi hesaplanır.

Kıymet İndeksi: Bir mal veya hizmetin kıymeti, fiyat ve miktarının çarpımıyla elde edilir. Kıymet indeksi,

$$I = \left(\frac{P_n Q_n}{P_0 Q_0} \right) 100$$

formülüyle hesaplanır.[1]



Bir malın fiyat ve miktarının çarpımıyla söz konusu malın kıymeti elde edilir. Kıymet verilerinden kıymet indeksi hesaplanır.

Örnek: Bir otomobil fabrikasında üretilen belirli bir tip otomobilin yıllar itibarıyla üretim miktarları ve fiyatları aşağıdaki gibidir. Üretim miktarları ile otomobilin birim fiyatları çarpılarak kıymet serisi oluşturulur.

Yıllar	2021	2022	2023	2024
Üretim Miktarı	2380	2450	2550	3500
Fiyatı (1.000 TL)	42	45	50	52
Kıymet (1.000 TL)	99.960	110.250	127.500	182.000

2021 yılı esas alındığında her yılın kıymeti 2021 yılındaki kıymete bölünerek 100 ile çarpıldığında kıymet indeksleri aşağıdaki gibi bulunur:

Yıllar	2021	2022	2023	2024
Kıymet İndeksi	100	110	128	182

2021 yılı esas alındığında söz konusu otomobil fabrikasında üretilen otomobillerin kıymeti 2022 yılında %10, 2023'te %28 ve 2024'te %82 artmıştır.

Zincirleme İndeks: Bir zaman serisinde zincirleme fiyat, miktar veya kıymet indeksini hesaplarken; indeksi hesaplanacak yıldaki fiyat, miktar veya kıymet bir önceki dönemin fiyat, miktar veya kıymetine bölünerek 100 ile çarpılır [2].

Örnek: Türkiye’de 2021 yılından 2011 yılına kadar olan dönemde yıl sonları itibarıyla 24 ayar altının gram fiyatı aşağıdaki gibidir:

Yıllar	2021	2022	2023	2024
Altın Fiyatı (gr/TL)	42,2772	52,9109 6	1954	95,4296 1

Her yılın altın fiyatı kendisinden bir önceki yılın fiyatına bölünüp 100 ile çarpıldığında zincirleme fiyat indeksi,

Yıllar	2022	2023	2024
Zincirleme Fiyat İndeksi	125142	133178	136152



Zincirleme indeksler, bir malın fiyat, miktar veya kıymetinin bir önceki yıla göre yüzde değişimini gösterir.

şeklinde hesaplanır. Buna göre 2022 yılındaki altın fiyatı 2021'e göre %42 artmışken, 2023 yılındaki altın fiyatı 2022'ye göre %78 ve 2024 yılındaki altın fiyatı ise 2023'e göre %52 artmıştır.

BİLEŞİK İNDEKSLER

Birden fazla mal veya hizmet kaleminin fiyat, miktar veya kıymetindeki yüzde değişmeyi incelemek istediğimizde bileşik indekslerden yararlanılır.

Tartısız İndeksler

Toplam Fiyat İndeksi

Birden fazla mal veya hizmet kaleminin fiyatındaki nispi değişimi incelemek istediğimizde bu indeksi kullanırız. Tartısız toplam fiyat indeksi,

$$I = \left(\frac{\sum p_n}{\sum p_0} \right) 100$$

formülü ile hesaplanır. Yani indeksi hesaplanacak dönemdeki mal veya hizmet fiyatlarının toplamı, temel dönemdeki mal veya hizmet fiyatları toplamına bölünerek 100 ile çarpılır [4].

Örnek: Beş yağ çeşidinin yıllık ortalama fiyatları aşağıdaki gibidir. Bu verilere dayanarak yağ için toplam fiyat indisini hesaplayalım.

Yağın Türü	2021	2022	2023	2024
Tereyağı	100	150	200	250
Ayçiçek yağı	30	50	60	75
Zeytinyağı	75	100	150	200
Yemeklik Margarin	50	60	70	80
Kahvaltılık Margarin	60	75	80	100
Toplam	315	435	560	705

2021 yılı esas alındığında diğer yılların toplam yağ fiyatları 2021 yılı toplam yağ fiyatına bölünerek 100 ile çarpılır ve toplam yağ fiyatları indeksi aşağıdaki gibi hesaplanır:

Yıllar	2021	2022	2023	2024
Toplam Fiyat İndeksi	100	138	178	224

Toplam Miktar İndeksi

Birden fazla mal veya hizmet kaleminin miktarındaki nispi değişimi incelemek istediğimizde bu indeks kullanılır. Tartısız toplam miktar indeksi,



Bileşik indeksler, birden fazla mal veya hizmet kaleminin fiyat, miktar veya kıymetindeki yüzde değişimi gösterir.



Toplam fiyat indeksi, birden fazla mal veya hizmet kaleminin fiyatındaki değişimi gösterir.



Toplam miktar indeksi, birden fazla mal veya hizmet kaleminin miktarındaki yüzde değişimi gösterir.

$$I = \left(\frac{\sum q_n}{\sum q_0} \right) 100$$

formülü yardımıyla hesaplanır. Yani indeksi hesaplanacak dönemdeki mal veya hizmet miktarlarının toplamı, temel dönemdeki mal veya hizmet miktarları toplamına bölünerek 100 ile çarpılır [1].

Örnek: Yıllar itibarıyla otomotiv sanayiinde üretilen motorlu araç sayısı aşağıdaki gibidir:

Araç Türü	2021	2022	2023	2024
Kamyon	12.766	11.581	16.679	16.906
Kamyonet	7.282	7.464	10.355	13.061
Otomobil	120.796	118.095	166.222	195.599
Otobüs	1.042	931	1.663	1.017
Minibüs ve Midibüs	7.868	7.915	12.668	14.567
Traktör	31.327	19.602	30.739	21.964
Toplam	181.081	165.588	238.326	263.114

2021 yılında üretilen toplam motorlu araç sayısı esas alındığında, diğer yıllarda üretilen toplam motorlu araç sayıları 2021 yılı değerine bölünerek 100 ile çarpılır ve toplam motorlu araç miktar indeksleri,

Yıllar	2021	2022	2023	2024
Motorlu Araçlar Miktar İndeksi	100	91	132	145

şeklinde elde edilir. 100'ün altındaki indeks sayıları azalışları ifade eder. 2022 yılı motorlu taşıt sayısı 2021 yılına göre %9 azalmıştır. Bununla birlikte 2021 yılına göre 2023 yılı motorlu taşıt sayısı %32 ve 2024 yılı motorlu taşıt sayısı %45 artmıştır.



Toplam kıymet indeksi, birden fazla mal veya hizmet kaleminin kıymetindeki yüzde değişimi gösterir.

Toplam Kıymet İndeksi

Birden fazla mal veya hizmet kaleminin kıymetindeki nispi değişimi incelemek istediğimizde bu indeksi kullanırız. Kıymet indeksi,

$$I = \left(\frac{\sum p_n q_n}{\sum p_0 q_0} \right) 100$$

formülü ile hesaplanır. Yani, indeksi hesaplanacak dönemdeki mal veya hizmet kıymetleri toplamı, temel dönemdeki mal veya hizmet kıymetleri toplamına bölünerek 100 ile çarpılır [2].

Örnek: 2008-2011 yılları itibarıyla bir ülkedeki arpa, buğday ve çavdar üretim miktarları (ton) ve satış fiyatları (1.000 TL) aşağıdaki tabloda gösterilmiştir:

	2021		2022		2023		2024	
Mallar	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar
Arpa	18.3	64.0	22.3	60.3	24.2	62.2	25.3	63.1
Buğday	24.0	75.0	28.7	88.7	32.0	91.0	38.7	94.4
Çavdar	9.8	12.1	11.4	14.4	13.9	17.1	14.4	18.3

2021 yılı temel alınarak 2022, 2023 ve 2024 yıllarına ait kıymet indeksi hesaplanırken,

2021 yılı fiyatı, p_0 , 2022, 2023 ve 2024 yılları fiyatları, p_n ,

2021 yılı miktarı, q_0 , 2022, 2023 ve 2024 yılları miktarları, q_n

ile gösterilir. 2008 yılı temel yıl olarak kabul edilirse 2009 yılına ait kıymet indeksi,

$$I_{2022} = \left(\frac{22.3(60.3) + 28.7(88.7) + 11.4(14.4)}{18.3(64) + 24(75) + 9.8(12.1)} \right) 100 = 131$$

olarak hesaplanır. Belirtilen üç tahıl cinsinin toplam değeri 2021'e göre 2022 yılında %31 artmıştır. Benzer şekilde 2021 yılı temel yıl olarak kabul edilirse 2023 ve 2024 yıllarına ait kıymet indeksi,

$$I_{2023} = \left(\frac{24.2(62.2) + 32(91) + 13.9(17.1)}{18.3(64) + 24(75) + 9.8(12.1)} \right) 100 = 151$$

$$I_{2024} = \left(\frac{25.3(63.1) + 38.7(94.4) + 14.4(18.3)}{18.3(64) + 24(75) + 9.8(12.1)} \right) 100 = 178$$

olarak hesaplanır. 2021 yılı temel yıl olarak kabul edildiğinde belirtilen tahıl ürünlerinin kıymeti 2023 yılında %51 ve 2024 yılda %78 artış göstermiştir.

Tartılı İndeksler

Fiyat indeksleri hesaplanırken miktarlar, miktar indeksleri hesaplanırken fiyatlar tartı olarak kullanılır.

Tartılı fiyat indeksleri

Laspeyres Fiyat İndeksi: Bir mal veya hizmet grubuna ilişkin fiyat indeksini hesaplarken, temel dönem miktarlarını tartı olarak alırsak Laspeyres fiyat indeksini elde ederiz. Bu indeksi,



Fiyat indeksi hesaplanırken miktar, miktar indeksi hesaplanırken fiyat tartı olarak kullanılır.

$$I = \left(\frac{\sum p_n q_0}{\sum p_0 q_0} \right) 100$$

formülü yardımıyla hesaplarız [3].

Örnek: Kıymet indeksi konusunda verilen örneğe dönelim. 2021 yılını temel alarak 2022, 2023 ve 2024 yıllarına ait Laspeyres fiyat indeksini,

$$I_{2022} = \left(\frac{22.3(64) + 28.7(75) + 11.4(12.1)}{18.3(64) + 24(75) + 9.8(12.1)} \right) 100 = 120$$

$$I_{2023} = \left(\frac{24.2(64) + 32(75) + 13.9(12.1)}{18.3(64) + 24(75) + 9.8(12.1)} \right) 100 = 133$$

$$I_{2024} = \left(\frac{25.3(64) + 38.7(75) + 14.4(12.1)}{18.3(64) + 24(75) + 9.8(12.1)} \right) 100 = 152$$

olarak hesaplarız. Belirtilen üç tahıl cinsinin fiyatları 2021 yılına göre 2022’de %20, 2023’te %33 ve 2024’te %52 artmıştır.

Paasche Fiyat İndeksi: Bir mal veya hizmet grubuna ilişkin fiyat indeksini hesaplarken, indeksi hesaplanacak dönemdeki miktarları tartı olarak kullanırsak Paasche fiyat indeksini elde ederiz. Paasche fiyat indeksinin formülü,

$$I = \left(\frac{\sum p_n q_n}{\sum p_0 q_n} \right) 100$$

şeklindedir [4].

Örnek: Kıymet indeksi konusunda verilen örneğin verilerine bakalım. Kıymet indeksi konusunda verilen örneğe dönelim. 2021 yılını temel alarak 2022, 2023 ve 2024 yıllarına ait Paasche fiyat indeksini,

$$I_{2022} = \left(\frac{22.3(60.3) + 28.7(88.7) + 11.4(14.4)}{18.3(60.3) + 24(88.7) + 9.8(14.4)} \right) 100 = 120$$

$$I_{2023} = \left(\frac{24.2(62.2) + 32(91) + 13.9(17.1)}{18.3(62.2) + 24(91) + 9.8(17.1)} \right) 100 = 133$$

$$I_{2024} = \left(\frac{25.3(63.1) + 38.7(94.4) + 14.4(18.3)}{18.3(63.1) + 24(94.4) + 9.8(18.3)} \right) 100 = 153$$

olarak hesaplarız. Belirtilen üç tahıl cinsinin fiyatları 2021 yılına göre 2022’de %20, 2023’te %33 ve 2024’te %53 artmıştır.

Fisher’in İdeal Fiyat İndeksi: Laspeyres fiyat indeksi ile Paasche fiyat indekslerinin geometrik ortalaması alındığında Fisher’in ideal fiyat indeksi hesaplanır. Fisher’in ideal fiyat indeksi,



Laspeyres fiyat indeksi hesaplanırken indeksi temel yılın miktarı tartı olarak kullanılır.



Paasche fiyat indeksi hesaplanırken indeksi hesaplanacak yılın miktarı tartı olarak kullanılır.



Fisher'in ideal fiyat indeksi, Laspeyres fiyat indeksi ile Paasche fiyat indeksinin geometrik ortalamasıdır.

$$I = \sqrt{\left(\frac{\sum p_n q_0}{\sum p_0 q_0} \right) \left(\frac{\sum p_n q_n}{\sum p_0 q_n} \right)} 100$$

formülü ile hesaplanır [5].

Örnek: Kıymet indeksi konusunda verilen örneğin verilerine bakalım. 2021 yılını temel aldığımızda 2022, 2023 ve 2024 yıllarına ait Paasche ve Laspeyres fiyat indeksleri,

İndeksler	2022	2023	2024
Paasche Fiyat İndeksi	120	133	152
Laspeyres Fiyat İndeksi	120	133	153

şeklinde hesaplanmıştı. Her yılın Laspeyres ve Paasche indeksleri çarpılıp karekökü alınırsa 2022 yılı ideal fiyat indeksi 120, 2023 yılı ideal fiyat indeksi 133 ve 2024 yılı ideal fiyat indeksi 153 olarak hesaplanır. Bu sonuçlara göre 2021 yılı temel alındığında belirtilen üç tahıl cinsinin fiyatları 2022'de %20, 2023'te %33 ve 2024'te %53 artmıştır.



Laspeyres miktar indeksi hesaplanırken indeksi temel yılın fiyatı tartı olarak kullanılır.

Tartılı miktar indeksleri

Laspeyres Miktar İndeksi: Bir mal veya hizmet grubuna ilişkin miktar indeksini hesaplarken, temel dönem fiyatlarını tartı olarak alırsak Laspeyres miktar indeksini elde ederiz. Bu indeksi,

$$I = \left(\frac{\sum q_n p_0}{\sum q_0 p_0} \right) 100$$

formülü yardımıyla hesaplarız [6].

Örnek: Kıymet indeksi konusunda verilen örneğe geri dönelim. 2021 yılını temel alarak 2022, 2023 ve 2024 yıllarına ait Laspeyres miktar indeksini,

$$I_{2022} = \left(\frac{60.3(18.3) + 88.7(24) + 14.4(9.8)}{64(18.3) + 75(24) + 12.1(9.8)} \right) 100 = 109$$

$$I_{2023} = \left(\frac{62.2(18.3) + 91(24) + 17.1(9.8)}{64(18.3) + 75(24) + 12.1(9.8)} \right) 100 = 113$$

$$I_{2024} = \left(\frac{63.1(18.3) + 94.4(24) + 18.3(9.8)}{64(18.3) + 75(24) + 12.1(9.8)} \right) 100 = 117$$

olarak hesaplarız. Belirtilen üç tahıl cinsinin miktarları 2021 yılına göre 2022'de %9, 2023'te %13 ve 2024'te %17 artmıştır.

Paasche Miktar İndeksi: Bir mal veya hizmet grubuna ilişkin miktar indeksini hesaplarken, indeksi hesaplanacak dönemdeki fiyatları tartı olarak kullanırsak Paasche miktar indeksini elde ederiz. Paasche miktar indeksinin formülü,



Paasche miktar indeksi hesaplanırken indeksi hesaplanacak yılın fiyatı tartı olarak kullanılır.

$$I = \left(\frac{\sum q_n p_n}{\sum q_0 p_n} \right) 100$$

şeklindedir [3].

Örnek: Kıymet indeksi konusunda verilen örneğin verilerine tekrar göz atalım. 2021 yılını temel alarak 2022, 2023 ve 2024 yıllarına ait Paasche miktar indeksini,

$$I_{2022} = \left(\frac{60.3(22.3) + 88.7(28.7) + 14.4(11.4)}{64(22.3) + 75(28.7) + 12.1(11.4)} \right) 100 = 109$$

$$I_{2023} = \left(\frac{62.2(24.2) + 91(32) + 17.1(13.9)}{64(24.2) + 75(32) + 12.1(13.9)} \right) 100 = 113$$

$$I_{2024} = \left(\frac{63.1(25.3) + 94.4(38.7) + 18.3(14.4)}{64(25.3) + 75(38.7) + 12.1(14.4)} \right) 100 = 117$$

olarak hesaplarız. Belirtilen üç tahıl cinsinin miktarları 2021 yılına göre 2022’de %9, 2023’te %13 ve 2024’te %17 artmıştır.

Fisher’in İdeal Miktar İndeksi: Bu indeks, Laspeyres miktar indeksi ile Paasche miktar indekslerinin geometrik ortalamasıdır. Fisher’in ideal miktar indeksi,

$$I = \sqrt{\left(\frac{\sum q_n p_0}{\sum q_0 p_0} \right) \left(\frac{\sum q_n p_n}{\sum q_0 p_n} \right)} 100$$

formülü ile hesaplanır [4].

Örnek: Kıymet indeksi konusunda verilen örneğin verilerine bakalım. 2021 yılını temel aldığımızda 2022, 2023 ve 2024 yıllarına ait Paasche ve Laspeyres miktar indeksleri,

İndeksler	2022	2023	2024
Paasche Miktar İndeksi	109	113	117
Laspeyres Miktar İndeksi	109	113	117

şeklinde hesaplanmıştı. Her yılın Laspeyres ve Paasche miktar indeksleri çarpılıp karekökü alınırsa 2022 yılı ideal miktar indeksi 109, 2023 yılı ideal miktar indeksi 113 ve 2024 yılı ideal miktar indeksi 117 olarak hesaplanır. Bu sonuçlara göre 2021 yılı temel alındığında belirtilen üç tahıl cinsinin miktarları 2022’de %9, 2023’te %13 ve 2024’te %17 artmıştır.

MEKÂN İNDEKSLERİ

Bir mal veya hizmete ait fiyat, miktar veya kıymetin yerleşim merkezleri itibarıyla gösterdiği nispi değişimi ortaya koymak için hazırlanan indekslere mekân indeksleri denir. Mekân indeksinin hesaplanmasında ilk safha, ilgili mekânlardaki



Mekân serilerinden elde edilen indekslere mekân indeksleri denir.

ölçüm değerlerinin ortalamasını bulmaktır. Daha sonra yerleşim merkezlerindeki fiyat, miktar veya kıymet ölçümleri bu ortalama bölünerek 100 ile çarpılır. Mekân indeksleri,

$$I = \left(\frac{X_i}{\bar{X}} \right) 100$$

formülü yardımıyla elde edilir. Formüldeki X_i indeksi hesaplanacak mekâna ait fiyat, miktar veya kıymeti ifade eder [6].

Örnek: 2024 yılında altı ilimizde meydana gelen trafik kazalarında ölenlerin sayısı aşağıdaki gibidir:

İller	Trafik Kazası
Ankara	325
İstanbul	315
Konya	300
Antalya	290
İzmir	223
Erzurum	34
Toplam	1487



Mekân serisindeki değerler, serideki değerlerin aritmetik ortalamasına bölünerek 100 ile çarpılır. Buna mekân indeksleri denir.

Altı ildeki trafik kazalarında ölen kişi sayısı il sayısına böldüğümüzde; il bazında trafik kazalarında ortalama kişi sayısı 247.83 olarak bulunur. İllerde meydana gelen trafik kazalarında ölen kişi sayıları bu ortalama bölünerek 100 ile çarpılırsa, illerde meydana genel trafik kazalarında ölenlerin indeksleri,

İller	Trafik Kazası İndeksi
Ankara	131
İstanbul	127
Konya	121
Antalya	117
İzmir	90
Erzurum	14

şeklinde hesaplanır. Bu sonuçlara göre 2024 yılında Ankara'da meydana gelen trafik kazalarında ölenlerin sayısı verilen altı ilin ortalamasından %31 daha fazladır. 2024 yılında Erzurum'da meydana gelen trafik kazalarında ölenlerin sayısı ise verilen illerin ortalamasından %86 daha azdır.



Bireysel Etkinlik

- Türkiye'de en fazla üretilen üç bitkisel yağ çeşidinin son iki yıldaki üretim miktarları ve fiyatlarını kullanarak tartılı fiyat indekslerini hesaplayınız ve yorumlayınız.
- Türkiye'de son on yılda gerçekleşen enflasyon oranlarını kullanarak zincirleme indeksleri elde ediniz. Böylece her yılın enflasyon oranının bir önceki yıla göre % kaç artış veya azalış gösterdiğini belirtiniz.
- Tartısız indekslerin dezavantajları neler olabilir? Tartılı indeklerdeki tartıların avantaj veya dezavantajları tartışınız.



Özet

- İndeksler, basit veya bileşik bir iktisadi olayın, zaman veya mekân itibarıyla gösterdiği değişimleri, bir nispet hâlinde ifade ederler. Değişikliklerin mutlak rakamlar yerine, yüzdeler hâlinde ifade edilmesi daha rahat anlaşılır. Meselâ, “A malının üretimi 2010 yılında 256500 ton iken, 2011’de 275700 ton olmuştur.” demek yerine, 2010 yılı üretimini 100 kabul ederek 2011 yılındaki üretimi 107,5 olarak tespit ederiz. Bu şekilde 2011 yılı üretiminin 2010’a göre %7,5’lik bir artış gösterdiği hemen anlaşılır.
- İndeksler zaman serilerinde temel yıla, mekân serilerinde ise ortalama değere göre üzerinde durulan dönemin veya mekânın değerdeki yüzde değişimi gösterir. 100’ün üzerindeki indeks sayıları temel değere göre artışları gösterir. Mesela indeks sayısı 110 olduğunda 100’den fazla kısım temel değere göre %10’luk artışın olduğunu gösterir. 100’ün altındaki indeks sayıları temel değere göre azalışları gösterir. Mesela indeks sayısı 90 olduğunda bunun 100’e tamamlayan kısım temel değere göre %10’luk azalışın olduğunu gösterir.
- Zaman serisinden elde edilen indeksler, herhangi bir mal veya hizmetin fiyat, miktar veya kıymetinde meydana gelen değişimi izleme kullanılabileceği gibi herhangi bir iktisadi göstergenin değerinde meydana gelen değişimi izlemede de kullanılabilir. Hatta birden fazla mal veya hizmetin fiyat, miktar veya kıymetinde meydana gelen değişimi izleme de kullanılabilir.
- Zaman serilerinde ilk yıl genellikle temel yıl olarak kabul edilir. Diğer yılların değerleri temel yıl değerine bölünüp 100 ile çarpıldığında diğer yıllara ait indeks değerleri hesaplanmış olur. Zaman serisinde her yılın değeri kendisinden bir önceki yılın değerine bölünüp 100 ile çarpıldığında zincirleme indeksler hesaplanmış olur.
- Bu ünite okunduğunda tek bir malın, hizmetin veya bir iktisadi göstergenin fiyat, miktar veya kıymetindeki yüzde değişim hesaplanabileceği (basit indeksler) gibi birden fazla mal veya hizmetin veya iktisadi göstergenin fiyat, miktar veya değerindeki yüzde değişim (bileşik indeksler) de hesaplanabilir.
- Bileşik indeksler hem tartılı hem de tartısız olarak hesaplanabilir. Tartısız bileşik indeksler hesaplanırken yalnızca indeksin hesaplanmasına konu olaran değerler kullanılır. Birden fazla mal veya hizmetin fiyatındaki değişimi gösteren indekslere bileşik fiyat indeksi denir. Birden fazla mal veya hizmetin miktarındaki değişimi gösteren indekslere bileşik miktar indeksi denir. Birden fazla mal veya hizmetin kıymetindeki değişimi gösteren indekslere bileşik kıymet indeksi denir.
- Tartılı bileşik fiyat indekslerinin hesaplanmasında miktarlar tartı olarak kullanılırken, tartılı bileşik miktar indekslerinin hesaplanmasında fiyatlar tartı olarak kullanılır. Laspeyres fiyat indeksinin hesaplanmasında temel yıl miktarları tartı olarak kullanılır. Paasche fiyat indeksinin hesaplanmasında indeksi hesaplanacak yılın miktarları tartı olarak kullanılır. Fisher’in ideal fiyat indeksi Paasche fiyat indeksi ve Laspeyres fiyat indeksinin geometrik ortalamasıdır. Laspeyres miktar indeksinin hesaplanmasında temel yıl fiyatları tartı olarak kullanılır. Paasche miktar indeksinin hesaplanmasında indeksi hesaplanacak yılın fiyatları tartı olarak kullanılır. Fisher’in ideal miktar indeksi Paasche miktar indeksi ve Laspeyres miktar indeksinin geometrik ortalamasıdır.
- Mekân serilerinden elde edilen indekslere mekân indeksi denir. Mekân indeksleri bir mekâna ait değerlerin mekân değerleri ortalamasına bölünmesiyle elde edilir. Mekân indeksi, bir mekândaki iktisadi mal veya hizmetin veya iktisadi göstergenin mekân değerleri ortalamasının yüzde kaç altında veya üstünde olduğunu gösterir.

DEĞERLENDİRME SORULARI

1. 2023 yılında beyaz peynirin kg fiyatı 60 TL iken 2024 yılında 110 TL olmuştur. 2023 yılı temel kabul edilirse 2024 yılında peynirin fiyat indeksi kaç olur?
a) 110
b) 120
c) 144
d) 183
e) 210
2. 2023 yılında beyaz peynirin kg fiyatı 60 TL iken 2024 yılında peynirin fiyatı 120 TL olmuştur. 2023 yılı temel kabul edilirse 2024 yılında peynirin fiyatı % kaç artmıştır?
a) 40
b) 50
c) 83
d) 100
e) 200
3. 2023 yılında beyaz peynirin kg fiyatı 60 TL iken 2024 yılında peynirin fiyatı 150 TL olmuştur. 2023 yılı temel kabul edilirse 2024 yılında peynirin fiyatı kaç kat artmıştır?
a) 1
b) 2
c) 2.5
d) 3
e) 4

Yıllar	2021	2022	2023	2024
Tereyağı (Ton)	210	250	330	404

4. Yukarıda yıllar itibarıyla bir yöremizde üretilen tereyağı miktarı (ton) verilmiştir. 2021 yılı temel kabul edilirse 2024 yılında tereyağı miktar indeksi kaç olur?
a) 52
b) 112
c) 130
d) 192
e) 209

Yıllar	2021	2022	2023	2024
Tereyağı (Ton)	210	250	330	404

5. Yukarıdaki tabloda yıllar itibarıyla bir yöremizde üretilen tereyağı miktarı (ton) verilmiştir. 2024 yılı için tereyağına ait zincirleme miktar indeksi kaç olur?
- 109
 - 122
 - 133
 - 142
 - 159

Yıllar	2021	2022	2023	2024
Tereyağı Fiyatı	5.0	5.5	5.9	6.3
Tereyağı Miktarı	210	250	330	404

6. Yukarıdaki tabloda yıllar itibarıyla bir yöremizde üretilen tereyağı fiyatı (\$/kg) ve miktarı (ton) verilmiştir. 2021 yılı temel kabul edilirse 2024 yılı tereyağı kıymet indeksi kaç olur?
- 199
 - 242
 - 254
 - 272
 - 289

Yıllar	2021	2022	2023	2024
Süt (\$/lt)	0.4	0.6	0.7	0.8
Tereyağı (\$/kg)	5.0	5.5	5.9	6.3
Peynir (\$/kg)	4.8	5.1	5.2	5.5
Yoğurt (\$/kg)	1.8	2.0	2.1	2.3

7. Yukarıdaki tabloda yıllar itibarıyla bir yöremizde satılan süt ve süt mamullerinin fiyatları verilmiştir. 2021 yılı temel kabul edilirse 2024 yılı süt ve süt mamulleri toplam fiyat indeksi kaç olur?
- 101
 - 114
 - 124
 - 132
 - 149

(8-10. soruları vakaya göre cevaplayınız.)

Vaka: Yıllar itibarıyla bir yöremizde satılan süt ve süt mamullerinin fiyatları ve miktarları aşağıdaki gibidir.

Yıllar/ Süt Ürünleri	2023		2024	
	Fiyat (\$)	Miktar (Ton)	Fiyat (\$)	Miktar (Ton)
Süt	0.4	80	0.7	90
Tereyağı	5.0	20	5.9	30
Peynir	4.8	40	5.2	50

8. 2023 yılı temel kabul edilirse 2024 yılına ait Laspeyres fiyat indeksi kaç olur?
- a) 111
b) 118
c) 124
d) 129
e) 136
9. 2023 yılı temel kabul edilirse 2024 yılına ait Paasche fiyat indeksi kaç olur?
- a) 101
b) 114
c) 117
d) 122
e) 129
10. 2023 yılı temel kabul edilirse 2024 yılına ait Fisher'in ideal fiyat indeksi kaç olur?
- a) 109
b) 111
c) 114
d) 118
e) 123

Cevap Anahtarı

1.e, 2.d, 3.c, 4.d, 5.b, 6.b, 7.c, 8.b, 9.c, 10.d

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İsmail Hakkı (2018). İşletmelerde Uygulamalı İstatistik (3. baskı). İstanbul: Alfa Yayınları.
- [2] Başar, Alaaddin, Erkan Oktay (2018). Uygulamalı İstatistik – II: Kısa Teorik Bilgiler ve Çözölmüş Problemler (9. baskı), Erzurum: Erzurum Kültür ve Eğitim Kitap Kirtasiye.
- [3] Daniel, Wayne W., James C. Terrell (1995). Business Statistics: For Management and Economics (7. baskı). Boston: Houghton Mifflin Company.
- [4] Gürtan, Kenan (1982). İstatistik ve Araştırma Metotları. İstanbul: İstanbul Üniversitesi Yayınları.
- [5] Köksal, Bilge Aloba (2003). İstatistik Analiz Metodları (6. baskı). İstanbul: Çağlayan Kitabevi.
- [6] Serper, Özer (2014). Uygulamalı İstatistik 1 (7. baskı). İstanbul: Ezgi Kitabevi.

İHTİMAL TEORİSİ



İÇİNDEKİLER

- Temel Kavramlar
- Toplama Kuralı
- Çarpma Kuralı
- İhtimal Dağılım Tablosu
- Beklenen Değer
- Bazı İhtimal Kuralları



HEDEFLER

- Bu üniteyi çalıştıktan sonra;
- İhtimal (olasılık) kavramını anlayabilecek,
- Birbirini engelleyen birden fazla olayın meydana gelme ihtimalini hesaplayabilecek,
- Birbirini engellemeyen birden fazla olayın birlikte aynı anda meydana gelme ihtimalini hesaplayabilecek,
- Basit veya bileşik bir olayın meydana gelme beklentisini belirleyebileceksiniz.

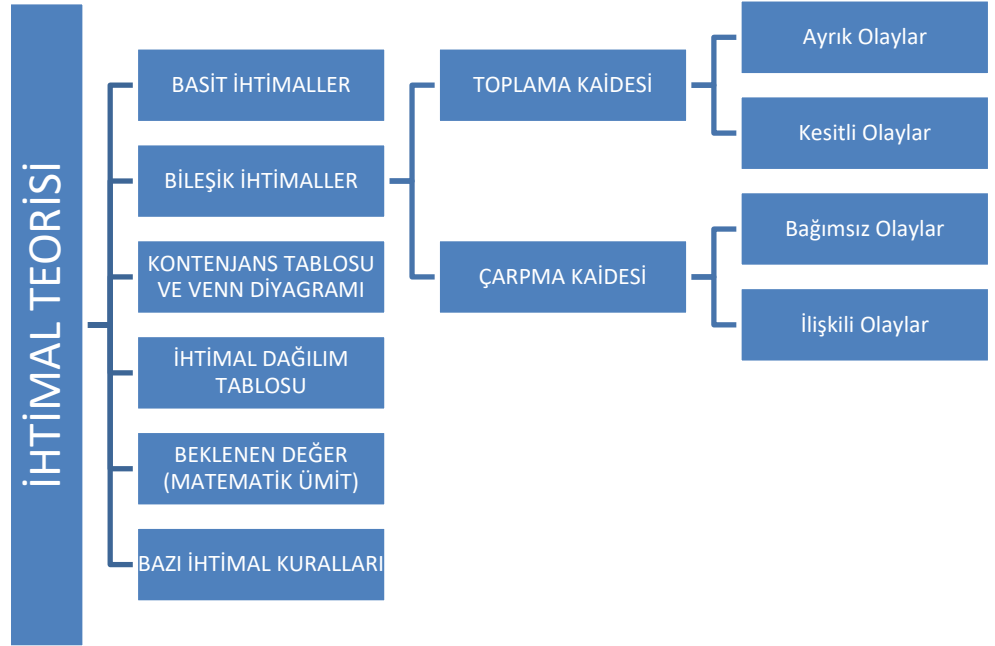


Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ

Prof. Dr. Erkan OKTAY

ÜNİTE 11



GİRİŞ

İhtimal (olasılık) kavramı hayatımızın her alanında karşımıza çıkabilecek olaylar için kullanılır. Bir olayın olması mümkün olduğu gibi olmaması da mümkün ise bu olay, ihtimale konu olan bir olaydır. “Meteoroloji verilerine göre yarın Erzurum il merkezinde büyük ihtimalle yağmur yağacak” denildiği zaman, yarın Erzurum’da yağmur yağmasının yağmamasına göre daha mümkün olduğu kastedilir [1].

N birim ihtiva eden bir anakütle içinde belli bir X özelliğini taşıyan n tane birim varsa, bu anakütleden rastgele bir birim alındığında bu birimin X özelliğini taşıması ihtimali,

$$P(X) = \frac{n}{N}$$

şeklinde hesaplanır. Bir X değişkeninin nispi frekansı ile bu olayın müşahade edilme ihtimali arasında yakın bir ilişki vardır. Bir olayın meydana gelme ihtimali 0 ile 1 arasında değişir. İhtimalin sıfır olması, söz konusu olayın meydana gelmesinin mümkün olmadığını, 1 olması ise olayın kesinlikle (yani %100) meydana geleceğini ifade eder. 0 ve 1 durumlarında ihtimalden bahsedilemez. 0’a yakın ihtimal zayıf ihtimal ve 1’e yakın ihtimal ise kuvvetli ihtimaldir [2].

Bir olayın mümkün bütün hâllerinin ihtimalleri toplamı 1’e eşittir. Mesela, bir olayın ancak A, B, C ve D gibi dört yolla meydana gelebilmesi ve bu yollara ait ihtimallerin de sırayla P_A , P_B , P_C ve P_D olması hâlinde,

$$P_A + P_B + P_C + P_D = 1$$

olur. Yine, X olayının meydana gelme ihtimalini p , meydana gelmeme ihtimalini,

$$1 - p = q$$

şeklinde tarif edersek,

$$p + q = 1$$

olur. Bu tür ihtimallere *birbirini tamamlayan ihtimaller* denir.

BASİT VE BİLEŞİK İHTİMALLER

Tek bir olayın sonuçları ile ilgili ihtimaller basit ihtimallerdir. Mesela; yarın yağmur yağması ihtimali, bir sınıftan tesadüfi olarak seçilen bir öğrencinin gözlüklü olması ihtimali gibi ihtimaller ayrı ayrı düşünüldüğünde basit birer ihtimaldir. İki veya daha fazla olayın birlikte vuku bulması ihtimali ise bileşik bir ihtimaldir. Aynı şekilde ikiden fazla olaydan bazılarının bazıları ile birlikte vuku bulması ihtimali de bileşik ihtimaldir. Bileşik ihtimal hesaplarına konu olan olaylar iki gruba ayrılır:

- Bir arada meydana gelebilen olaylar,
- Birbirini engelleyen olaylar [3].



Bir olayın olması mümkün olduğu gibi olmaması da mümkünse bu olay, ihtimale konu olan bir olaydır.



0 ile 1 arasındaki sayılar ihtimal sayıları olarak kabul edilebilir. 1’e yakın ihtimal sayıları, olması daha mümkün olayları gösterir.



İki ya da daha fazla olayın birlikte gerçekleşmesi durumu birleşik ihtimallerin konusudur.

ÖRNEK UZAYI



İstatistiki bir olayın mümkün olan bütün sonuçlarının oluşturduğu sete, örnek uzayı denir.

İstatistiki bir olayın mümkün olan bütün sonuçlarının oluşturduğu sete, örnek uzayı denir ve S ile gösterilir [3]. Örnek uzayındaki her bir sonuç, söz konusu örnek uzayının bir elemanıdır. Örnek uzayı sınırlı sayıda elemana sahipse, karışmamaları için elemanlar birbirlerinden virgülle ayrılıp parantez içerisinde gösterilebilir. Madeni bir para atıldığında mümkün olan iki sonuçla karşılaşacağımız için örnek uzayı,

$$S = \{Y, T\}$$

şeklinde yazılabilir. Rastgele seçilen bir mamul, ortalama gramajdan eksik veya fazla olabilir. Bu durumda örnek uzayı,

$$S = \{E, F\}$$

şeklinde bir zarın atılması deneyinde örnek uzayı,

$$S = \{1, 2, 3, 4, 5, 6\}$$

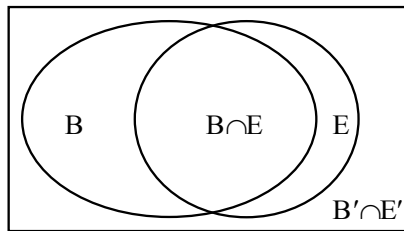
kümesinden oluşmaktadır.

KONTENJANS TABLOLARI VE VENN DİYAGRAMLARI

Örnek uzayını başlıca iki yolla gösterebiliriz. İncelenecek olayları çapraz sınıflandırma yoluyla kontenjans tablolarında gösteririz [4]. Mesela, memurlar arasında kredi kartı kullanımını yaygınlaştırmaya çalışan kredi kartı şirketleri bir yıl sonunda memurlar arasından tesadüfi olarak 200'ünü seçerek bunlara banka kredi kartı ve/veya seyahat ve eğlence kredi kartı kullanıp kullanmadıklarını sormuş olsunlar. Alınan cevaplar aşağıdaki kontenjans tablosunda gösterilebilir:

	Seyahat ve Eğlence Kredi Kartı	
Banka Kredi Kartı	Evet	Hayır
Evet	60	60
Hayır	15	65

Olayların birlikte ve ayrı ayrı vuku bulma durumlarını, Venn diyagramı dediğimiz iç içe daireler yardımıyla da gösterebiliriz. Kontenjans tablosunda gösterilebilen olayları Venn diyagramı yardımıyla da gösterebiliriz. Yalnız banka kredi kartına sahip olanlar, B ; yalnızca seyahat ve eğlence kredi kartına sahip olanlar, E ; hem banka hem de seyahat kredi kartına sahip olanlar, $B \cap E$; hem banka hem de seyahat kredi kartı olmayanlar ise $B' \cap E'$ ile gösterilmek üzere Venn diyagramı aşağıdaki gibi çizilebilir [4]:



İki olayın mümkün sonuçlarını kontenjans (iki yönlü tablo) tablolarında gösterebiliriz.



İkiden fazla olayın mümkün sonuçlarını Venn diyagramları ile gösterebiliriz.

B olayının gerçekleşme ihtimali $P(B)$, E olayının gerçekleşme ihtimali $P(E)$, B ve E olaylarının ikisinin birden gerçekleşmesi ihtimali $P(B \cap E)$, B ve E olaylarının ikisinin birden gerçekleşmeme ihtimali $P(B' \cap E')$, B veya E olaylarından birinin gerçekleşmesi ihtimali $P(B \cup E)$ 'dir.



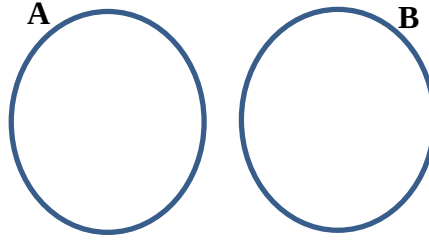
Birbirinden ayrık olaylardan birinin veya diğerinin meydana gelmesi ihtimali, olayların ayrı ayrı meydana gelmesi ihtimallerinin toplamına eşittir.

TOPLAMA KAİDESİ

$X_1, X_2, \dots, X_n$ birbirini engelleyen n tane olay ve bu olayların meydana gelme ihtimalleri de sırayla $P_1, P_2, \dots, P_n$ olmak üzere, bu olaylardan birinin veya diğerinin meydana gelme ihtimali,

$$P_1 + P_2 + \dots + P_n$$

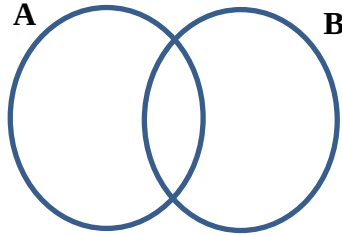
olur. Bu kaideye **toplama kaidesi** denir [5]. Toplama kaidesine konu olan olaylar ikiye ayrılır: Birbirinden ayrık olaylar, birbirinden ayrık olmayan veya ara kesitli olaylar.



Birlikte meydana gelmeleri mümkün olmayan (birbirinden ayrık) A ve B gibi iki olaydan birinin veya diğerinin meydana gelmesi ihtimali,

$$P(A \cup B) = P(A) + P(B)$$

şeklinde dir. A ve B olayları bazı hâllerde birlikte de meydana gelebilir. Bu tip olaylara birbirinden ayrık olmayan veya ara kesitli olaylar denir.



Bu durumda, bu olaylardan birinin veya diğerinin meydana gelme ihtimali,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

şeklinde hesaplanır. Yani olayların birlikte meydana gelme ihtimali, olayların ayrı ayrı meydana gelme ihtimallerinin toplamından çıkarılır.

Örnek 1: Bir müşteri $P(A) = 0.15$ ihtimalle yeşil ve $P(B) = 0.23$ ihtimalle beyaz otomobil satın alacaktır. Söz konusu müşterinin bu renk arabalardan birini satın alma ihtimali,

$$P(A \cup B) = P(A) + P(B) = 0.15 + 0.23 = 0.38$$

olarak tespit edilir.



Birbirinden ayrık olmayan olaylardan birinin veya diğerinin meydana gelmesi ihtimali, olayların ayrı ayrı meydana gelmesi ihtimallerinin toplamından olayların bir arada meydana gelmesi ihtimalinin çıkarılması ile elde edilir.

Örnek 2: Bir öğrencinin matematik dersinden geçme ihtimali, $P(A) = 2/3$, muhasebe dersinden geçmesi ihtimali $P(B) = 4/9$ ve her iki dersten geçme ihtimali $P(A \cap B) = 1/4$ ise söz konusu öğrencinin bahsedilen derslerin birisinden geçme ihtimali,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2/3 + 4/9 - 1/4 = 0.8611$$

olarak hesaplanır.

ÇARPMA KAİDESİ



Birbirinden bağımsız olayların birlikte meydana gelmesi ihtimali, olayların ayrı ayrı meydana gelmesi ihtimallerinin çarpımına eşittir.

Çarpma kaidesine konu olan olaylar ikiye ayrılır: Birbirinden bağımsız olaylar ve birbirinden bağımsız olmayan olaylar. Bir olayın vuku bulması bir başka olayın gerçekleşme şansına bağlı değilse, bu gibi olaylara **bağımsız olaylar** denir [5]. $X_1, X_2, \dots, X_n$ gibi n tane bağımsız olayın ihtimallerini $P_1, P_2, \dots, P_n$ ile gösterirsek, bu n olayın birlikte meydana gelme ihtimali,

$$(P_1) \cdot (P_2) \cdot \dots \cdot (P_n)$$

olur. Bu kaideye **çarpma kaidesi** denir. A ve B gibi iki bağımsız olayın birlikte vuku bulması ihtimali,

$$P(A \cap B) = P(A) \cdot P(B)$$

şeklinde ifade edilir.

Örnek 1: Bir bozuk para ve bir zar aynı anda atıldığında paranın tura ve zarın iki gelmesi ihtimali,

$$P(A \cap B) = P(A) \cdot P(B) = \left(\frac{1}{2}\right) \left(\frac{1}{6}\right) = \left(\frac{1}{12}\right) = 0.0833$$

olarak hesaplanır.



Birbirine bağımlı olayların birlikte meydana gelmesi ihtimali, bağımsız olayın meydana gelmesi ihtimali ile bağımsız olayın gerçekleşmesi şartına bağlı olarak bağımlı olayın meydana gelmesi ihtimalinin çarpımına eşittir.

Bir B olayının vuku bulması A olayının vuku bulması ihtimaline bağlı ise B olayının ihtimali **şartlı ihtimal**dir ve $P(B|A)$ şeklinde gösterilir. A ve B olaylarının birlikte gerçekleşmesi ihtimali,

$$P(A \cap B) = P(A) \cdot P(B|A)$$

şeklinde hesaplanır. A olayına bağlı olarak B'nin gerçekleşme ihtimali ise,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

formülü ile elde edilir [5].

Örnek 2: Bir mamul partisinde 3 kusurlu ve 17 kusursuz mamul bulunmaktadır. İmalattan tesadüfi olarak iki mamul satın alan bir müşteri, aldığı mamullerin ikisinin de kusurlu olması % kaç ihtimalledir?

Birinci mamulün kusurlu olması ihtimali $P(A) = 3/20$, birinci mamulün kusurlu olması şartına bağlı olarak ikinci mamulün kusurlu olması ihtimali $P(B|A) = 2/19$ olduğuna göre, iki hadisenin birlikte meydana gelmesi ihtimali,

$$P(A \cap B) = P(A) \cdot P(B \setminus A) = \left(\frac{3}{20}\right) \left(\frac{2}{19}\right) \cong 0.0158$$

şeklinde hesaplanır.

Örnek 3: Uçakların tam zamanında kalkması ihtimali $P(K) = 0.83$, gideceği yere zamanında varması ihtimali $P(i) = 0.82$ ve tam zamanında kalkma ve inmesi ihtimali $P(K \cap i) = 0.78$ olması hâlinde, tam zamanında kalkan bir uçağın tam zamanında inmesi ihtimali,

$$P(i \setminus K) = \frac{P(K \cap i)}{P(K)} = \frac{0.78}{0.83} \cong 0.94$$

ve tam zamanında inen bir uçağın tam zamanında kalkması ihtimali ise,

$$P(K \setminus i) = \frac{P(K \cap i)}{P(i)} = \frac{0.78}{0.82} \cong 0.95$$

şeklinde hesaplanır.

İHTİMAL DAĞILIM TABLOSU

Bir X olayının meydana gelmesinde mümkün olan hâller; $X_1, X_2, \dots, X_n$ ve bu hâllerin meydana gelme ihtimalleri de sırayla; $P_1, P_2, \dots, P_n$ ise söz konusu olaya ait ihtimal dağılım tablosu aşağıdaki gibi olur [2].

X_i	X_1	X_2		X_n	Toplam
$P(X_i)$	P_1	P_2		P_n	1

Örnek 1: Bir fabrikanın imal ettiği civataların %20'si kusurludur. Fabrikanın imalatından şansa bağlı olarak 3 civata seçildiğinde kusurlu civata sayısına ait ihtimal dağılım tablosunu düzenleyelim. L kusurlu mamul ve S, kusursuz mamul olmak üzere seçilecek bir mamulün kusurlu olması ihtimali, $P(L) = 0.20$ ve kusursuz olması ihtimali, $P(S) = 1 - 0.20 = 0.80$ 'dir. Bununla birlikte üçer birimlik mamuller seçildiği için kusurlu mamul dağılımı belirlenirken; $P(X = 0)$, üç mamulden hiçbirisinin kusurlu olmaması ihtimalini; $P(X = 1)$, üç mamulden birisinin kusurlu olması ihtimalini; $P(X = 2)$, üç mamulden ikisinin kusurlu olması ihtimalini; $P(X = 3)$, üç mamulün üçünün de kusurlu olması ihtimalini gösterir.

$$P(X = 0) = P(SSS) = (0.8)(0.8)(0.8) = 0.512$$

$$P(X = 1) = P(LSS) + P(SLS) + P(SSL) = 3[(0.2)(0.8)(0.8)] = 0.384$$

$$P(X = 2) = P(LLS) + P(LSL) + P(SLL) = 3[(0.2)(0.2)(0.8)] = 0.096$$

$$P(X = 3) = P(LLS) = (0.2)(0.2)(0.2) = 0.008$$

Üç birimlik mamul örneğinde kusurlu sayılarına ilişkin ihtimal dağılım tablosu aşağıdaki gibidir.

X_i	0	1	2	3	Toplam
$P(X_i)$	0.512	0.384	0.096	0.008	1



Bileşik ihtimal hesaplamalarında “veya” ve “ya da” bağlacının karşılığı toplama kaidesi iken, “ve” ve “hem...hem...” bağlacının karşılığı çarpma kaidesidir.



Bir olayın mümkün sonuçlarına ait ihtimallerin gösterildiği tabloya ihtimal dağılım tablosu denir.

Örnek 2: Dört çocuklu bir ailede erkek çocuklara ait ihtimal dağılımı tablosunu düzenleyelim.

Hiç erkek çocuk olmaması ihtimali,
 $P(KKKK)$

1 erkek çocuk olması ihtimali,
 $P(EKKK+KEKK+KKEK+KKKE)$

2 erkek çocuk olması ihtimali,
 $P(EEEK+EEKK+KEEK+KEKE+KEEK)$

3 erkek çocuk olması ihtimali,
 $P(EEEE+EEKE+EKEE+EEEE)$

4 erkek çocuk olması ihtimali,
 $P(EEEE)$

Buna göre ihtimal dağılım tablosu,

X_i	0	1	2	3	4	Toplam
$P(X_i)$	1/16	4/16	6/16	4/16	1/16	16/16

şeklinde elde edilir.

BEKLENEN DEĞER (MATEMATİK ÜMİT)

n adet denemede X_1 olayı P_1 ihtimalle, X_2 olayı P_2 ihtimalle, X_n olayı P_n ihtimalle meydana geliyorsa, X_1 'in beklenen değeri $E(X_1) = nP_1$, X_2 'nin beklenen değeri $E(X_2) = nP_2$, X_n 'in beklenen değeri $E(X_n) = nP_n$ 'dir [4].

Örnek 1: İhtimal dağılım tablosu açıklanırken verilen bir numaralı örneğe geri dönelim. Bir firma söz konusu fabrikadan üçer birimlik 1000 paket mal satın aldığı anda,

$$E(X = 0) = nP_0 = 1000(0.512) = 512$$

pakette hiç kusurlu mamul gözlememeyi beklerken,

$$E(X = 1) = nP_1 = 1000(0.384) = 384$$

pakette bir kusurlu mamul,

$$E(X = 2) = nP_2 = 1000(0.096) = 96$$

pakette de iki kusurlu mamul bekler. Nihayet,

$$E(X = 3) = nP_3 = 1000(0.008) = 8$$

pakette ise mamullerin tamamının kusurlu olmasını bekler.

Örnek 2: Bir çantada 2 tane beyaz ve 3 tane siyah top vardır. Dört şahıs (A, B, C, D) isim sırasına göre bu çantadan birer top çekmekte ve çektikleri topları çantaya geri bırakmamaktadır. Beyaz topu ilk defa çeken şahsa 100000 TL verilmesi halinde A, B, C ve D şahıslarının matematik ümidini bulunuz.



Bir olayın beklenen değeri, o olayın gerçekleşmesi ihtimali ile toplam deneme sayısının çarpımına eşittir.

Beyaz topu A'nın çekmesi ihtimali,

$$P(A) = \frac{2}{5}$$

ve A'nın matematik ümidi, $E(A) = \left(\frac{2}{5}\right)100000 = 40000$ -TL'dir. A'nın çekmeyip

B'nin çekmesi ihtimali,

$$P(\overline{A}B) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right) = \frac{3}{10}$$

ve B'nin matematik ümidi, $E(B) = (0.3)100000 = 30000$ 'dir. A ve B'nin çekemeyip C'nin çekmesi ihtimali,

$$P(\overline{A}\overline{B}C) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = \frac{1}{5}$$

ve C'nin matematik ümidi, $E(C) = (0.2)100000 = 20000$ -TL'dir. A, B ve C şahıslarının beyaz topu çekemeyip D'nin çekmesi ihtimali,

$$P(\overline{A}\overline{B}\overline{C}D) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right)1 = \frac{1}{10}$$

bulunur. D'nin matematik ümidi, $E(D) = (0.1)100000 = 10000$ TL'dir.

ÖNEMLİ BAZI İHTİMAL KURALLARI



Aynı anda vuku bulmaları imkânsız olan birbirinden farklı k adet olay n defa tekrarlanırsa, mümkün sonuç sayısı k^n olur.

1. Kural: Aynı anda vuku bulmaları imkânsız olan birbirinden farklı k adet olay n defa tekrarlanırsa, mümkün sonuç sayısı,

$$k^n$$

olur [5]. Bir para 10 kez atıldığında mümkün sonuç sayısı, $2^{10} = 1024$ 'dür.

2. Kural: İlk denemede k_1 , ikinci denemede k_2 ve n inci denemede k_n adet olayla karşılaşıyorsak mümkün sonuç sayısı,

$$(k_1)(k_2).....(k_n)$$

şeklinde hesaplanır [2]. Mesela bir lokantada akşam yemeği için servis yapılacak 4 çeşit salata, 10 çeşit çorba, 3 çeşit meşrubat ve 6 çeşit tatlı varsa mümkün akşam yemeği sayısı,

$$(4)(10)(3)(6) = 720$$

olarak hesaplanır.

3. Kural: Sıra önemli olduğunda n adet olay,

$$n! = n(n-1).....(1)$$

yolla vuku bulabilir [5]. $n!$, n faktöriyel diye okunur. $0!$, 1'e eşittir. Mesela 6 kitap,

$$6! = (6)(5)(4)(3)(2)(1) = 720$$

farklı dizilişle rafa yerleştirilebilir.



İlk denemede k_1 , ikinci denemede k_2 ve n inci denemede k_n adet olayla karşılaşıyorsak mümkün sonuç sayısı k değerlerinin çarpımına eşittir.



Sıra önemli olduğundan
adet olay $n!$ yolla
meydana gelir.



Sıra önemli olduğunda
 n adet olaydan X adedi,
 n 'in X 'li permütasyonu
kadar farklı yolla
meydana gelir.



Sıra önemli
olmadığında n adet
olaydan X adedi, n 'in
 X 'li kombinasyonu
kadar farklı yolla
meydana gelir.

4. Kural: Sıra önemli olduğunda n adet olaydan X adedi,

$$\frac{n!}{(n - X)!}$$

farklı yolla birlikte vuku bulabilirler [5]. Buna *permütasyon* kuralı denir. Mesela elimizde 6 ders kitabı olmakla birlikte kitaplıkta bu kitaplardan 4 tanesini koyabilecek kadar yer varsa, sıra önemli olduğunda, kitaplar kitaplığı,

$$\frac{6!}{(6 - 4)!} = \frac{6!}{2!} = \frac{(6)(5)(4)(3)(2)(1)}{(2)(1)} = 360$$

farklı şekilde dizilebilir.

5. Kural: Sıra önemli olmadığında n adet olaydan X adedi,

$$\binom{n}{X} = \frac{n!}{X!(n - X)!}$$

kadar farklı yolla birlikte vuku bulabilirler [5]. Bu kurala *kombinasyon* kuralı denir. Sıra önemli olmadığında 6 ders kitabından 4'ü,

$$\frac{6!}{4!(6 - 4)!} = \frac{6!}{4!2!} = \frac{(6)(5)(4)(3)(2)(1)}{(4)(3)(2)(1)(2)(1)} = 15$$

farklı şekilde kitaplığa yerleştirilebilir.



Bireysel Etkinlik

- Bir olayın nispi frekansı ile o olayın gelecekte meydana gelme ihtimali arasında bir ilişki var mıdır? Araştırınız.
- Bir olayın beklenen değerinin hesaplanmasının avantajlarını araştırınız.
- İhtimal hesaplamalarının istatistik bilimi açısından önemini araştırınız.



Özet

- İhtimal (olasılık) kavramı hayatımızın her alanında karşımıza çıkabilecek olaylar için kullanılır. Bir olayın olması mümkün olduğu gibi olmaması da mümkün ise bu olay ihtimale konu olan bir olaydır. Bir olay kesin olarak meydana gelecektse veya kesin olarak meydana gelmeyecekse ihtimalden bahsedilemez. Bir olayın olması mümkün olduğu gibi olmaması da mümkünse bu tür olaylara ihtimalli olaylar denir.
- Bir olayın meydana gelme ihtimali söz konusu olayın meydana gelme sayısının o olayla birlikte meydana gelmesi mümkün sonuç sayısına bölümüdür.
- Tek bir olayın sonuçları ile ilgili ihtimaller basit ihtimallerdir. Mesela; yarın yağmur yağması ihtimali, bir sınıftan tesadüfi olarak seçilen bir öğrencinin gözlüklü olması ihtimali gibi ihtimaller ayrı ayrı düşünüldüğünde basit birer ihtimaldir. İki veya daha fazla olayın birlikte vuku bulması ihtimali ise bileşik bir ihtimaldir. Aynı şekilde ikiden fazla olaydan bazılarının bazıları ile birlikte vuku bulması ihtimali de bileşik ihtimaldir. Bileşik ihtimal hesaplarına konu olan olaylar iki gruba ayrılır:
 - Bir arada meydana gelebilen olaylar,
 - Birbirini engelleyen olaylar.
 - Birbirini engelleyen birden fazla olaydan bir veya daha fazlasının meydana gelme ihtimali bulunurken toplama kuralından yararlanılır.
 - Birbirini engellemeyen bir veya daha fazla olayın birlikte meydana gelme ihtimali hesaplanırken çarpma kuralından faydalanılır.
- İstatistiki bir olayın mümkün olan bütün sonuçlarının oluşturduğu sete örnek uzayı denir ve S ile gösterilir.
- Bir olayın beklenen değeri o olayın gerçekleşmesi ihtimali ile toplam deneme sayısının çarpımına eşittir.
- Bir olayın mümkün gerçekleşme yollarının ve bu yolların ihtimallerinin gösterildiği tabloya ihtimal dağılım tablosu denir.
- Aynı anda vuku bulmaları imkânsız olan birbirinden farklı k adet olay n defa tekrarlanırsa, mümkün sonuç sayısı k^n olur.
- İlk denemede k_1 , ikinci denemede k_2 ve n inci denemede k_n adet olayla karşılaşıyorsak mümkün sonuç sayısı k değerlerinin çarpımına eşittir.
- Sıra önemli olduğunda n adet olay $n!$ yolla meydana gelir.
- Sıra önemli olduğunda n adet olaydan X adedi, n 'in X 'li permütasyonu kadar farklı yolla meydana gelir.
- Sıra önemli olmadığında n adet olaydan X adedi, n 'in X 'li kombinasyonu kadar farklı yolla meydana gelir.

DEĞERLENDİRME SORULARI

1. Bir ihtimal sonucu aşağıdakilerden hangisi olabilir?
 - a) -1.2551
 - b) -0.2551
 - c) 0.2551
 - d) 1.2551
 - e) $\frac{3}{2}$
2. Kuvvetli ihtimali aşağıdaki sonuçlardan hangisi gösterir?
 - a) 0.05
 - b) 0.45
 - c) 0.55
 - d) 0.65
 - e) 0.95
3. Bir fabrikada üretilen 100 mamulden 10'u kusurlu, 20 tanesi bakıma ihtiyaç duyan mal ve 70 tanesi sağlamdır. Bu fabrikadan tesadüfi olarak bir mal seçilirse bu malın sağlam olma ihtimali kaçtır?
 - a) 0.10
 - b) 0.20
 - c) 0.50
 - d) 0.70
 - e) 0.90
4. Bir fabrikada üretilen 100 mamulden 10'u kusurlu, 20 tanesi bakıma ihtiyaç duyan mal ve 70 tanesi sağlamdır. Bu fabrikadan tesadüfi olarak bir mal seçilirse bu malın kusurlu olmaması ihtimali kaçtır?
 - a) 0.10
 - b) 0.20
 - c) 0.50
 - d) 0.70
 - e) 0.90
5. Bir fabrikada üretilen 100 mamulden 10'u kusurlu, 20 tanesi bakıma ihtiyaç duyan mal ve 70 tanesi sağlamdır. Bu fabrikadan tesadüfi olarak 1000 mal seçilirse bu mallardan kaç tanesinin sağlam olması beklenir?
 - a) 100
 - b) 200
 - c) 500
 - d) 600
 - e) 700

6. Ahmet elindeki 10 TL ile %10 ihtimalle simit, %30 ihtimalle poğaç ve %60 ihtimalle tost alacaktır. Ahmet'in elindeki para ile simit veya poğaç alma ihtimali kaçtır?
- 0.30
 - 0.40
 - 0.60
 - 0.80
 - 0.90
7. Ahmet'in İstatistik dersinden başarılı olma ihtimali 0.65, matematik dersinden başarılı olma ihtimali 0.70 ve her iki dersten başarılı olma ihtimali 0.50'dir. Ahmet'in İstatistik veya matematik dersinden başarılı olma ihtimali kaçtır?
- 0.40
 - 0.50
 - 0.65
 - 0.75
 - 0.85

(8., 9. ve 10. soruları vakaya göre cevaplayınız.)

Vaka: Sigara içme ile akciğer kanserine yakalanma arasındaki ilişkiyi açıklamaya çalışan bir araştırmacı elde ettiği bulguları aşağıdaki kontenjans tablosunda özetlemiştir:

Sigara İçme	Akciğer Kanserine Yakalanma		Toplam
	Yakalananlar	Yakalanmayanlar	
İçenler	400	200	600
İçmeyenler	100	300	400
Toplam	500	500	1000

8. Araştırmaya göre tesadüfi olarak seçilen bir kişinin sigara içiyor olma ihtimali kaçtır?
- 0.40
 - 0.50
 - 0.60
 - 0.70
 - 0.80
9. Araştırmaya göre tesadüfi olarak seçilen bir kişinin hem sigara içiyor olma hem de akciğer kanserine yakalanmış olma ihtimali kaçtır?
- 0.20
 - 0.30
 - 0.40
 - 0.50
 - 0.60

10. Araştırmaya göre tesadüfi olarak seçilen bir kişinin sigara içiyor olma veya akciğer kanserine yakalanmış olma ihtimali kaçtır?
- a) 0.40
 - b) 0.60
 - c) 0.70
 - d) 0.80
 - e) 0.90

Cevap Anahtarı

1.c, 2.e, 3.d, 4.e, 5.e, 6.b, 7.e, 8.c, 9.c, 10.c

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İsmail Hakkı (2018). İşletmelerde Uygulamalı İstatistik (3. baskı). İstanbul: Alfa Yayınları.
- [2] Başar, Alaaddin, Erkan Oktay (2018). Uygulamalı İstatistik – II: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Erzurum Kültür ve Eğitim Kitap ve Kirtasiye.
- [3] Daniel, Wayne W., Terrell, James C. (1995). Business Statistics: For Management and Economics (7. baskı). Boston: Houghton Mifflin Company.
- [4] Gürtan, Kenan (1982). İstatistik ve Araştırma Metodları. İstanbul: İstanbul Üniversitesi Yayınları.
- [5] Serper, Özer (2014). Uygulamalı İstatistik I (7. baskı). İstanbul: Ezgi Kitabevi.

KESİKLİ TESADÜFİ DEĞİŞKENLER VE DAĞILIMLARI



İÇİNDEKİLER

- Kesikli Tesadüfi Değişkenler
- Kesikli Tesadüfi Değişkenlerin Dağılımı
 - Binom Dağılımı
 - Poisson Dağılımı



HEDEFLER

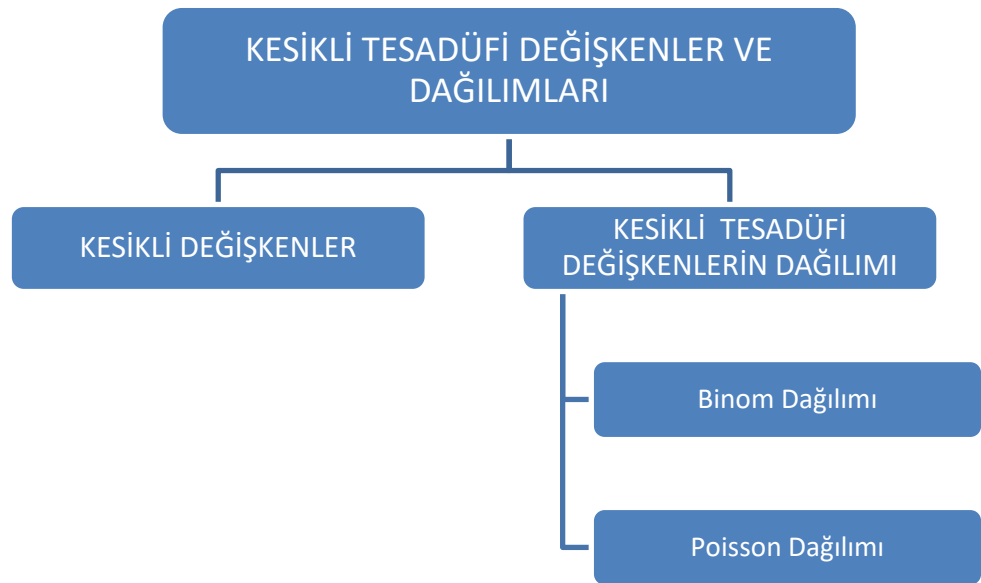
- Bu üniteyi çalıştıktan sonra;
 - Kesikli tesadüfi değişken kavramını anlayabilecek,
 - Kesikli tesadüfi değişkenlere ait dağılımların ortalama ve standart sapmasını hesaplanabilecek,
 - Binom dağılımı ve binom dağılımının karakteristiklerini kavrayacak,
 - Binom dağılımı gösteren bir olayın sonuçlarını hesaplayabilecek,
 - Poisson dağılımı ve poisson dağılımının karakteristiklerini kavrayabilecek,
 - Poisson dağılımı gösteren bir olayın sonuçlarını hesaplayabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Prof. Dr. Erkan
OKTAY

ÜNİTE
12



GİRİŞ

Bir önceki ünite de ihtimaller teorisinden bahsedilmişti. Bu ünite de ise kesikli tesadüfi değişkenler ve bu değişkenlere ait ihtimal dağılımlarından bahsedilecektir.

Tekrarlanan olaylara ilişkin sonuçların kesikli değişken değerleri olmasından dolayı, bu tür olayların sonuçlarının gerçekleşme ihtimallerine ait dağılımlara kesikli ihtimal dağılımları denir [1].

Kesikli değişken belirli değerler arasında sadece tam sayı değerler alabilen değişkendir. Mesela 0 ila 5 değerleri arasında 0 ve 5 dâhil edilirse sadece 0, 1, 2, 3, 4 ve 5 belirli tam sayı değerleri vardır. Kesikli değişken değerleri genellikle sayımla elde edilirler. Mesela bir derslikteki öğrenci sayısı, bir otobüsteki yolcu sayısı, bir ailedeki çocuk sayısı, bir para beş kez atıldığında yazı gelme sayısı ve bir sinema salonundaki izleyici sayısı kesikli değişken değerlerine verilebilecek örneklerdendir [2].

KESİKLİ TESADÜFİ DEĞİŞKENLER

Kesikli değişken, sonuçları sayımla elde edilen değişkendir. Bu sebeple kesikli değişkenlere ait sonuçlar yalnızca belirli tam sayı değerleri alabilir. Mesela bir köyde yaşayan ailelerdeki çocuk sayıları Tablo 12.1’de gösterilmiştir.

Tablo 12.1. Bir Köydeki Ailelerin Çocuk Sayılarına Göre Dağılımı

Çocuk Sayısı (X_i)	Aile Sayısı	Nispi Frekans
0	50	0.05
1	130	0.13
2	220	0.22
3	410	0.41
4	150	0.15
5	40	0.04
Toplam	1000	1.00

Tablo 12.1’de gösterilen değişken ailelerdeki çocuk sayısıdır. Bu değişken sayımla elde edilen değerleri gösterir. Bir ailedeki çocuk sayısı 0 olabilir, 1 olabilir; fakat 1.5 olamaz. Çocuk sayısı belirli tam sayı değerlerle ifade edilebilir. Bu örneğimizdeki değişkenimizin alabileceği muhtemel sonuçlar 0, 1, 2, 3, 4, ve 5’tir. Değişkenin sonuçlarından birinin gözlenmesi şansa bağlı olarak ortaya çıkar. Bu sebeple bu değişkene tesadüfi değişken denir. Her bir mutlak frekans toplam frekansa bölünerek nispi frekanslar elde edilmiştir. Nispi frekans terimi mevcut durum ve geçmiş için kullanılır. Olayın gelecekte meydana gelme durumundan bahsedildiğinde nispi frekanslara ihtimal değerleri nazarıyla bakılabilir.

Tablo 12.1’deki tesadüfi değişkenin sonuçları kesikli olduğu için bu değişkene, kesikli tesadüfi değişken denir. Mesela bahse konu olan köydeki ailelerden biri tesadüfi olarak seçilirse ailenin iki çocuklu bir aile olması ihtimali 0.22’dir.



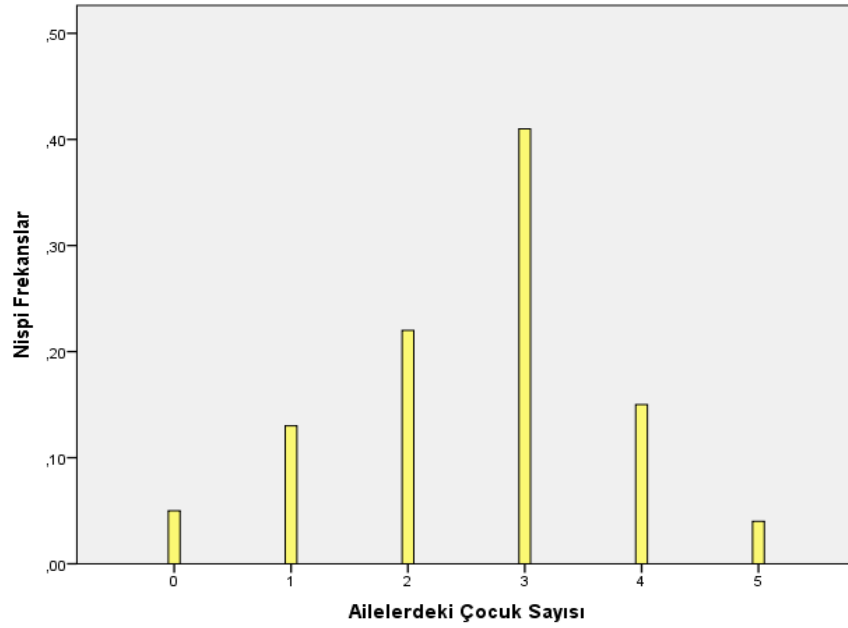
Kesikli değişken, belirli değerler arasında sadece tam sayı değerler alabilen değişkendir.



Kesikli değişken, sonuçları sayımla elde edilen değişkendir.



Beş çocuklu ailelerdeki erkek çocuk sayısı dağılımı kesikli dağılımdır. Sonuçları; 0, 1, 2, 3, 4 ve 5’tir.



Görsel 12.1. Bir Köydeki Ailelerin Çocuk Sayılarına Göre Dağılımı

Tablo 12.1'deki değişkene ait sonuçlar X-Y koordinat sistemine taşınırsa Görsel 12.1'deki grafik elde edilir. Tablo 12.1'de de görüleceği gibi ailelerdeki çocuk sayısı tesadüfi değişkeninin sadece 0, 1, 2, 3, 4 ve 5 değerleri vardır ve bu değerler için ihtimallerden bahsedilebilir.



X kesikli değişkenine ait mümkün bütün sonuçları ve bu sonuçların meydana gelme ihtimallerini gösteren dağılımlara kesikli ihtimal dağılımları denir.

KESİKLİ İHTİMAL DAĞILIMLARI

X değişkeni tesadüfi bir değişken olmak üzere, X değişkenine ait mümkün bütün sonuçları ve bu sonuçların meydana gelme ihtimallerini gösteren dağılımlara kesikli ihtimal dağılımları denir [1]. Mesela yukarıda bahsi geçen köyden tesadüfi olarak bir aile seçildiğinde ailedeki çocuk sayısına ait ihtimal dağılım tablosu aşağıdaki gibidir. Unutmayınız olayın gelecekte tekrar edebilmesi durumunda Tablo 12.2'deki ihtimaller Tablo 1'deki nispi frekanslardan elde edilmiştir.

Tablo 12.2. Bir Köydeki Ailelerin Çocuk Sayılarının İhtimal Dağılımı

Çocuk Sayısı (X_i)	İhtimal, $P(X_i)$
0	0.05
1	0.13
2	0.22
3	0.41
4	0.15
5	0.04
Toplam	1.00

Tablo 12.2'e göre bahse konu olan köyden tesadüfi olarak bir aile seçilirse bu ailede 0 çocuk olması ihtimali 0.05, 1 çocuk olması ihtimali 0.13, ..., 5 çocuk olması ihtimali 0.04'tür.

Tablo 12.2’de de görüleceği üzere ihtimal dağılım tablosundaki ihtimaller 0 ile 1 arasındadır. Olayın tüm sonuçları ilgili tabloda gösterilmiştir ve bu sonuçların gerçekleşmesi ihtimallerinin toplamı 1’dir.

Tablo 12.2’deki verilere göre bahse konu olan köyden tesadüfi olarak bir aile seçilirse, ailede en fazla bir çocuk olması ihtimali $P(X \leq 1) = P(X=0) + P(X=1) = 0.05 + 0.13 = 0.18$ ’dir.

Tablo 12.2’deki verilere göre bahse konu olan köyden tesadüfi olarak bir aile seçilirse, ailede en az üç çocuk olması ihtimali $P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) = 0.41 + 0.15 + 0.04 = 0.6$ ’dır.

Tablo 12.2’deki verilere göre bahse konu olan köyden tesadüfi olarak bir aile seçilirse ailede üçten fazla çocuk olması ihtimali $P(X > 3) = P(X=4) + P(X=5) = 0.15 + 0.04 = 0.19$ ’dur.

Kesikli değişkene ait ihtimal dağılımının ortalaması μ ile gösterilir. Ortalamaya beklenen değer de denir ve $E(X)$ ile gösterilir. Kesikli tesadüfi değişkenin ortalaması, kesikli değişken değerleri ile bu değişken değerlerinin ihtimallerinin çarpılıp toplanması sonucu elde edilir. Kesikli tesadüfi değişkenin ortalaması,

$$\mu = E(X_i) = \sum_{i=0}^n X_i P(X_i)$$

formülü yardımıyla elde edilir [4]. Tablo 12.2’deki kesikli değişkene ait sonuçları ve sonuçların gerçekleşmesi ihtimallerini kullanarak dağılıma ait ortalamayı elde edebiliriz.

Tablo 12.3. Ailelerdeki Çocuk Sayıları Dağılımının Ortalaması

Çocuk Sayısı (X_i)	$P(X_i)$	$X_i P(X_i)$
0	0.05	0(0.05)=0.00
1	0.13	1(0.13)=0.13
2	0.22	2(0.22)=0.44
3	0.41	3(0.41)=1.23
4	0.15	4(0.15)=0.60
5	0.04	5(0.04)=0.20
Toplam	1.00	$\sum_{i=0}^n X_i P(X_i)=2.60$

$$\mu = \sum_{i=0}^n X_i P(X_i) = 2.6$$

Tablo 12.3’e göre bahsi geçen köydeki ailelerde ortalama 2.6 çocuk bulunmaktadır. Bir başka ifadeyle tesadüfi olarak bir aile seçildiğinde bu ailede ortalama 2.6 çocuk bulunmasını bekleriz. Üzerinde durulan değişken kesikli



Kesikli tesadüfi değişkenin ortalaması, kesikli değişken değerleri ile bu değişken değerlerinin ihtimallerinin çarpılıp toplanması sonucu elde edilir.

değişken olmasına rağmen ortalama veya beklenen değer 2.6 gibi kesikli olmayan bir değer alabilir.

Kesikli tesadüfi değişkenin standart sapması, kesikli ihtimal dağılımının değişkenliğinin bir ölçüsüdür ve σ sembolü ile gösterilir. Standart sapmanın büyük olması kesikli değişken değerlerinin daha geniş bir aralıkta dağılım gösterdiğini ifade eder.

Kesikli tesadüfi değişkenin standart sapması,

$$\sigma = \sqrt{\sum_{i=0}^n X_i^2 P(X_i) - \mu^2}$$

veya

$$\sigma = \sqrt{\sum_{i=0}^n (X_i - \mu)^2 P(X_i)}$$

formüllerinden biri ile hesaplanabilir [5]. Tablo 12.3'deki kesikli değişkene ait sonuçları ve sonuçların gerçekleşmesi ihtimallerini kullanarak dağılıma ait standart sapmayı elde edebiliriz.

Tablo 12.4. Ailelerdeki Çocuk Sayıları Dağılımının Ortalaması

(X_i)	$P(X_i)$	X_i^2	$X_i^2 P(X_i)$
0	0.05	0	0(0.05)=0.00
1	0.13	1	1(0.13)=0.13
2	0.22	4	4(0.22)=0.88
3	0.41	9	9(0.41)=3.69
4	0.15	16	16(0.15)=2.40
5	0.04	25	25(0.04)=1.00
Toplam	1.00		$\sum_{i=0}^n X_i^2 P(X_i)=8.10$

Tablo 12.4'teki $\sum_{i=0}^n X_i^2 P(X_i)=8.10$ değerini ve daha önce elde edilen $\mu = 2.6$ değerini kullanarak,

$$\sigma = \sqrt{\sum_{i=0}^n X_i^2 P(X_i) - \mu^2} = \sqrt{8.1 - 2.6^2} = \sqrt{8.1 - 6.76} = 1.16$$

olarak bulunur. Bahsi geçen köydeki ailelerin çocuk sayılarına göre dağılımı incelenirken ailelerdeki ortalama çocuk sayısının 2.6 olduğunu ve dağılımın standart sapmasının 1.16 olduğunu unutmayınız.



Standart sapmanın büyük olması, kesikli değişken değerlerinin daha geniş bir aralıkta dağılım gösterdiğini ifade eder.



Bir X olayının meydana gelmesinde sadece iki hâl söz konusu ise bu olayın binom dağılımı gösterdiği söylenir.

Binom dağılımı

Bir X olayının meydana gelmesinde sadece iki hâl söz konusu ise bu olayın binom dağılımı gösterdiği söylenir. Uygulamada başarılı veya başarısız, kusurlu veya kusursuz, yazı veya tura, erkek veya kız gibi iki kategorili olaylar binom dağılımı gösterirler. Binom dağılımında başarılı olma ihtimali p ; başarısızlık ihtimali ise $1-p$ ile gösterilir. Ancak şunu belirtelim ki gerçekte başarı, üzerinde durulan olayın meydana gelmesi; başarısızlık ise meydana gelmemesidir. Mesela bir yazı-tura oyununda, yazı gelme olayı üzerinde duruyorsak p yazı gelme ihtimalini, $1-p$ yazı gelmeme yani tura gelme ihtimalini gösterir. Tura gelme olayı ile ilgilendiğimiz takdirde p tura gelme ihtimalini ifade eder [4].

Bir olayın binom olayı olabilmesi için p ihtimalinin olaydan olaya değişmemesi gerekir. Bu da ancak iki şekilde olabilir. Sonlu anakütlelerden iadeli çekilişlerin yapılması veya sonsuz anakütlelerden çekilişlerin yapılması gerekir.

Üzerinde durulan olayın n denemede x defa meydana gelmesi ihtimali,

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n$$

eşitliğiyle hesaplanır. Formülde,

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

şeklinde açılır. x değerleri 0, 1, 2, ..., n gibi kesikli değerler alabileceğinden ve sadece bu değerler için nokta ihtimalleri hesaplanabileceğinden binom dağılımı kesikli bir ihtimal dağılımıdır [6].

Binom dağılımının karakteristikleri

Binom dağılımının parametresi p 'dir. Bu dağılımın ortalaması, varyansı, standart sapması, moment çarpıklık ve basıklık ölçüleri aşağıdaki eşitliklerle hesaplanır:

$$\text{Ortalama } (\mu) = np$$

$$\text{Varyans } (\sigma^2) = np(1-p)$$

$$\text{Standart sapma } (\sigma) = \sqrt{np(1-p)}$$

$$\text{Moment çarpıklık ölçüsü } (\alpha_3) = \frac{1-2p}{\sqrt{np(1-p)}}$$

$$\text{Moment basıklık ölçüsü } (\alpha_4) = 3 + \frac{1-6p(1-p)}{np(1-p)}$$

$p = 0.5$ olduğunda binom dağılımı simetriktir. $p > 0.5$ olduğunda dağılım sola çarpık, $p < 0.5$ olduğunda ise sağa çarpıktır [3].



Binom dağılımının parametresi p 'dir. $p = 0.5$ olduğunda binom dağılımı simetriktir. $p > 0.5$ olduğunda dağılım sola çarpık, $p < 0.5$ olduğunda ise sağa çarpıktır.

Örnek 1: Bir proste üretilen mamullerin %15'inin kusurlu olduğu biliniyor. Bu prosten şansa bağlı olarak alınan 3 birimlik bir mamul örneğinde en az iki mamulün kusurlu çıkma ihtimali,

$$\begin{aligned}
 P(X \geq 2) &= P(X = 2) + P(X = 3) \\
 P(X \geq 2) &= \binom{3}{2} 0.15^2 (1 - 0.15)^{3-2} + \binom{3}{3} 0.15^3 (1 - 0.15)^{3-3} \\
 P(X \geq 2) &= \frac{3!}{2!(3-2)!} 0.15^2 0.85 + \frac{3!}{3!(3-3)!} 0.15^3 0.85^0 \\
 P(X \geq 2) &= 0.0574 + 0.0034 = 0.0608 \cong \%6.08
 \end{aligned}$$

olarak hesaplanır.

Örnek 2: Standart dışı mamul oranının %10 olduğu bilinen bir prosten alınan 4 birimlik mamul örneğinde en çok iki mamulün standart dışı olma ihtimali,

$$\begin{aligned}
 P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
 P(X \leq 2) &= \binom{4}{0} 0.10^0 (1 - 0.10)^{4-0} + \binom{4}{1} 0.10^1 (1 - 0.10)^{4-1} + \binom{4}{2} 0.10^2 (1 - 0.10)^{4-2} \\
 P(X \leq 2) &= \frac{4!}{0!(4-0)!} 0.10^0 0.90^4 + \frac{4!}{1!(4-1)!} 0.10^1 0.90^3 + \frac{4!}{2!(4-2)!} 0.10^2 0.90^2 \\
 P(X \leq 2) &= 0.6561 + 0.2916 + 0.0486 = 0.9963 = \%99.63
 \end{aligned}$$

olur.

Örnek 3: Erkek çocuk doğma ihtimalini %50 kabul ederek, dört çocuklu bir ailede en az bir erkek çocuk olması ihtimalini bulunuz.

Çözüm: $n = 4$, $p = 0.5$ olduğu için,

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1 - P(X = 0)$$

$$P(X = 0) = \binom{4}{0} 0.5^0 (1 - 0.5)^{4-0} = 0.0625 \text{ 'tir. Böylece,}$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.0625 = 0.9375 \text{ olarak elde edilir.}$$

Örnek 4: Bir makinenin imal ettiği civataların %20'si kusurlu çıkıyorsa, şansa bağlı olarak seçilen 4 civatadan;

- a) Birisinin kusurlu çıkma ihtimalini,
- b) Hiçbirinin kusurlu çıkmama ihtimalini,
- c) En fazla ikisinin kusurlu çıkma ihtimalini bulunuz.

Çözüm: $p = 0.20$ ve $n = 4$

$$\text{a) } P(X = 1) = \binom{4}{1} (0.2)(1 - 0.2)^{4-1} = 0.4096$$

$$\text{b) } P(X = 0) = \binom{4}{0} (0.2)^0 (1 - 0.2)^4 = 0.4096$$

c) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$ 'dir.

(a) şıkkında $P(X = 1) = 0.4096$ ve (b) şıkkında $P(X = 0) = 0.4096$ bulunmuştur.

$$P(X = 2) = \binom{4}{2} (0.2)^2 (1 - 0.2)^{4-2} = 0.1536$$

olduğuna göre

$$P(X \leq 2) = 0.4096 + 0.4096 + 0.1536 = 0.9728$$

olarak elde edilir.

Örnek 5: Bir sigortacı hepsi aynı yaşta ve sıhhatli 5 kişiyle sigorta mukavelesi yapmıştır. *Yaşama müddeti ihtimal tablosuna* göre bu şahısların yaşında olan birisinin 2/3 ihtimalle 30 yıl yaşayacağı tahmin edilmektedir. Bu 5 kişinin;

- a) Hepsinin,
- b) En az üçünün,
- c) Sadece ikisinin ve
- d) En az birisinin 30 yıl yaşamaları ihtimalini bulunuz.

Çözüm:

$$a) P(X = 5) = \binom{5}{5} \left(\frac{2}{3}\right)^5 \left(1 - \frac{2}{3}\right)^{5-5} = 0.1317$$

$$b) P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$

$$P(X = 3) = \binom{5}{3} \left(\frac{2}{3}\right)^3 \left(1 - \frac{2}{3}\right)^{5-3} \cong 0.3292$$

$$P(X = 4) = \binom{5}{4} \left(\frac{2}{3}\right)^4 \left(1 - \frac{2}{3}\right)^{5-4} \cong 0.3292$$

$$P(X \geq 3) = 0.3292 + 0.3292 + 0.1317 \cong 0.7901$$

$$c) P(X = 2) = \binom{5}{2} \left(\frac{2}{3}\right)^2 \left(1 - \frac{2}{3}\right)^{5-2} = 0.1646$$

$$d) P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{5}{0} \left(\frac{2}{3}\right)^0 \left(1 - \frac{2}{3}\right)^{5-0} = 1 - 0.0041 = 0.9959$$

Örnek 6: $p = 0.7$ ve $n = 60$ olan bir binom dağılımı için,

- a) Ortalamayı,
- b) Standart sapmayı,
- c) Moment çarpıklık katsayısını,
- d) Moment basıklık katsayısını hesaplayınız.

Çözüm: $p = 0.7, n = 60$

a) $\bar{X} = np = 60(0.7) = 42$

b) $\sigma = \sqrt{np(1-p)} = \sqrt{60(0.7)(0.3)} \cong 3.55$

c) $\alpha_3 = \frac{1-2p}{\sigma} = \frac{1-2(0.7)}{3.55} \cong -0.11$ (hafifçe sola çarpık)

d) $\alpha_4 = 3 + \frac{1-6p(1-p)}{\sigma^2} = 3 + \frac{1-6(0.7)(1-0.7)}{(3.55)^2} \cong 2.98$ (hafifçe basık)

Poisson Dağılımı

Poisson dağılımı da binom dağılımı gibi kesikli bir ihtimal dağılımıdır. Poisson olayları binom olaylarına çok benzer. Binom dağılımından farklı olarak bu dağılımda üzerinde durulan olayın meydana gelme ihtimali çok küçüktür. n 'in büyümesi p 'nin de küçülmesi hâlinde, binom formülü yerine poisson formülü kullanılır. Daha net bir ifadeyle, np değeri 5'ten küçük olduğunda binom dağılımı poisson dağılımına dönüşür. Belirli bir poisson olayının ihtimali,

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots, n$$

formülü ile hesaplanır [2]. Burada λ , poisson dağılımının ortalamasıdır. Bu dağılımın varyansı da λ 'ya eşittir. Poisson dağılımının ortalaması $\lambda = np$ formülü yardımıyla hesaplanır. Poisson formülündeki e sabiti, yaklaşık olarak 2.71828 değerine eşittir [6].

Poisson dağılımının karakteristikleri

Poisson dağılımının parametresi λ 'dır.

Ortalama (μ) = λ

Varyans (σ^2) = λ

Standart sapma (σ) = $\sqrt{\lambda}$

Moment çarpıklık ölçüsü (α_3) = $\frac{1}{\lambda}$

Moment basıklık ölçüsü (α_4) = $3 + \frac{1}{\lambda}$

Poisson dağılımı normalden daha diktir ve sağa çarpıktır. λ büyüdükçe dağılım simetrikleşir [1].

Örnek 1: Bir otomobil galerisine ayda ortalama 150 müşteri gelmektedir. Herhangi bir günde dükkanını açmayan galeri sahibi % kaç ihtimalle en az üç müşteriyi kaçırmıştır? (1 ayda 30 gün olduğunu farz ediniz.)



Binom dağılımından farklı olarak bu dağılımda üzerinde durulan olayın meydana gelme ihtimali çok küçüktür.



Poisson dağılımının parametresi λ 'dır. λ , poisson dağılımının ortalamasıdır. Bu dağılımın varyansı da λ 'ya eşittir.

Aylık λ değeri 150 olduğuna göre günlük λ değeri $150/30 = 5$ 'tir. 3 veya daha fazla müşteriyi kaçırma ihtimali hesaplanırken ya 3'ten sonsuza kadar olan ihtimalleri hesaplayıp toplayacağız veya X 'in 0, 1 ve 2 olması ihtimallerini hesaplayıp topladıktan sonra 1'den çıkaracağız. İkinci yol hem daha kısa ve güvenilir olduğu için galericinin üç veya daha fazla müşteriyi kaçırma ihtimali,

$$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

$$P(X \geq 3) = 1 - \left[\frac{e^{-5} 5^0}{0!} + \frac{e^{-5} 5^1}{1!} + \frac{e^{-5} 5^2}{2!} \right] = 1 - (0.0067 + 0.0337 + 0.0842)$$

$$P(X \geq 3) = 0.8754 = \%87.54$$

olarak bulunur.

Örnek 2: Bir ülkede havalanan her 1000 uçaktan biri mecburi iniş yapmaktadır. Buna göre rastgele seçilen bir ayda, 4000 uçaktan en çok ikisinin mecburi iniş yapması ihtimali,

$$p = 1/1000, n = 4000, \lambda = np = 0.001(4000) = 4 \text{ olduğu için,}$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 2) = \left[\frac{e^{-4} 4^0}{0!} + \frac{e^{-4} 4^1}{1!} + \frac{e^{-4} 4^2}{2!} \right]$$

$$P(X \leq 2) = 0.0183 + 0.0733 + 0.1465 = 0.2381$$

olarak hesaplanır.

Örnek 3: Bir şirketin imal ettiği pillerin %3'ünün kusurlu olması hâlinde, 100 pilden meydana gelen bir örnekte;

a) 0

b) 1

c) 3

d) 3'ten daha fazla

e) 2 veya daha fazla pilin kusurlu olma ihtimalini bulunuz.

$$\text{Çözüm: } p = 0.03, \quad n = 100, \quad \lambda = np = 3$$

$$\text{a) } P(X = 0) = \frac{e^{-3} 3^0}{0!} = 0.0498$$

$$\text{b) } P(X = 1) = \frac{e^{-3} 3^1}{1!} = 0.1494$$

$$\text{c) } P(X = 3) = \frac{e^{-3} 3^3}{3!} = 0.2240$$

$$\text{d) } P(X = 2) = \frac{e^{-3} 3^2}{2!} = 0.2240$$

$$P(X > 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

$$P(X > 3) = 1 - (0.0498 + 0.1494 + 0.2240 + 0.2240) = 0.3528$$

$$e) P(X \geq 2) = 1 - [P(0) + P(1)] = 1 - (0.0498 + 0.1494) = 0.8008$$

Örnek 4: İstatistiklere göre bir ülkede 100000 kişiden 2'si kaza neticesi boğulmaktadır. Bu ülkenin 200000 nüfuslu bir şehrinde söz konusu kaza sayısının yılda;

a) 0

b) 2

c) 4 ile 6 arasında

d) 3'ten az olma ihtimalini bulunuz.

$$\text{Çözüm: } p = \frac{2}{100000}, \quad n = 200000, \quad \lambda = 200000 \left(\frac{2}{100000} \right) = 4$$

$$a) P(X = 0) = \frac{e^{-4} 4^0}{0!} = 0.0183$$

$$b) P(X = 2) = \frac{e^{-4} 4^2}{2!} = 0.1465$$

$$c) P(4 \leq X \leq 6) = P(X = 4) + P(X = 5) + P(X = 6)$$

$$P(4 \leq X \leq 6) = \frac{e^{-4} 4^4}{4!} + \frac{e^{-4} 4^5}{5!} + \frac{e^{-4} 4^6}{6!} = 0.1954 + 0.1563 + 0.1042 = 0.4559$$

$$d) P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X < 3) = \frac{e^{-4} 4^0}{0!} + \frac{e^{-4} 4^1}{1!} + \frac{e^{-4} 4^2}{2!} = 0.0183 + 0.0733 + 0.1465 = 0.2381$$

Örnek 5: Bir hastanenin acil servisine günde ortalama 60 hasta gelmektedir. Buna göre, günün belli bir saatinde acil servise en az iki hasta gelmesi ihtimalini bulunuz.

Çözüm: Günde 60 hasta geldiğine göre, 1 saatte ortalama olarak $60/24 = 2.5$ hasta geliyor demektir. Yani, $\lambda = 2.5$ olur. Buna göre,

$$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$$

$$P(X \geq 2) = 1 - \left[\frac{e^{-2.5} 2.5^0}{0!} + \frac{e^{-2.5} 2.5^1}{1!} \right] = 1 - (0.0821 + 0.2052) = 0.7127$$

olarak hesaplanır.



Bireysel Etkinlik

- Kesikli tesadüfi değişken kavramını tartışınız. Çevrenizde gözlenebilecek kesikli tesadüfi değişkenlerden örnekler veriniz.
- Binom olaylarına örnekler veriniz.
- Poisson olaylarına örnekler veriniz.
- Binom ve Poisson dağılımlarını karşılaştırınız.



Özet

- Tekrarlanan olaylara ilişkin sonuçların kesikli değişken değerleri olması hâlinde, bu tür olayların sonuçlarının gerçekleşme ihtimallerine ait dağılımlara kesikli ihtimal dağılımları denir.
- Kesikli değişken belirli değerler arasında sadece tam sayı değerler alabilen değişkendir. Mesela 0 ile 5 değerleri arasında 0 ve 5 dâhil edilirse sadece 0, 1, 2, 3, 4 ve 5 gibi belirli tamsayı değerleri vardır. Kesikli değişken değerleri genellikle sayımla elde edilirler. Mesela bir derslikteki öğrenci sayısı, bir otobüsteki yolcu sayısı, bir ailedeki çocuk sayısı, bir para beş kez atıldığında yazı gelme sayısı, bir sinema salonundaki izleyici sayısı kesikli değişken değerlerine verilebilecek örneklerdendir.
- Sonuçları belirli tam sayı değerleri alan değişkenlere kesikli değişken denir. Kesikli değişken, sonuçları sayımla elde edilen değişkendir. Bu sebeple kesikli değişkenlere ait sonuçlar yalnızca belirli tam sayı değerler alabilir.
- Kesikli tesadüfi değişkenlere ait dağılımların ortalaması $\mu = \sum_{i=0}^n 1_{(i=0)} \cdot i \cdot P(X=i)$ ve standart sapması $\sigma = \sqrt{\sum_{i=0}^n 1_{(i=0)} \cdot i^2 \cdot P(X=i) - \mu^2}$ formülü ile hesaplanır.
- Binom olayları iki sonuçlu olaylardır. Binom olayları tekrarlanabilen olaylardır. Olayın gerçekleşmesi ihtimali denemeden demeye değişmez. X bir binom dağılımı gösteren bir değişken ise olayın n defa tekrarlanması durumunda olayın sonuçlarına ait ihtimaller $P(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$ formülü ile hesaplanır. Binom dağılımının parametresi p'dir. Bu dağılımın ortalaması (μ) = np ve standart sapması (σ) = $\sqrt{np(1-p)}$ eşitliklerle hesaplanır.
- p = 0.5 olduğunda binom dağılımı simetriktir. p > 0.5 olduğunda dağılım sola çarpık, p < 0.5 olduğunda ise sağa çarpıktır.
- Poisson olayları binom olaylarına çok benzer. Binom olayından farkı olayın bir denemede gerçekleşmesi ihtimali oldukça küçüktür.
- n'in büyümesi p'nin de küçülmesi hâlinde, binom formülü yerine poisson formülü kullanılır. Daha net bir ifadeyle, np değeri 5'ten küçük olduğunda binom dağılımı poisson dağılımına dönüşür.
- x bir poisson dağılımı gösteren bir değişken ise olayın n defa tekrarlanması durumunda olayın sonuçlarına ait ihtimaller $P(X=i) = \frac{e^{-\lambda} \lambda^i}{i!}$ formülü ile hesaplanır.
- Poisson formülündeki e sabiti, yaklaşık olarak 2.71828 değerine eşittir.
- Poisson dağılımının ortalaması (μ) = np = λ ve standart sapması (σ) = $\sqrt{\lambda}$ eşitliklerle hesaplanır.

DEĞERLENDİRME SORULARI

(1- 3. soruları aşağıdaki vakaya göre çözünüz.)

Vaka: Bir şehirde yaşayan ailelerde bulunan araba sayılarına göre aşağıdaki ihtimal dağılımı hazırlanmıştır.

Tablo A. Araba Sahipliğine Göre Bir Şehirdeki Ailelerin Dağılımı

Araba Sayısı (X_i)	$P(X_i)$
0	0.45
1	0.40
2	0.15
Toplam	1.00

- Tesadüfi olarak seçilen bir ailede bir ya da daha az araba olması ihtimali aşağıdakilerden hangisidir?
 - 0.40
 - 0.45
 - 0.65
 - 0.75
 - 0.85
- Tablo A'daki verilere göre söz konusu şehirdeki ailelerde ortalama kaç araba vardır?
 - 0.3
 - 0.4
 - 0.7
 - 1.1
 - 1.4
- Tablo A'daki verilere göre söz konusu şehirdeki ailelerdeki araba sayısının standart sapması kaçtır?
 - 0.59
 - 0.65
 - 0.71
 - 0.85
 - 0.92
- Bir zar üç kez atıldığında mümkün sonuç sayısı kaçtır?
 - 167
 - 188
 - 208
 - 216
 - 239

5. Meyvelerin servis edildiği bir masada 10 adet portakal, 12 adet elma, 8 adet mandalina ve 15 adet erik vardır. Bu masadan çıkarılabilecek meyve tabağı sayısı kaçtır?
 - a) 7800
 - b) 8400
 - c) 10500
 - d) 12000
 - e) 14400
6. Elimizde rafa dizilecek 7 kitap vardır. Bu kitapları kaç farklı dizilişle rafa dizebiliriz?
 - a) 4800
 - b) 5040
 - c) 5240
 - d) 5450
 - e) 5720
7. Elimizde hepsi aynı kalınlıkta olan 5 kitap vardır. Ancak rafımızda bunlardan sadece 4'ünü koyacak kadar yer vardır. Sıra önemli olduğunda 5 kitaptan dördü kaç farklı dizilişle rafa yerleştirilir?
 - a) 120
 - b) 133
 - c) 154
 - d) 165
 - e) 180
8. Elimizde hepsi aynı kalınlıkta olan 5 kitap vardır. Ancak rafımızda bunlardan sadece 4'ünü koyacak kadar yer vardır. Sıra önemli olmadığında 5 kitaptan dördü kaç farklı dizilişle rafa yerleştirilir?
 - a) 4
 - b) 5
 - c) 6
 - d) 7
 - e) 8
9. Yapılan bir araştırmada, $p=0.8$ ve $n=95$ olan bir binom dağılımı için hesaplanacak standart sapma değeri aşağıdakilerden hangisidir?
 - a) 1.30
 - b) 2.40
 - c) 3.90
 - d) 4.60
 - e) 5.80

10. Bir fabrikanın üretim bandında herhangi bir günde üretilen ürünlerin %5'inin kusurlu olduğu bilinmektedir. Binom dağılımı gösterdiği farz edilen bu ürünlerden şansa bağlı olarak seçilen üç üründen hiç birinin kusurlu çıkmaması ihtimali aşağıdakilerden hangisidir?
- a) 0.58
 - b) 0.63
 - c) 0.72
 - d) 0.86
 - e) 0.94

Cevap Anahtarı

1.e, 2.c, 3.c, 4.d, 5.e, 6.b, 7.a, 8.b, 9.c, 10.d

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İsmail Hakkı (2018). İşletmelerde Uygulamalı İstatistik (3. baskı) İstanbul: Alfa Yayınları.
- [2] Başar, A., Oktay, E. (2018). Uygulamalı İstatistik – I: Kısa Teorik Bilgiler ve Çözülmüş Problemler (12. Baskı) Erzurum: Erzurum Kültür Eğitim Kitap ve Kirtasiye.
- [3] Daniel, Wayne W., James C. Terrell (1995). Business Statistics: For Management and Economics (7. baskı). Boston: Houghton Mifflin Company.
- [4] Gürtan, Kenan (1982). İstatistik ve Araştırma Metodları. İstanbul: İstanbul Üniversitesi Yayınları.
- [5] Serper, Ö. (2014). Uygulamalı İstatistik I (7. baskı). İstanbul: Ezgi Kitabevi.
- [6] Yüzer, A. F. vd. (2011). İstatistik. Eskişehir: Anadolu Üniversitesi Açık Öğretim Fakültesi Yayınları.

SÜREKLİ TESADÜFİ DEĞİŞKENLER VE NORMAL DAĞILIM



İÇİNDEKİLER

- Sürekli İhtimal Dağılımları
- Normal Dağılımın Karakteristikleri
- Standart Normal Dağılım
- Standart Normal Eğri Alanları



HEDEFLER

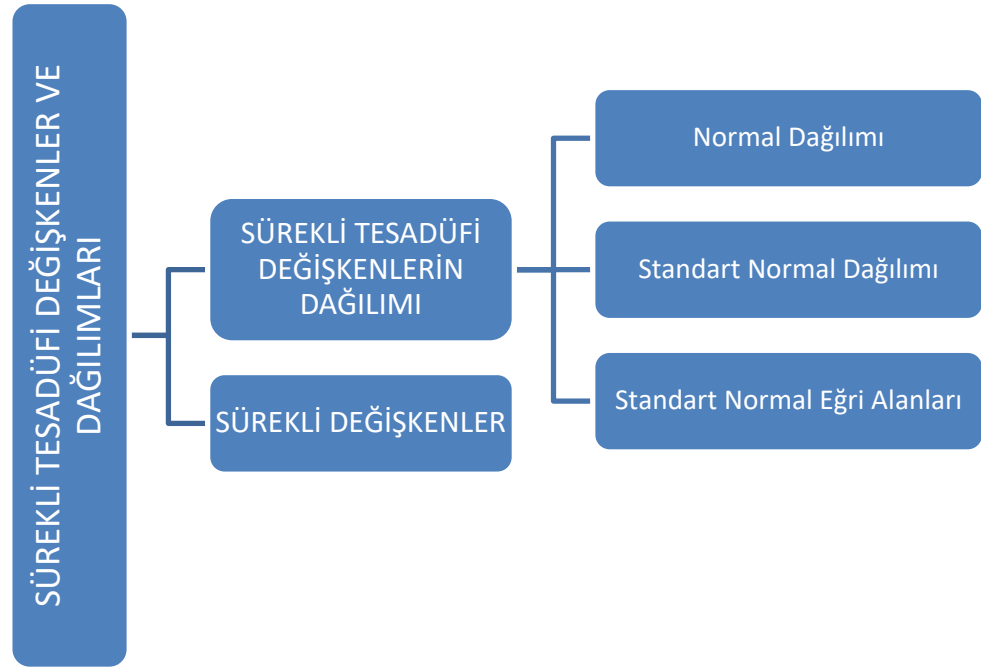
- Bu üniteyi çalıştıktan sonra;
 - Sürekli tesadüfi değişken kavramını açıklayabilecek,
 - Normal dağılım ve normal dağılımın karakteristiklerini kavrayabilecek,
 - Normal eğri alanları ile ilgili hesaplamalar yapabilecek,
 - Normal dağılan değişkenlere ait ihtimal hesaplamaları yapabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Prof. Dr. Erkan
OKTAY

ÜNİTE
13



GİRİŞ



Sürekli değişken, belirli değerler arasında sonsuz sayıda değer alabilen değişkendir.

Bir önceki ünite de kesikli tesadüfi değişkenler ve bu değişkenlere ait bazı ihtimal dağılımlarından bahsedilmişti. Bu ünite de ise sürekli tesadüfi değişkenler ve bu değişkenlere ait dağılımlardan en çok uygulama alanı olan normal dağılımdan bahsedilmiştir.

Sürekli değişken, belirli değerler arasında sonsuz sayıda değer alabilen değişkendir [1]. Mesela 1 ile 2 değerleri arasında 1.02, 1.14, 1.268, 1.8812, 1.9988, ... 1.99998 gibi sonsuz sayıda ölçüm değeri vardır. Sürekli değişken değerleri genellikle ölçülebilen ve tartılabilen değişkenlerin aldıkları değerlerdir. Ölçüm ve tartım aletinin hassasiyetine bağlı olarak virgülden sonraki basamaktaki sayılar şekillenir. Mesela bazı teraziler virgülden sonraki birinci sayıyı gösterecek şekilde ayarlanmıştır. Bir kişinin ağırlığı belirlenirken 65.4 kg ve 76.8 kg gibi sayılarla gösterilir. Bazı teraziler de tarttıkları maddenin kıymetinin büyüklüğüne bağlı olarak daha hassas ölçüm yapmaları için tasarlanmışlardır [2].

SÜREKLİ İHTİMAL DAĞILIMLARI



Sürekli tesadüfi değişkenler, sonuçları ölçüm ve tartımla elde edilmiş, belirli iki değer arasında sonsuz sayıda değer alabilen değişkenlerdir.

Sürekli tesadüfi değişkenler, sonuçları ölçüm ve tartımla elde edilmiş, belirli iki değer arasında sonsuz sayıda değer alabilen değişkenlerdir. Sürekli değişken değerlerinden birinin gözlenmesi ihtimali bu sebeple sonsuzda bir, yani sıfırdır. Bu sebeptendir ki sürekli değişken değerlerini belirli aralıklarda gösterebiliriz. Bu şekilde bir sürekli değişken değerinin belirli bir aralıkta gözlenmesi ihtimali hesaplanabilir.

Tablo 13.1. AUÖF’de Okuyan Erkek Öğrencilerin Ağırlıklarına Göre Dağılımı

<u>Ağırlıklar (kg)</u>	<u>Öğrenci Sayısı</u>	<u>Nispi Frekans</u>
50-60	800	0.08
60-70	2,200	0.22
70-80	4,000	0.40
80-90	2,300	0.23
90-100	700	0.07
Toplam	10,000	1.00

Mesela Atatürk Üniversitesi Açık ve Uzaktan Öğretim Fakültesi’nde okuyan 10,000 erkek öğrencinin ağırlıklarına göre dağılımı Tablo 13.1’deki gibi olsun. Öğrencilerin ağırlıkları 50 kg ile 100 kg arasında değişmektedir. Öğrencilerin ağırlıklarının ortalaması 74.9 kg ve standart sapması 10.3 kg olarak hesaplanmıştır. Öğrencilerin ağırlıkları hassas terazi ile belirlenirse hiçbirinin ağırlığı diğerine eşit olmayacaktır. Bu durumda öğrencilerin ağırlıkları belirli aralıklarda gösterilecek ve bu aralıkta yer alan öğrenci sayısı belirtilecektir.

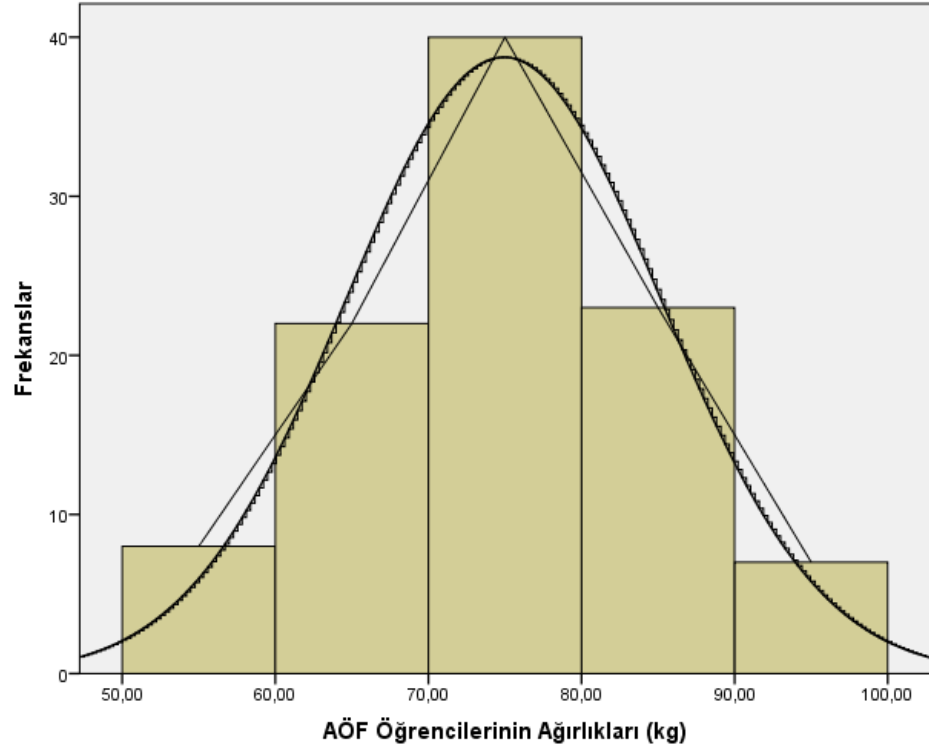
Gruplamada hesap kolaylığı bakımından gruplar genellikle eşit büyüklükte alınır. Bununla beraber uygulamada farklı büyüklükte gruplardan meydana gelen seriler yanında, ilk yahut son sınıfları (veya her ikisi de) açık olan gruplara da rastlanmaktadır. Gruplandırılmış bir seride matematiksel işlemler yapabilmemiz için, her bir sınıfı bir sayı ile temsil etmemiz gerekir. Bu temsili değerler her sınıfın



Histogram, bir diktörtgenler dizisidir. Bu diktörtgenlerin tabanları, gruplandırılmış serideki her bir sınıfın sınıf büyüklüğünü; yükseklikleri ise sınıf frekansını gösterir.

orta noktasındadır ve sınıf değeri olarak adlandırılır. Bu değerler sınıflandırılmış serideki X değerleri gibi işlem görürler [1].

Histogram, bir diktörtgenler dizisidir. Bu diktörtgenlerin tabanları, gruplandırılmış serideki her bir sınıfın sınıf büyüklüğünü; yükseklikleri ise sınıf frekansını gösterir [2]. Görsel 13.1’de Atatürk Üniversitesi Açık ve Uzaktan Öğretim Fakültesi’nde okuyan 10,000 erkek öğrencinin ağırlıklarına göre dağılımını gösteren bir frekans histogramı gösterilmiştir:



Görsel 13.1. AUÖF’de Okuyan Erkek Öğrencilerin Ağırlıklarına Göre Dağılımı

Frekans histogramını oluşturan diktörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe frekans poligonu denir. Histogramların kapladığı alan ile poligonun altında kalan alan birbirine eşittir. Frekans poligonu kırık çizgilerden oluşur. Frekans poligonundaki kırık çizgilerden minimum sapma ile geçen eğri ise frekans eğrisidir. Dikey ekseninde nispi frekanslar gösterildiğinde elde edilen eğriye *ihtimal dağılım eğrisi* denir [3].

İhtimal dağılım eğrisinin altında kalan alan 1’e eşittir. Bir başka ifadeyle X tesadüfi değişkene ait tüm sonuçların gözlenmesi ihtimalleri toplamı 1’e eşittir. Sürekli değişkenine ait bir değer belirlenmesi ihtimali 0 ile 1 arasındadır. Sürekli tesadüfi değişkenin ihtimal dağılım eğrisine ihtimal yoğunluk *fonksiyonu* da denir [4].

NORMAL DAĞILIMIN KARAKTERİSTİKLERİ

İstatistik analizinin temelini teşkil eden normal dağılım sürekli bir ihtimal dağılımıdır. Yani normal dağılımı meydana getiren birimler ölçme yoluyla elde edilmiş verilerdir ve $-\infty$ ile $+\infty$ arasında sonsuz sayıda değer alabilirler. Normal



Frekans histogramını oluşturan diktörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe frekans poligonu denir.

dağılımın moment çarpıklık katsayısı $\alpha_3 = 0$ 'dır. Yani normal dağılım simetriktir. Normal dağılımın moment basıklık katsayısı ise $\alpha_4 = 3$ 'tür. Diğer bütün dağılımların basıklık ölçüleri bu katsayı ile mukayese edilir. Normal dağılım eğrisi aşağıdaki fonksiyonla temsil edilir:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu_x}{\sigma_x} \right)^2}$$

Bu fonksiyondaki e, yaklaşık olarak 2.71828'e eşit matematiksel sabit; π 3.14159'a eşit matematiksel sabit; μ anakütle ortalaması; σ_x anakütle standart sapması ve X herhangi bir sürekli tesadüfi değişkendir. Görüldüğü gibi anakütle ortalaması ve standart sapması bilinen X değerleri için ihtimal hesabı yapılabilir. Bu yüzden normal dağılımın μ_x ve σ_x gibi iki parametresi vardır [4].

Normal eğri altındaki toplam alan 1'e eşittir. Normal dağılımda herhangi bir X sürekli değişkeninin nokta tahmini sıfırdır. Çünkü normal eğri altında sonsuz sayıda X noktaları söz konusudur. Bu yüzden ancak herhangi bir X değerinin X_1 ile X_2 arasında bulunma ihtimali hesaplanabilir. Bunun hesaplanması için fonksiyonun X_1 'den X_2 'ye integralinin alınması gerekir. Bununla birlikte nokta tahmini yapılacağı zaman, verilen X sürekli değişkenine 0.5 değeri eklenip çıkarılarak bir aralık tarif edilir ve daha sonra tarif edilen aralığın ihtimali hesaplanabilir [5].

Anakütle ortalaması ve standart sapmasının farklı olduğu her problem için ayrı bir integrasyon işlemi uygulamak gerekecektir. Ayrıca bu eğrinin integral hesapları da ileri matematik işlemlere dayanmakta ve fazla zaman almaktadır. Bu sebeple bütün problemlerde kullanılabilecek standart bir fonksiyon geliştirilmiştir.

STANDART NORMAL DAĞILIM

Normal dağılım fonksiyonundaki

$$\frac{x - \mu_x}{\sigma_x}$$

ifadesini Z ile gösterirsek, Z değerleri dağılımının ortalaması $\mu_z = 0$ ve standart sapması $\sigma_z = 1$ 'e eşitlendiğinde; normal dağılım, standart normal dağılıma dönüşür. Bu durumda Z değişkeninin standart normal dağılım fonksiyonu,

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} Z^2}$$

gibi daha basit bir şekil alır. Standart normal dağılımın bu özelliğinden hareketle tablolar geliştirilmiştir [4]. Ünitinin sonunda bu tablolardan birisi yer almaktadır. Ünite sonundaki Ek Tablo 1, verilen bir Z değeri ile $Z = 0$ aralığına düşen alanı bulmamızı sağlar.

Tablodaki Z değerleri virgülden sonra iki basamak yürütülecek şekilde takdim edilmiştir. Virgülden sonraki birinci basamağa kadar olan Z değeri tablonun birinci sütunundan bulunur. Bu değer bakacağımız satırı gösterir. Z değerinin

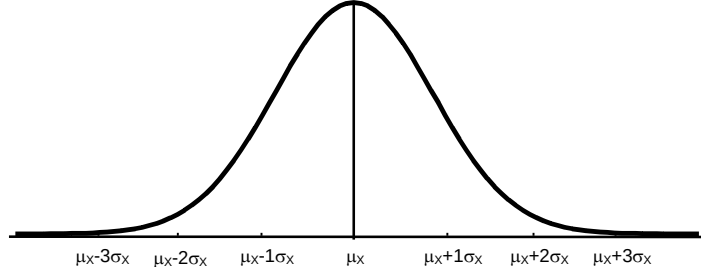


İstatistik analizinin temelini teşkil eden normal dağılım, sürekli bir ihtimal dağılımıdır.



Z değerleri dağılımının ortalaması $\mu_z = 0$ ve standart sapması $\sigma_z = 1$ 'e eşitlendiğinde normal dağılım, standart normal dağılıma dönüşür.

virgülden sonraki ikinci basamağında yer alan değerini diğer sütunların başlarında arayacağız. Bu değer bakacağımız sütunu gösterir. Bakacağımız satır ve sütunun kesiştiği yerdeki ihtimal 0 ile Z değeri arasında bir Z değeri gözlemleme ihtimalini gösterir [6].



Standart normal eğride toplam alanın %68.27'si $\pm 1\sigma_x$, ve %95.54'ü $\pm 2\sigma_x$ ve %99.73'ü $\pm 3\sigma_x$ aralığında bulunur [3].

STANDART NORMAL EĞRİ ALANLARI

Standart normal dağılım için hazırlanan tablolardan yararlanabilmek için verilen X değerlerinin standart Z değerlerine dönüştürülmesi gerekir. X sürekli değişkeninin gözlenmesi ihtimali bu Z değerlerinin gösterdiği alandan hesaplanır. Üitenin sonunda verilen Z tablosundaki Z değerleri 0 ile 3.09 arasındadır. Tablonun birinci sütununda virgülden sonra bir basamak yürütülmüş Z değerleri vardır. Virgülden sonraki ikinci basamak diğer sütunlardadır. Mesela $Z = 0$ ile $Z = 2.14$ arasında bir Z değeri görme ihtimalini bulmak istediğimizde Görsel 13.2'de görüldüğü üzere tablonun ilk sütunundaki 2.1 değerinin bulunduğu satır ile başında 0.04 bulunan sütunun kesişme noktasındaki değere bakarız. $Z = 2.14$ 'ün gösterdiği normal eğri alanı 0.4838'dir [5].



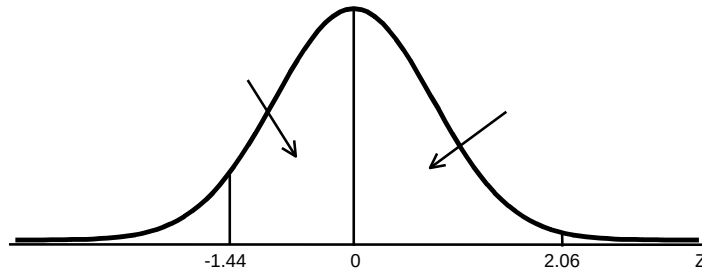
Standart normal eğride toplam alanın %68.27'si $\pm 1\sigma_x$, ve %95.54'ü $\pm 2\sigma_x$ ve %99.73'ü $\pm 3\sigma_x$ aralığında bulunur.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916

Görsel 13.2. Z = 0 ile Z = 2.14 Arasında Bir Z Değeri Görülme İhtimalinin Ek Tablo 1'den Bulunması

Örnek 1: Z = -1.44 ile Z = 2.06 arasındaki alanı bulunuz.

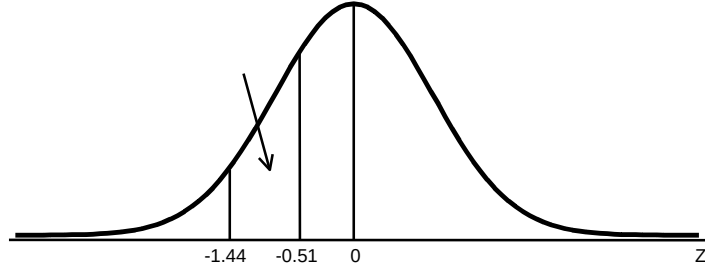
Z değerleri tablosu yardımıyla söz konusu alanı kolayca bulabiliriz. Önce tablodan Z = -1.44'ün gösterdiği alanı tespit ederiz. Bu alan Z = -1.44 ile Z = 0 arasındaki alandır. Daha sonra Z = 2.06 ile Z = 0 arasındaki alanı buluruz. O hâlde Z = -1.44 ile Z = 2.06 arasındaki alan az önce tespit edilen iki alanın toplamına eşittir. Aşağıdaki grafik bu alanı göstermektedir:



$$P(-1.44 \leq Z \leq 2.06) = 0.4251 + 0.4803 = 0.9054$$

Örnek 2: Z = -1.44 ile Z = -0.51 arasındaki alanı bulunuz.

Önce Z = -1.44'ün daha sonra Z = -0.51'in gösterdiği alanı tespit ederiz. Büyük olan alandan küçük olan alan çıkarıldığında istenen alan bulunmuş olur. Aşağıdaki grafik bu alanı göstermektedir:

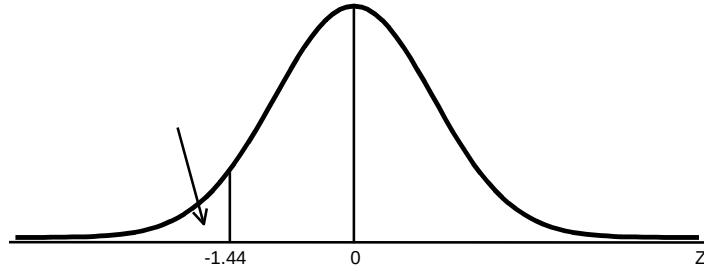


$$P(-1.44 \leq Z \leq -0.51) = 0.4251 - 0.1950 = 0.2301$$

Eğer iki pozitif Z değeri arasındaki alan istenirse yine büyük olan alandan küçük olan alan çıkarılarak istenen alan bulunur.

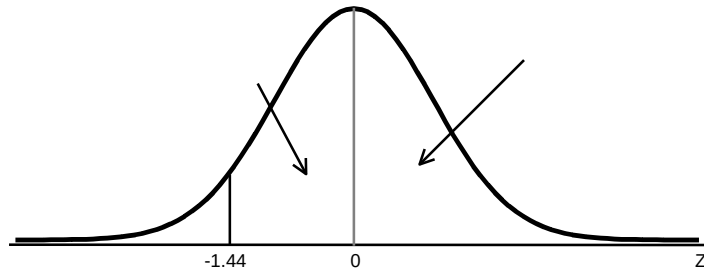
Örnek 3: $P(Z \leq -1.44)$ değerini bulunuz.

Normal eğri altındaki alan 1'e eşittir. Eğri tam ortadan ikiye bölündüğünde sol taraftaki alan 0.5'e eşit olur. $Z = -1.44$ ile $Z = 0$ arasındaki alan 0.4251 olduğuna göre, $P(Z \leq -1.44) = 0.5 - 0.4251 = 0.0749$ olur. Bu alanı gösteren grafik aşağıdadır.



Örnek 4: $P(Z \geq -1.44)$ değerini bulunuz.

$Z = -1.44$ ile $Z = 0$ arasındaki alan 0.4251'dir. Normal eğrinin diğer yarısı 0.5'e eşit olduğuna göre, istenen alan bu iki alanın toplamıdır. Aşağıdaki grafik bu alanı göstermektedir:



$$P(Z \geq -1.44) = 0.4251 + 0.5 = 0.9251$$

Örnek 5: Bir imalat prosesinde üretilen mamullerin ortalama ağırlığı 5 kg ve standart sapması 0.15 kg'dır. Söz konusu imalat prosesinden tesadüfi olarak seçilen bir malın 5.05 kg'dan fazla olma ihtimali % kaçtır?

$X = 5.05$ 'in standart değeri,

$$\frac{X - \mu_x}{\sigma_x} = \frac{5.05 - 5}{0.15} \cong 0.33 \text{ 'tür. Buna göre aranan ihtimal,}$$

$$P(X \geq 5.05) = P(Z \geq 0.33) = 0.5 - 0.1293 = 0.3707$$

olarak hesaplanır.

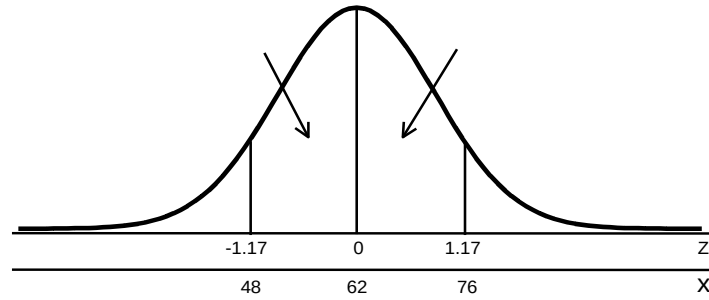
Örnek 6: Bir okuldaki 1200 öğrencinin ortalama ağırlığı 62 kg ve ağırlıkların standart sapması 12 kg'dır. Ağırlıkların normal dağılım gösterdiğini farz ederek kaç öğrencinin ağırlığının 48 kg ile 76 kg arasında olduğunu hesaplayınız.

$\mu = 62$ kg, $\sigma = 12$ kg olduğuna göre,

$$Z_1 = \frac{48 - 62}{12} = -1.17, \quad Z_2 = \frac{76 - 62}{12} = 1.17$$

$$P(48 < X < 76) = P(-1.17 < z < 1.17) = 2[P(0 < z < 1.17)]$$

$$P(48 < X < 76) = 2(0.3790) = 0.7580$$



$E(48 < X < 76) = n[P(48 < X < 76)] = 1200(0.7580) = 909.6 \cong 910$ öğrencinin ağırlığının 48 ile 76 kilo arasında olması beklenir.

Örnek 7: Bir biyoloji sınavında öğrencilere 10 soru sorulmuş ve bu sorulara verilen doğru cevap sayısına göre puanlar 0, 1, 2, ..., 10 şeklinde verilmiştir. Söz konusu sınavında ortalama puan 6.7 ve standart sapma 1.2 puan olarak tespit edilmiştir. Puanların normal dağılım gösterdiğini farz ederek;

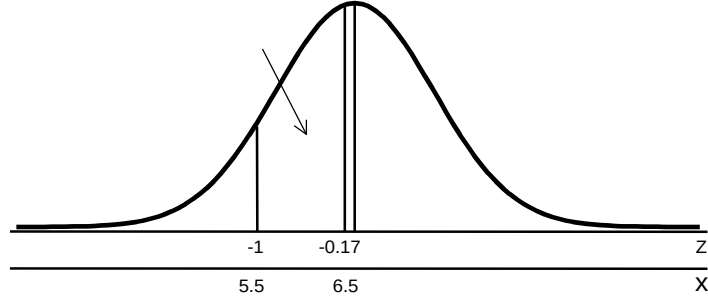
- Puanı 6 olan öğrencilerin yüzdesini,
- Sınıftaki en düşük %10 puan içerisindeki en yüksek puanı,
- Sınıftaki en yüksek %10 puan içerisindeki en düşük puanı

bulunuz.

a) Normal dağılımda nokta tahmini yapılamaz. Ancak X değişkeni süreklilik düzeltilmesine tabi tutulursa tahmin yapılabilir. Kesikli değişken değerine 0.5 değeri ilave edilerek ve çıkarılarak süreklilik düzeltilmesi yapılabilir. Bu sebeple $X = 6$ değerinden 0.5 çıkarılarak $X_1 = 5.5$ değeri ve 0.5 değeri ilave edilerek $X_2 = 6.5$ bulunur ve bu değerler arasında bir değer bulma ihtimali hesaplanabilir.

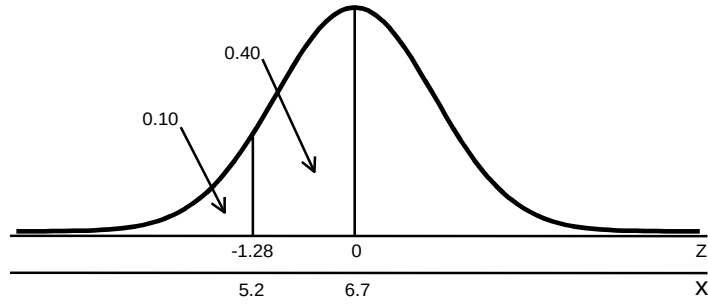
$$P(5.5 < X_i < 6.5) = P(-1 < Z_i < -0.17)$$

$$P(5.5 < X_i < 6.5) = 0.3413 - 0.0675 = 0.2738$$



Puanı 6 olan öğrencilerin yüzdesi %27.38'dir.

b) En düşük %10 içerisindeki en yüksek puanı Z cetvelindeki 0.4000'e isabet eden Z değerini tespit ederek hesaplarız. Z tablosunda 0.40'luk alana tekabül eden Z değeri -1.28'dir.

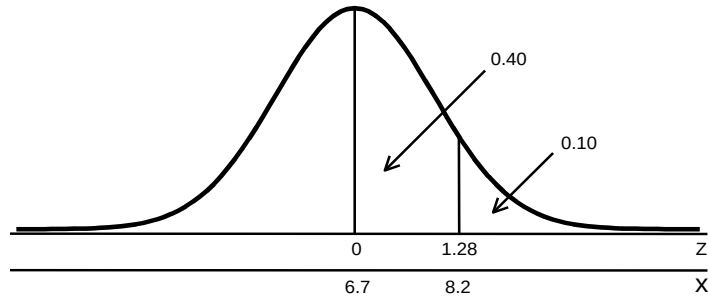


Z formülünde bilinenler yerine yazıldığında,

$$-1.28 = \frac{X - 6.7}{1.2} \Rightarrow X \approx 5.2$$

olarak hesaplanır.

c) En yüksek %10 içerisindeki en düşük puanı Z cetvelindeki 0.4000'e isabet eden Z değerini tespit ederek hesaplarız. Z tablosunda 0.40'luk alana tekabül eden Z değeri 1.28'dir.

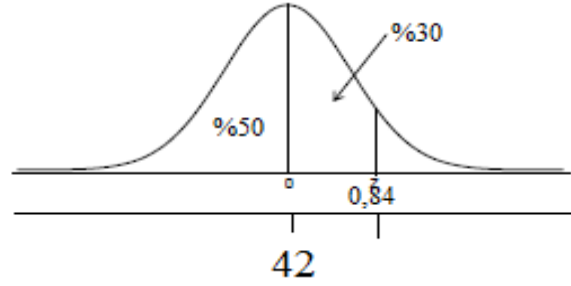


Z formülünde bilinenler yerine yazıldığında,

$$1.28 = \frac{X - 6.7}{1.2} \Rightarrow X \approx 8.2$$

olarak hesaplanır.

Örnek 8: Bir şirkette çalışan personelin yaş ortalaması 42 ve yaşların standart sapması 8'dir. Yaşların normal dağıldığı farz edildiğinde bu şirkette en yaşlı %20'yi oluşturan personelden en genci kaç yaşındadır?



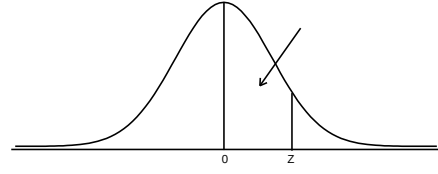
$\mu = 42$ ve $\sigma = 8$ olduğuna göre 42 yaş tam %50'ye tekabül etmektedir. Soruda %80'e tekabül eden yaş sorulduğundan ortalamanın %30 üzerinde olan Z değerinin bulunması gerekmektedir. Z tablosuna bakıldığında 0.3'e en yakın olan değer 0.84 değeridir. O halde,

$$Z = \frac{X - \mu_x}{\sigma_x} = \frac{X - 42}{8} = 0.84$$

ise $X = 48.72$ olacaktır.

Ek Tablo 1: Normal Eğri Alanları

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$$



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990



Bireysel Etkinlik

- Merkezî limit teoremi nedir? Çeşitli örneklerle merkezî limit teoreminin çalışmasını açıklayınız.
- Sürekli tesadüfi değişken kavramını tartışınız. Çevrenizde gözlenebilecek sürekli tesadüfi değişkenlerden örnekler veriniz.



Özet

- İncelenecek bir çok sürekli değişkene ait dağılımın normal dağılıma uygun olduğu kabul edilmektedir.
- Sürekli değişken, belirli değerler arasında sonsuz sayıda değer alabilen değişkendir. Mesela 1 ile 2 değerleri arasında 1.02, 1.14, 1.268, 1.8812, 1.9988, ... 1.99998 gibi sonsuz sayıda ölçüm değeri vardır. Sürekli değişken değerleri genellikle ölçülebilen ve tartılabilen değişkenlerin aldıkları değerlerdir. Ölçüm ve tartım aletinin hassasiyetine bağlı olarak virgülden sonraki basamaktaki sayılar şekillenir. Mesela bazı teraziler virgülden sonraki birinci sayıyı gösterecek şekilde ayarlanmıştır. Bir kişinin ağırlığı belirlenirken 65.4 kg, 76.8 kg gibi sayılarla gösterilir. Bazı teraziler de tarttıkları maddenin kıymetinin büyüklüğüne bağlı olarak daha hassas tartı yapmaları için tasarlanmışlardır.
- Sürekli tesadüfi değişkenler; sonuçları ölçüm ve tartımla elde edilmiş, belirli iki değer arasında sonsuz sayıda değer alabilen değişkenlerdir. Sürekli değişken değerlerinden birinin gözlenmesi ihtimali bu sebeple sonsuzda bir, yani sıfırdır. Bu sebeptendir ki sürekli değişken değerlerini belirli aralıklarda gösterebiliriz. Bu şekilde bir sürekli değişken değerinin belirli bir aralıkta gözlenmesi ihtimali hesaplanabilir.
- Normal dağılan değişkenlere ait değerler standart değerlere dönüştürülebilir. Sürekli değişken değerleri standart değerlere dönüştürüldükten sonra ihtimal hesaplamalarına konu olabilir. Standart normal dağılım için hesaplanmış olan hazır tablolar kullanılarak sürekli bir değişkene ait ihtimal hesaplamaları yapılabilir. Dağılımın ortalaması ve standart sapması biliniyorsa standart değerler kullanılarak ihtimalin hesaplanmasına konu olan sürekli değişken değeri de hesaplanabilir.
- Normal dağılım fonksiyonundaki
- $Z = \frac{X_i - \mu}{\sigma}$ değerleri dağılımının ortalaması $\mu_z = 0$ ve standart sapması $\sigma_z = 1$ 'e eşitlendiğinde; normal dağılım, standart normal dağılıma dönüşür. Bu durumda Z değişkeninin standart normal dağılım fonksiyonu,
- $(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$ gibi daha basit bir şekil alır. Standart normal dağılımın bu özelliğinden hareketle tablolar geliştirilmiştir. Ünitinin sonunda bu tablolardan birisi yer almaktadır. Ünite sonundaki Ek Tablo 1, verilen bir Z değeri ile $Z = 0$ aralığına düşen alanı bulmamızı sağlar. Tablodaki Z değerleri virgülden sonra iki basamak yürütülecek şekilde takdim edilmiştir. Virgülden sonraki birinci basamağa kadar olan Z değeri tablonun birinci sütunundan bulunur. Bu değer bakacağımız satırı gösterir. Z değerinin virgülden sonraki ikinci basamağında yer alan değerini diğer sütunların başlarında arayacağız. Bu değer bakacağımız sütunu gösterir. Bakacağımız satır ve sütunun kesiştiği yerdeki ihtimal 0 ile Z değeri arasında bir Z değeri gözlemlene ihtimalini gösterir.

DEĞERLENDİRME SORULARI

1. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre; Z değerinin 0 ile 1.22 arasında bir değere eşit olması ihtimali aşağıdakilerden hangisidir?
 - a) 0.3432
 - b) 0.3888
 - c) 0.4132
 - d) 0.4617
 - e) 0.4945
2. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre; Z değerinin -1.12 ile 0 arasında bir değere eşit olması ihtimali aşağıdakilerden hangisidir?
 - a) 0.2939
 - b) 0.3212
 - c) 0.3461
 - d) 0.3686
 - e) 0.3888
3. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre; $P(0.45 \leq Z \leq 0.98)$ ihtimali aşağıdakilerden hangisidir?
 - a) 0.1629
 - b) 0.2014
 - c) 0.2237
 - d) 0.2434
 - e) 0.2917
4. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre; $P(-1.03 \leq Z \leq -0.48)$ ihtimali aşağıdakilerden hangisidir?
 - a) 0.1642
 - b) 0.2214
 - c) 0.2451
 - d) 0.2649
 - e) 0.2953
5. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre; $P(-1.03 \leq Z \leq 0.98)$ ihtimali aşağıdakilerden hangisidir?
 - a) 0.5892
 - b) 0.6123
 - c) 0.6426
 - d) 0.6850
 - e) 0.7213

6. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre;
 $P(Z \leq 0.98)$ ihtimali aşağıdakilerden hangisidir?
a) 0.6993
b) 0.7352
c) 0.7742
d) 0.8032
e) 0.8365
7. Z, standart normal dağılıma sahip tesadüfi bir değişken olduğuna göre;
 $P(Z \geq 0.98)$ ihtimali aşağıdakilerden hangisidir?
a) 0.1212
b) 0.1635
c) 0.1963
d) 0.2234
e) 0.2683
8. Standart normal dağılımda ortalama etrafındaki 2.5 standart sapma sınırları arasında bir Z değeri gözlemlene ihtimali, bir başka ifadeyle $P(-2.5\sigma \leq \mu \leq 2.5\sigma)$ ihtimali kaçtır?
a) 0.8719
b) 0.8912
c) 0.9123
d) 0.9452
e) 0.9876
9. Atatürk Üniversitesi Açık ve Uzaktan Öğretim Fakültesi'nde okuyan erkek öğrencilerin boylarının 175 cm ortalama ve 10 cm standart sapma ile normal dağılım gösterdiği varsayılın. Bahsedilen öğrencilerden biri tesadüfi olarak seçildiğinde öğrencinin boyunun 170 cm'den uzun olması ihtimali kaçtır?
a) 0.5964
b) 0.6314
c) 0.6915
d) 0.7237
e) 0.7581

10. Atatürk Üniversitesi Açık ve Uzaktan Öğretim Fakültesi'nde okuyan erkek öğrencilerin boylarının 175 cm ortalama ve 10 cm standart sapma ile normal dağılım gösterdiği varsayalım. Bahsedilen öğrencilerden biri tesadüfi olarak seçildiğinde öğrencinin boyunun 160 cm ile 180 cm arasında olması ihtimali kaçtır?
- a) 0.6247
 - b) 0.6721
 - c) 0.7073
 - d) 0.7412
 - e) 0.7839

Cevap Anahtarı

1.b, 2.d, 3.a, 4.a, 5.d, 6.e, 7.b, 8.e, 9.c, 10.a

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İ. H. (2018). İşletmelerde Uygulamalı İstatistik (3. baskı). İstanbul: Alfa Yayınları.
- [2] Başar, A., Oktay, E. (2018). Uygulamalı İstatistik – I: Kısa Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap Kirtasiye.
- [3] Daniel, Wayne W., James C. Terrell (2011). Business Statistics: For Management and Economics (7. baskı). Boston: Houghton Mifflin Company.
- [4] Gürtan, Kenan (1982). İstatistik ve Araştırma Metodları. İstanbul: İstanbul Üniversitesi Yayınları.
- [5] Serper, Ö. (2014). Uygulamalı İstatistik I (7. baskı). İstanbul: Filiz Kitabevi.
- [6] Yüzer, A. F. vd. (2011). İstatistik. Eskişehir: Anadolu Üniversitesi Açık Öğretim Fakültesi Yayınları.

KESİKLİ İHTİMAL DAĞILIMLARIN NORMAL DAĞILIMA YAKLAŞIMI



İÇİNDEKİLER

- Kesikli Dağılımların Normale Yaklaşımı
- Binom Dağılımının Normale Yaklaşımı
- Poisson Dağılımının Normale Yaklaşımı



HEDEFLER

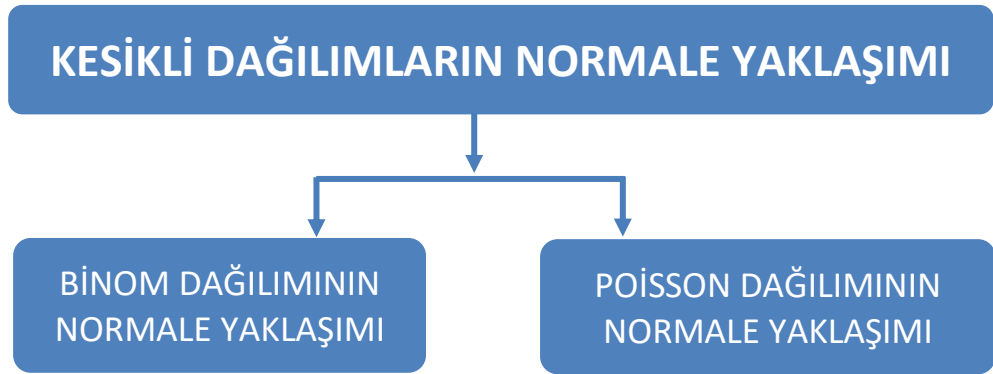
- Bu üniteyi çalıştıktan sonra;
 - Kesikli dağılımların normal dağılıma hangi şartlar altında yaklaştığı anlaşılabilecek,
 - Normale yaklaşan binom olayları için ihtimallerin hesaplanışını kavrayabilecek,
 - Normale yaklaşan poisson olayları için ihtimallerin hesaplanışını kavrayabileceksiniz.



Atatürk Üniversitesi
Açıköğretim Fakültesi

İSTATİSTİĞE GİRİŞ
Prof. Dr. Erkan
OKTAY

ÜNİTE
14





Örnek hacimleri büyük olduğunda binom ve poisson formülleri yardımıyla kesikli değişken değerlerine ait ihtimalleri hesaplamak, uzun ve yorucu hesaplamalar gerektirir.



Ortalama (μ_x) ve standart sapma (σ_x) normal dağılımın parametreleridir.



Örnek hacmi yeterince büyük olduğunda kesikli olayların ortalama ve standart sapması standart değer formülünde yazılarak olayın gerçekleşmesi ihtimali hesaplanabilir.

GİRİŞ

12. ünite de kesikli tesadüfi değişkenler ve bu değişkenlere ait bazı ihtimal dağılımlarından bahsedilmişti. 13. ünite de sürekli tesadüfi değişkenler ve bu değişkenlere ait sürekli dağılımlarından en önemlisi normal dağılımdan bahsedilmişti. Bu ünite de ise kesikli ihtimal dağılımlarından binom ve poisson dağılımının normal dağılıma yaklaşımından bahsedilmiştir. Örnek hacimleri küçük olduğunda binom ve poisson dağılımı formülleri kullanılarak kesikli değişken değerlerine ait ihtimaller hesaplanabilmektedir [1]. Örnek hacimleri büyük olduğunda binom ve poisson formülleri yardımıyla kesikli değişken değerlerine ait ihtimalleri hesaplamak, uzun ve yorucu hesaplamalar gerektirir [2].

KESİKLİ DAĞILIMLARIN NORMALE YAKLAŞIMI

Örnek hacmi n 'in büyük olduğu hallerde kesikli ihtimal dağılımlarına ait formüllerin kullanılması uzun hesaplamalar gerektirir. Örnek hacmi yeterince büyük olduğunda, X değerlerinin dağılımı normal dağılıma yaklaşır. Standart normal değerleri bulmayı sağlayan,

$$Z = \frac{X - \mu_x}{\sigma_x}$$

formülündeki μ_x ve σ_x normal dağılımın parametreleridir. Bu formülde normal dağılım parametreleri yerine gerekli şartları sağlayan kesikli ihtimal dağılımlarının ortalama ve standart sapmaları yazıldığında, kesikli ihtimal değerlerinin kolaylıkla hesaplanabilmesini sağlayacak Z değerleri elde edilir [1].

Kesikli ihtimal dağılımları ile kesikli değişkenin belirli bir değeri alması ihtimali hesaplanabilmektedir. Ancak normal dağılım söz konusu olduğunda belirlenen iki değer arasında sonsuz sayıda değişken değeri olduğu için sürekli değişkenin belli bir değere eşit olması ihtimali sonsuzda bir olur. Yani sıfır olur. Bu sebeple normal dağılımda nokta değerinin ihtimalini bulabilmek için, verilen X değerine 0.5 ilave edilip çıkarılarak belirli bir aralığın tarif edilmesi gerekir. Buna *süreklilik düzeltmesi* denir. Standart Z değerleri hesaplanırken, gerekli hallerde X değişkenine 0.5 ilave edilir ve/veya çıkarılır [3].

Mesela X değişkeni kesikli bir değişken ise, $X=3$ olması ihtimali nasıl bulunacaktır? Bu durumda X değişkenine 0.5 değeri ilave edilip çıkarılarak bir aralık bulunacaktır. Bu aralık 2.5 ile 3.5 aralığı olacaktır. Kesikli X değişkeninin 3'e eşit olması ihtimali denildiğinde X değişkeninin 2.5 ile 3.5 arasında olması ihtimali kastedilmektedir.

BİNOM DAĞILIMININ NORMALE YAKLAŞIMI

12. ünite de binom olayı ile ilgili şu bilgiler verilmişti:

- Binom olaylarının sonuçları kesikli tesadüfi değişkenlerle ifade edilir.

- Binom olayları iki sonuçlu olaylardır. Sonuçlardan birinin gözlenmesi ihtimali p ile gösterilirse diğer sonucun gözlenmesi ihtimali $1 - p$ ile ifade edilir.
- Olayların gözlenmesi ihtimali p denemeden denemeye değişiklik göstermez.
- Olayların tekrarlanması birbirinden bağımsızdır. Olay n defa tekrarlandığından x 'in 0 'dan başlayıp n 'e kadar olan sonuçları için sonuçların gerçekleşmesi ihtimali

$$P(x_i) = \binom{n}{x_i} p^{x_i} (1 - p)^{n-x_i}$$

formülü ile hesaplanır. Örnek hacmi büyük olduğunda bu formül yardımıyla ihtimal hesaplamak oldukça zordur [4].

Örnek hacimlerinin 30'dan büyük ve p 'nin 0.5'e yakın olduğu hallerde binom problemleri genellikle normale yaklaştırma yolu ile çözülmektedir. Örnek hacmi ile p değerinin çarpımı 5'e eşit veya daha büyük olursa binom dağılımı normale yaklaşır. Simetrik binom dağılımlarında, n küçük dahi olsa normale yaklaşım binom formülüne yakın sonuçlar verdiği halde, p 'nin 0'a veya 1'e yaklaşması halinde normale yaklaştırma yoluyla elde edilen sonuçlara güvenilmez. Zira bu hallerde eğri sağa veya sola çarpık olacağından normal eğri cetveliyle hesaplanacak ihtimaller, gerçek ihtimallerden sapma gösterecektir. Şu var ki, n 'in çok büyük değerleri için p 'nin çok küçük (veya büyük) olması halinde bile normale yaklaştırma yoluyla gerçeğe yakın sonuçlar elde edilebilir.

X değerlerine ait ihtimallerin hesaplanmasında standart Z değeri formülündeki anakütle ortalaması ve standart sapmasının yerine binom dağılımının ortalama ve standart sapması yazılır. Böylece standart Z değeri formülü,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

şeklini alır. Gerekğinde X değişkeninden 0.5 çıkarılır ve/veya ilave edilir [2].

Mesela bir hastanenin kadın doğum polikliniğinde doğum için bekleyen 20 kadın olsun. Yeni doğacak bir çocuğun kız olması ihtimali 0.5 ve erkek olması ihtimali 0.5'tir. Bu olay bir binom olayıdır. Gerçekleşecek 20 doğundan 0 ile 20 arasında kız çocuğu olması ihtimali binom formülü yardımıyla aşağıdaki Tablo 14.1 deki gibi bulunur:

Tablo 14.1. Yirmi doğum vakasında kız çocuğu olması ihtimalleri

Kız Çocuğu Sayısı (X_i)	Olayın İhtimali, $P(X_i)$	Kız Çocuğu Sayısı (X_i)	Olayın İhtimali, $P(X_i)$
0	0.0000	11	0.1602
1	0.0000	12	0.1201
2	0.0002	13	0.0739
3	0.0011	14	0.0370



X kesikli değişken değerine 0.5 ilave edilip çıkarılarak belirli bir aralığın tarif edilmesi gerekir. Buna süreklilik düzeltmesi denir.



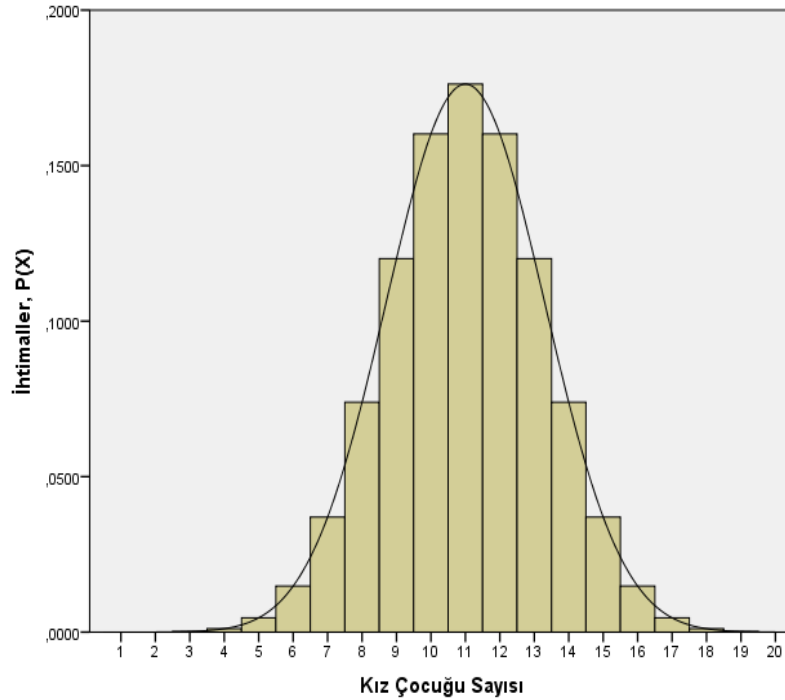
X kesikli değişkeni binom dağılımı gösterirken, n yeterince büyük olduğunda ($n \geq 30$) binom olayı normale yaklaşır.

4	0.0046	15	0.0148
5	0.0148	16	0.0046
6	0.0370	17	0.0011
7	0.0739	18	0.0002
8	0.1201	19	0.0000
9	0.1602	20	0.0000
10	0.1762		



$n \geq 30$ olduğunda binom dağılımının ortalaması ve standart sapması kullanılarak binom olayının ihtimalleri normale yaklaştırılarak hesaplanır.

Binom olayının sonuçları ile ihtimalleri X-Y koordinat sistemine taşındığında Grafik 14.1'deki gibi bir histogram elde edilir. Bu histogramdaki sütunların tepe noktalarında sınıf değerleri işaretlendiğinde söz konusu noktalardan minimum sapma ile geçen frekans eğrisi grafikteki gibi olur. Görüleceği üzere bu eğri normal eğriye oldukça benzemektedir.



Grafik 14.1. Yirmi Doğum Vakasında Kız Çocuğu Olması İhtimalleri

Şimdi binomun normale yaklaşımı yardımıyla yukarıdaki olayın sonuçlarından bazılarını hesaplayalım. Önce binom dağılımının ortalamasını,

$$\mu = np = 20(0.5) = 10$$

ve standart sapmasını,

$$\sigma = \sqrt{np(1-p)} = \sqrt{20(0.5)(1-0.5)} \cong 2.24$$

olarak elde ederiz. Şimdi doğacak 20 çocuktan 12 tanesinin kız olması ihtimalini hesaplayalım. 12 sayısı bir tam sayıdır. Normal dağılımda bir değer gözlenmesi ihtimali sonsuzda bir, yani sıfırdır. Bu sebeple bu değeri nokta olmaktan çıkarıp bir aralık haline getirmemiz gerekir. Bunu yapmak için 12 sayısından önce 0.5'i çıkaracağız sonra ekleyeceğiz. Böylece 12 sayısının yerine ihtimalini hesaplayacağımız aralık 11.5 ve 12.5 aralığı olacaktır. Bir sonraki adımda binom

dağılımının ortalamasını ve standart sapmasını kullanarak 11.5 ve 12.5 değerlerini standart değere dönüştüreceğiz. 11.5 değeri,

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{11.5 - 10}{2.24} \cong 0.67$$

değerine dönüşürken; 12.5 değeri,

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{12.5 - 10}{2.24} \cong 1.12$$

değerine dönüşür. Şimdi binomun normale yaklaşımını kullanarak doğacak 20 çocuktan 11.5 ile 12.5 arasında kız çocuğu doğması ihtimali,

$$P(11.5 \leq X_i \leq 12.5) = P(0.67 \leq Z_i \leq 1.12) = 0.3686 - 0.2486 = 0.12$$

olarak elde edilir. Bu sonuç binom formülü yardımıyla elde edilen sonuca çok yakındır. Konuyu örneklerle daha da açabiliriz.

Örnek 1: Bir fabrikanın ürettiği mamullerin %10'u standartlara uyma-maktadır. Bu fabrikanın imalatından alınan 150 birimlik mamul örneğinde 10 mamulün standartlara uymaması ihtimali % kaçtır?

Dağılımın ortalaması $\mu = np = 150(0.10) = 15$ ve

Standart sapması $\sigma = \sqrt{np(1-p)} = \sqrt{150(0.10)(1-0.10)} \cong 3.67$ 'dir.

Normal dağılımda herhangi bir noktanın ihtimali sıfırdır. Bu yüzden $X = 10$ değerine 0.5 ilave edilip çıkarılarak $X_1 = 10.5$ ve $X_2 = 9.5$ değerleri elde edilir. Daha sonra bu X değerleri,

$$Z_1 = \frac{10.5 - 15}{3.67} \cong -1.23$$

$$Z_2 = \frac{9.5 - 15}{3.67} \cong -1.50$$

standart değerlerine dönüştürülür. Buna göre aranan ihtimal,

$$P(X = 10) = 0.4332 - 0.3907 = 0.0425$$

olarak hesaplanır.

Örnek 2: Bir önceki örneğe göre 10'dan az mamulün standartlara uymaması ihtimali kaçtır?

10'dan az denildiği için ihtimal hesabına 10 dahil değildir. 10'un alt sınırı 9.5'te biteceği için standart Z değerinin hesaplanmasında 10 yerine 9.5 alınır. 9.5'in standart değeri -1.50'dir. Dolayısıyla 10'dan az mamulün standartlara uymaması ihtimali,

$$P(X < 10) = P(Z < -1.50) = 0.5 - 0.4332 = 0.0668$$

olarak elde edilir.

Örnek 3: Bir madeni paranın 200 kere atılması halinde;

a) 80 ile 120 arasında yazı gelme (80 ve 120 dahil),

b) 90'dan az yazı gelme,

c) 85'ten az veya 115'ten çok yazı gelme,

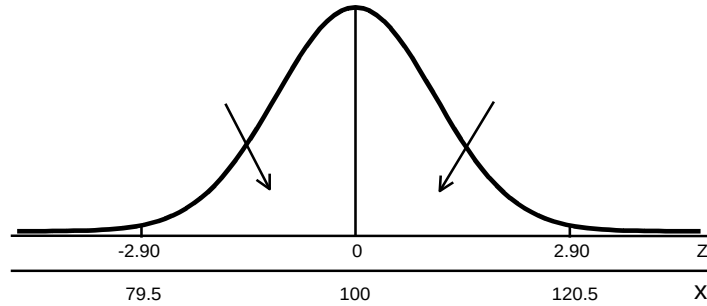
d) 100 tane yazı gelme ihtimalini bulunuz.

$$n = 200, p = 0.5, \mu = np = 200(0.5) = 100,$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{200(0.5)(1-0.5)} = 7.07$$

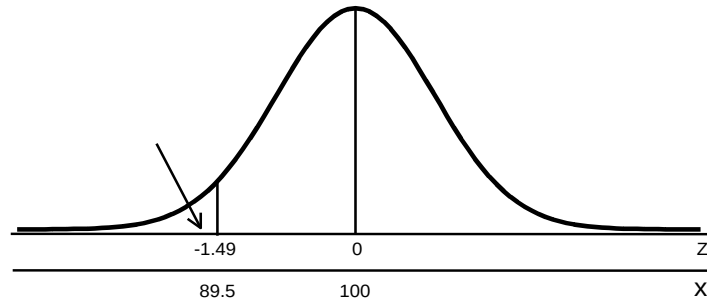
$$\text{a) } Z_1 = \frac{79.5 - 100}{7.07} = -2.90 \text{ ve } Z_2 = \frac{120.5 - 100}{7.07} = 2.90$$

$$P(79.5 < X < 120.5) = P(-2.90 < Z < 2.90) = 2(0.4981) = 0.9962$$



$$Z_1 = \frac{89.5 - 100}{7.07} = -1.49$$

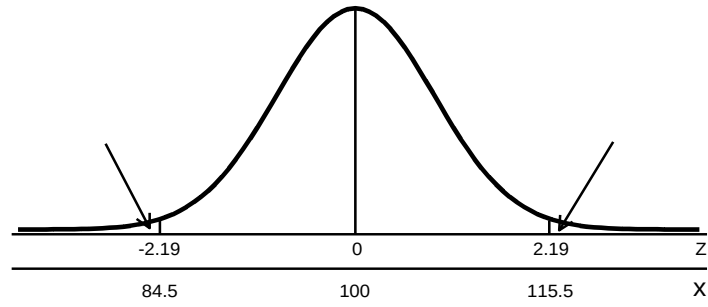
$$P(X < 89.5) = P(Z < -1.49) = 0.5 - 0.4319 = 0.0681$$



$$\text{c) } Z_1 = \frac{84.5 - 100}{7.07} = -2.19 \text{ ve } Z_2 = \frac{115.5 - 100}{7.07} = 2.19$$

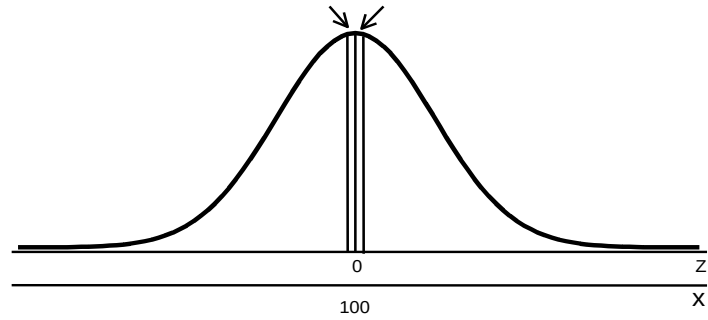
$$P(X < 84.5) + P(X > 115.5) = P(Z < -2.19) + P(Z > 2.19)$$

$$P(X < 84.5) + P(X > 115.5) = 2(0.5 - 0.4857) = 0.0286$$



$$d) Z_1 = \frac{99.5 - 100}{7.07} = -0.07 \text{ ve } Z_2 = \frac{100.5 - 100}{7.07} = 0.07$$

$$P(99.5 < X < 100.5) = P(-0.07 < Z < 0.07) = 2(0.0279) = 0.0558$$



Örnek 4: Bir doğru – yanlış testinde herhangi bir öğrencinin;

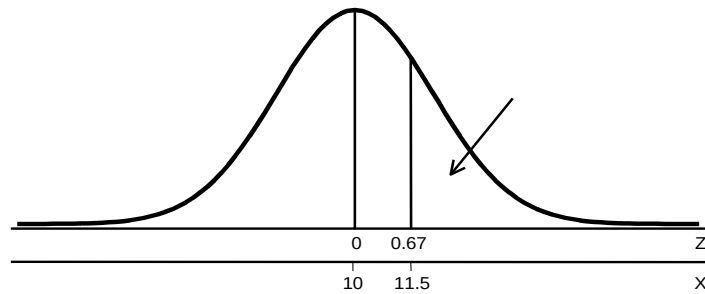
a) 20 sorudan 12'sine veya daha fazlasına,

b) 40 sorudan 24'üne veya daha fazlasına doğru cevap verme ihtimalini bulunuz.

a) $n = 20$, $p = 0.5$, $\mu = np = 20(0.5) = 10$,

$$\sigma = \sqrt{np(1-p)} = \sqrt{20(0.5)(1-0.5)} = 2.24, Z = \frac{11.5 - 10}{2.24} = 0.67$$

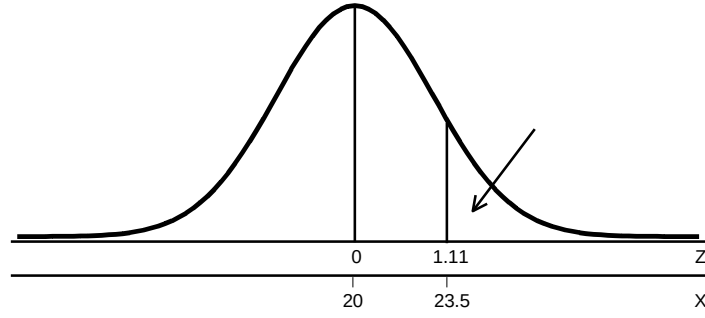
$$P(X > 11.5) = P(Z > 0.67) = 0.5 - 0.2486 = 0.2514$$



b) $n = 40$, $p = 0.5$, $\mu = np = 40(0.5) = 20$,

$$\sigma = \sqrt{40(0.5)(1-0.5)} = 3.16, Z = \frac{23.5 - 20}{3.16} = 1.11$$

$$P(X > 23.5) = P(Z > 1.11) = 0.5 - 0.3665 = 0.1335$$



Örnek 5: Bir fabrikanın imal ettiği ampullerin %10'u kusurlu çıkmaktadır. Bu ampullerden alınan 200 birimlik bir şans örneğinde;

a) En fazla 30,

b) 30 ile 50 arası,

c) 55 veya daha fazla ampulün kusurlu olması ihtimalini bulunuz.

$$n = 200, p = 0.1, \mu = np = 200(0.1) = 20,$$

$$\sigma = \sqrt{200(0.1)(1 - 0.1)} = 4.24$$

$$a) Z = \frac{30.5 - 20}{4.24} = 2.48$$

$$P(X < 30.5) = P(Z < 2.48) = 0.5 + 0.4934 = 0.9934$$

$$b) Z_1 = \frac{30.5 - 20}{4.24} = 2.48 \text{ ve } Z_2 = \frac{49.5 - 20}{4.24} = 6.96$$

Z tablosunda $Z = 3.9$ ve daha büyük değerlerin karşılığı 0.5000'dir.

$$P(30 \leq X \leq 50) = P(2.48 < Z < 6.96) = 0.5 - 0.4934 = 0.0066$$

$$c) Z = \frac{54.5 - 20}{4.24} = 8.14$$

$$P(X > 54.5) = P(Z > 8.14) = 0.5 - 0.5 = 0$$

POISSON DAĞILIMININ NORMALE YAKLAŞIMI

Poisson olayları yapıları itibariyle binom olaylarına çok benzer. Binom olayını anlatırken söylenenlerin tümü poisson olayları için de söylenebilir. Ancak poisson olayları binom olaylarının özel bir halidir. Poisson olayları nadir rastlanan olayların ihtimalleridir.

Olay n defa tekrarlandığından x 'in 0'dan başlayıp n 'e kadar olan sonuçları için sonuçların gerçekleşmesi ihtimali,

$$P(x_i) = \frac{e^{-\lambda} \lambda^x}{x!}$$



Poisson olayları binom olaylarının özel bir halidir. Poisson olayları nadir rastlanan olayların ihtimalleridir.

formülü ile hesaplanır [5]. Ancak örnek hacmi büyük olduğunda bu formül yardımıyla ihtimal hesaplamak oldukça zordur.

Poisson dağılımının ortalaması λ 'dır. Bu değer verilmediğinde olayın tekrarlanma sayısı n ve olayın bir defada gözlenmesi ihtimali p çarpılarak elde edilir. Poisson dağılımının standart sapması ise $\sqrt{\lambda}$ 'dır.



$\lambda \geq 5$ olduğunda poisson dağılımının ortalaması ve standart sapması kullanılarak P poisson olayının ihtimalleri normale yaklaştırılarak hesaplanır.

Büyük örnek hacimleri ve çok küçük p değerleri kullanıldığında λ , 5'e eşit veya büyük olursa poisson dağılımı normale yaklaşır. Poisson ihtimalleri normal dağılım varsayımı altında hesaplanırken, standart Z değeri formülündeki anakütle ortalaması ve standart sapmasının yerine poisson dağılımının aritmetik ortalama ve standart sapması yazılır. Bu durumda standart Z değeri,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

formülü yardımıyla hesaplanır. X, süreklilik düzeltmesine tabi tutulacaktır [6].

Örnek 1: Bir otomobil fabrikasında malzeme yokluğu sebebiyle günde, ortalama 12 kez üretim durmaktadır. Buna göre, rastgele seçilen bir günde malzeme yokluğu sebebiyle üretimin 15 kez veya daha az defa durması ihtimalini hesaplayınız.

Ortalama, $\lambda = 12$ ve standart sapma, $\sqrt{\lambda} = \sqrt{12} \cong 3.46$ 'dır. 15 veya az defa üretimin durması ihtimali sorulduğu için 15 ihtimal bölgesindedir. 15'in üst sınırı 15.5 olduğundan bu noktanın standart değeri,

$$Z = \frac{15.5 - 12}{\sqrt{12}} = 1.01$$

olur. Bu noktanın sol tarafındaki alan aranan ihtimaldir. Buna göre 15 veya az defa üretimin durması ihtimali,

$$P(X \leq 15) = P(Z \leq 1.01) = 0.5 + 0.3438 = 0.8438 \text{ olarak hesaplanır.}$$

Örnek 2: Bir önceki örnekte, malzeme yokluğu sebebiyle bir günde tam 12 defa üretimin durması ihtimali % kaçtır?

$$X_1 = 11.5,$$

$$X_2 = 12.5$$

$$Z_1 = \frac{11.5 - 12}{\sqrt{12}} \cong -0.14$$

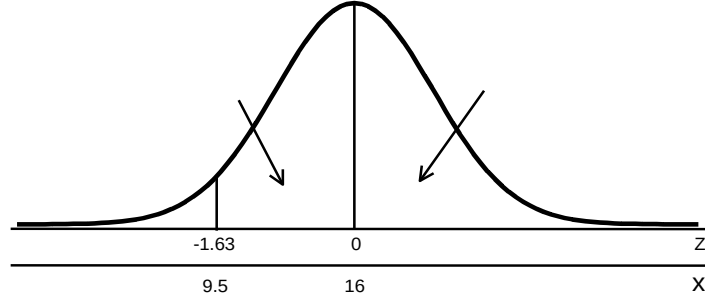
$$Z_2 = \frac{12.5 - 12}{\sqrt{12}} \cong 0.14$$

$$P(X = 12) = P(-0.14 < Z < 0.14) = 0.0557 + 0.0557 = 0.1114 = \%11.14$$

Örnek 3: Bir otomobil galerisi ayda ortalama 16 araba satmaktadır. Buna göre rastgele seçilen bir ay itibarıyla söz konusu galeride 10 veya daha fazla araba satılması ihtimalini hesaplayınız.

$$\lambda = 16, \sqrt{\lambda} = \sqrt{16} = 4, X \geq 10, Z = \frac{X - \lambda}{\sqrt{\lambda}} = \frac{9.5 - 16}{4} = -1.63$$

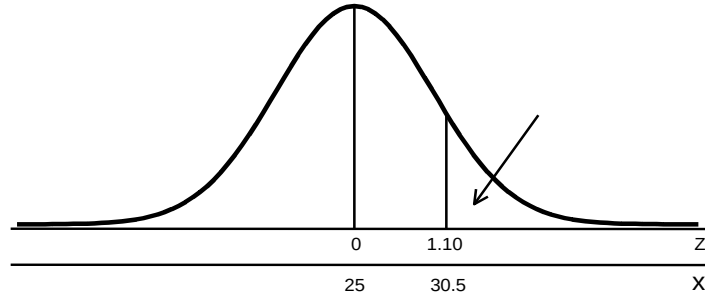
$$P(X > 9.5) = P(Z > -1.63) = 0.4484 + 0.5 = 0.9484$$



Örnek 4: Bir yıl itibarıyla belirli bir kavşakta ortalama 25 kaza olmaktadır. Gelecek yıl bu kavşakta 30'dan fazla kaza olması ihtimali % kaçtır?

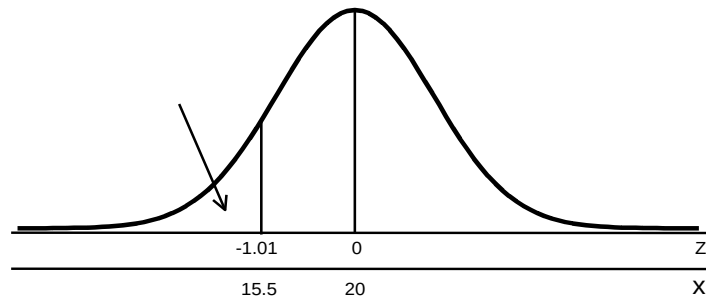
Çözüm: $\lambda = 25$, $\sqrt{\lambda} = \sqrt{25} = 5$, $X > 30$, $Z = \frac{30.5 - 25}{5} = 1.10$

$$P(X > 30) = P(Z > 1.10) = 0.5 - 0.3643 = 0.1357 = \%13.57$$



Örnek 5: Bir hastanenin acil servisine günde ortalama 20 hasta gelmektedir. Hastane yöneticileri tamirat sebebiyle acil servisi bir günlüğüne kapatacaklardır. Bu durumda en fazla 15 hastanın zarar görmesi ihtimalini hesaplayınız.

$$\lambda = 20, \sqrt{\lambda} = \sqrt{20} = 4.47, X \leq 15, Z = \frac{15.5 - 20}{4.47} = -1.01$$



$$P(X \leq 15) = P(Z < -1.01) = 0.5 - 0.3438 = 0.1562$$

Örnek 6: Bir bankanın A şubesinde bir günde ortalama 10 adet ihtiyaç kredisi kullanılmaktadır. Tesadüfi olarak seçilen bir günde bu şubece 12 veya daha az ihtiyaç kredisi kullanılması ihtimali nedir?

$\lambda = 10 > 5$ olduğu için, bu durumda poisson dağılımı normale yaklaşmıştır. Böylece ortalama, $\lambda = 10$ ve standart sapma, $\sqrt{\lambda} = \sqrt{10} = 3.16$ olur. 12 veya daha

az ihtiyaç kredisi kullanılması ihtimali sorulduğu için 12 ihtimal bölgesindedir. 12'in üst sınırı 12.5 olduğu için bu noktanın standart değeri,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} = \frac{12.5 - 10}{3.16} = 0.79$$

olur. Bu noktanın solunda kalan alan aranan ihtimaldir. Buna göre 12 veya daha az ihtiyaç kredisi kullanılması ihtimali,

$$P(X \leq 12) = P(Z \leq 0.79) = 0.5 + 0.2852 = 0.7852 \text{ olur.}$$

**Bireysel Etkinlik**

- Süreklilik düzeltmesi kavramını tartışınız. Süreklilik düzeltmesi niçin gereklidir?
- Kesikli değişkenler hangi durumlarda süreklilik düzeltmesine tabi tutulur?
- Kesikli değişken değerine ait ihtimaller hesaplanırken kesikli değişkene hangi durumda 0.5 değeri eklenir, hangi durumda çıkarılır?



Özet

- Bu bölümde, kesikli olasılık dağılımlarının belirli koşullar altında normal dağılıma nasıl yaklaştığı ayrıntılı biçimde ele alınmaktadır. İstatistik biliminin temel konularından biri olan bu yaklaşım, özellikle **binom** dağılımı ve **poisson dağılımı gibi kesikli** dağılımlarının olasılık hesaplamalarında karşılaşılan zorlukları azaltmayı amaçlar. Küçük örnek hacimlerinde binom veya poisson formülleri doğrudan kullanılabilirken, örnek hacmi büyüdükçe bu hesaplamalar karmaşıklaşır. Dolayısıyla, büyük örneklemeler için kesikli dağılımların normal dağılıma yaklaştırılması istatistik hesaplamalarda önemli kolaylık sağlar.
- Normal dağılım, ortalama (μ) ve standart sapma (σ) parametreleriyle tanımlanan simetrik ve sürekli bir dağılımdır. Büyük örnek hacimlerinde kesikli dağılımların frekans eğrileri biçim olarak normal dağılım eğrisine oldukça benzer hale gelir. Ancak normal dağılımda olasılıklar sürekli değerler arasında tanımlandığı için, belirli bir noktanın olasılığı sıfırdır. Bu nedenle, kesikli değişkenlerin belirli bir değeri alma olasılığını hesaplamak için **süreklilik düzeltmesi** uygulanır. Bu düzeltmede değişkenin değerine 0.5 eklenip çıkarılarak bir aralık oluşturulur ve hesaplama bu aralık üzerinden yapılır.
- **Binom dağılımının normale yaklaşımı**, örnek hacminin büyük (genellikle $n \geq 30$) ve başarı olasılığının (p) 0.5'e yakın olduğu durumlarda kullanılabilir. Binom dağılımı iki olasılıklı olaylara (örneğin "başarı" ve "başarısızlık") uygulanır ve deneylerin birbirinden bağımsız olması gerekir. Binom dağılımının ortalaması $\mu = np$, standart sapması ise $\sigma = \sqrt{np(1-p)}$ formülleriyle bulunur. Eğer $np \geq 5$ ve $n(1-p) \geq 5$ koşulları sağlanıyorsa binom dağılımı normale yaklaşır. Bu durumda olasılıklar standart normal değişken (Z) aracılığıyla hesaplanabilir. Ancak p değerinin 0'a veya 1'e aşırı yakın olduğu hallerde dağılım çarpık hale gelir ve normal yaklaşım geçerliliğini kaybeder.
- Uygulamalı örneklerde, binom formülüyle hesaplanan sonuçlar ile normal yaklaşım sonuçlarının birbirine oldukça yakın olduğu görülmektedir. Örneğin, 20 doğumdan 12'sinin kız çocuk olması olasılığı her iki yöntemle hesaplandığında farkın ihmal edilebilir düzeyde olduğu gözlemlenmiştir.
- **Poisson dağılımının normale yaklaşımı** ise nadir olayların analizinde geçerlidir. Poisson dağılımı, binom dağılımının özel bir hali olarak kabul edilir ve düşük olasılıklı ancak çok sayıda tekrar eden olayları temsil eder. Ortalama olay sayısı $\lambda \geq 5$ olduğunda poisson dağılımı normale yakınsar. Bu durumda ortalama $\mu = \lambda = np$, standart sapma $\sigma = \sqrt{\lambda}$ alınır. Poisson olasılıkları da süreklilik düzeltmesiyle birlikte standart normal değişken kullanılarak hesaplanır.
- Örneğin, bir fabrikada üretimin günde ortalama 12 kez durduğu varsayıldığında, üretimin 15 veya daha az kez durma olasılığı normal yaklaşım yardımıyla yaklaşık olarak hesaplanmaktadır. Bu durumda süreklilik düzeltmesiyle X 'in 15,5 ve daha az olma durumu Z değeri dönüşümüyle normal dağılım kullanılarak hesaplanır.
- Sonuç olarak, kesikli dağılımların normale yaklaşımı, istatistikte hem teorik açıklamalarda hem de pratik hesaplamalarda büyük önem taşır. Bu yaklaşım, özellikle binom ve poisson dağılımlarında büyük örnek hacimleriyle çalışırken karmaşık formüllerin yerine normal dağılım tablolarının kullanılmasına olanak tanıyarak hesaplama süreçlerini kolaylaştırır. Böylece, olasılık tahminleri daha hızlı, daha sade ve yeterli doğrulukta elde edilir.

DEĞERLENDİRME SORULARI

(1-6. soruları örneğe bağlı olarak cevaplayınız.)

Örnek 1: Tuğla üretimi yapan bir tesiste üretilen tuğlaların %15'i kusurludur. Kusur oranları dağılımının binoma uygun olduğu farz edilmektedir. Bu tesisten 250 adet tuğla satın alınmıştır.

1. Bu tuğla fabrikasından satın alınan 250 tuğladan ortalama kaç tanesi kusurludur?
 - a) 15.0
 - b) 30.5
 - c) 37.5
 - d) 45.5
 - e) 50.5
2. Bu tuğla fabrikasından 250 birimlik satın almalar yapılırsa kusurlu sayısı dağılımının standart sapması kaç olur?
 - a) 1.32
 - b) 3.52
 - c) 4.33
 - d) 5.65
 - e) 6.68
3. Bu tuğla fabrikasından 250 tuğla satın alındığında 40 veya daha az tuğlanın kusurlu çıkması ihtimali hesaplanırken 40 değeri hangi sürekli değere dönüşür?
 - a) 39.05
 - b) 39.50
 - c) 40.05
 - d) 40.50
 - e) 40.55
4. Bu tuğla fabrikasından 250 tuğla satın alındığında 40 veya daha az tuğlanın kusurlu çıkması ihtimali kaçtır?
 - a) 0.7019
 - b) 0.7514
 - c) 0.8474
 - d) 0.8726
 - e) 0.8957

5. Bu tuğla fabrikasından 250 tuğla satın alındığında 40'tan az tuğlanın kusurlu çıkması ihtimali hesaplanırken 40 değeri hangi sürekli değere dönüşür?
- a) 39.05
 - b) 39.50
 - c) 39.55
 - d) 40.05
 - e) 40.55
6. Bu tuğla fabrikasından 250 tuğla satın alındığında 40'tan az tuğlanın kusurlu çıkması ihtimali kaçtır?
- a) 0.6368
 - b) 0.7002
 - c) 0.7362
 - d) 0.7932
 - e) 0.8366

(7-10. soruları örneğe bağlı olarak cevaplayınız.)

Örnek 2: Bir ihbar hattına günde ortalama 25 adet yanlış ihbar telefonu gelmektedir. Günlük ihbarlar telefonu sayılarının poisson dağılımı gösterdiği farz edilmektedir.

7. Günlük yanlış ihbar sayılarına ait dağılımın standart sapması kaçtır?
- a) 3.5
 - b) 4.0
 - c) 4.5
 - d) 5.0
 - e) 5.5
8. Tesadüfi olarak seçilen bir günde 24'ten az yanlış ihbar gelmiş olması ihtimali hesaplanırken 24 değeri hangi sürekli değere dönüşür?
- a) 23.05
 - b) 23.50
 - c) 23.55
 - d) 24.05
 - e) 24.50
9. Tesadüfi olarak seçilen bir günde 24'ten az yanlış ihbar gelmiş olması ihtimali kaçtır?
- a) 0.2012
 - b) 0.2435
 - c) 0.3012
 - d) 0.3319
 - e) 0.3821

10. Tesadüfi olarak seçilen bir günde 26 veya daha az yanlış ihbar gelmiş olması ihtimali hesaplanırken 26 değeri hangi sürekli değere dönüşür?
- a) 25.05
 - b) 25.50
 - c) 26.55
 - d) 26.05
 - e) 26.50

Cevap Anahtarı

1.c, 2.d, 3.d, 4.a, 5.b, 6.a, 7.d, 8.b, 9.e, 10.e

YARARLANILAN KAYNAKLAR

- [1] Armutlulu, İ. H. (2018). İşletmelerde Uygulamalı İstatistik (3. baskı). İstanbul: Alfa Yayınları.
- [2] Başar, A., Oktay, E. (2018). Uygulamalı İstatistik – I: Kısa Bilgiler ve Çözülmüş Problemler (12. baskı). Erzurum: Kültür ve Eğitim Kitap Kirtasiye.
- [3] Daniel, Wayne W., James C. Terrell (2011). Business Statistics: For Management and Economics (7. baskı). Boston: Houghton Mifflin Company.
- [4] Gürtan, Kenan (1982). İstatistik ve Araştırma Metodları. İstanbul: İstanbul Üniversitesi Yayınları.
- [5] Serper, Ö. (2014). Uygulamalı İstatistik I (7. baskı). İstanbul: Ezgi Kitabevi.
- [6] Yüzer, A. F. vd. (2011). İstatistik. Eskişehir: Anadolu Üniversitesi Açık Öğretim Fakültesi Yayınları.