TANIMLAYICI İSTATİSTİKLERİN HESAPLANMASI

(MERKEZİ EĞİLİM ÖLÇÜLERİ)

TANIMLAYICI İSTATİSTİKLER (DESCRIPTIVE STATISTICS)

Değişken hakkında kolay bilgi edinmeyi sağlayan, verileri özetleyen, örnek ya da toplum ile ilgili olarak değişkenin merkezi eğilimi, yayılımı ve dağılımı gibi belirli özelliklerini yansıtan tipik değerlere belirtici istatistikler (Descriptive Statistics) adı verilir.

TANIMLAYICI İSTATİSTİKLER
(DESCRIPTIVE STATISTICS)

Belirtici istatistikler iki ana gruba ayrılarak incelenirler.

1- Merkezi Eğilim Ölçüleri ya da Yer Ölçüleri
(Central Tendancy Measures, Measures of Location)

2- Dağılım Ölçüleri (Yayılma/Serpilme Ölçüleri)
(Measures of Variation, Measures of Dispersion)

MERKEZİ EĞİLİM ÖLÇÜLERİ

Örnek ya da Toplumda birimlerin, değişkenin gözlenme aralığında (X_{min}-X_{max}) hangi değerler etrafında toplanma eğilimi gösterdiğini belirten ölçülere merkezi eğilim ölçüleri adı verilir.

Merkezi eğilim ölçüleri;

- Ortalama (Mean, Average)
- Ortanca Değer (Medyan, Median)
- Tepe Değeri (Moda Değer, Mode)
- Dörttebirlik Değerler (Çeyrekler, Quartils)
- Yüzdelik Değerler (Percentils)

ORTALAMA

Değişkenin gözlenme aralığında (X_{min}-X_{max}) hangi değer etrafında yığılma gösterdiğini, birimlerin incelenen değişken yönünden gözlenme değerlerinin hangi değere doğru eğilim gösterdiğini belirten ölçüye ortalama adı verilir.

Ortalama, gözlenen değerlerin hesaplamada ele alınış biçimine göre dört değişik biçimde hesaplanır ve hesaplanma tarzına göre farklı isimler alır.

1- Aritmetik Ortalama (Arithmetic Mean, Average)

2- Geometrik Ortalama (Geometric Mean)

3- Harmonik Ortalama (Harmonic Mean)

4- Ağırlıklı Ortalama (Weighted Mean)

ARİTMETİK ORTALAMA (ORTALAMA)

Gözlem değerleri toplamının birim sayısına bölünmesi ile hesaplanan ortalamadır.

Aritmetik Ortalama sadece Ortalama sözcüğü ile de ifade edilir. Değişkenin adına göre (X, Y,...,Z) aşağıdaki gibi gösterilir.

$$\overline{\mathbf{X}}, \quad \overline{\mathbf{Y}}, \quad \overline{\mathbf{Z}}$$

ORTALAMANIN HESAPLANMASI

Ortalama, verilerin durumuna göre iki farklı biçimde ele alınır ve farklı formüller yardımı ile hesaplanır.

a- Sınıflanmamış Veriler

(Ham, Dizi halindeki veriler)

b- Sınıflanmış Veriler

(Tablolaştırılmış, Gruplanmış Dizi halindeki veriler)

SINIFLANMAMIŞ VERİLERDE ORTALAMA

n birimin ölçülen, tartılan X değerleri bir sınıflama yapılmadan verilmiş ise bu değerlerin ortalaması, değerler toplamının birim sayısına bölünmesi ile hesaplanır.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad i = 1,2,\ldots,n$$

SINIFLANMIŞ VERİLERDE ORTALAMA

n birimin ölçülen tartılan X değerleri eşit aralıklı k sınıflı bir tablo biçiminde verilmiş ise, her bir sınıftaki frekanslar (f_i) ile o sınıfın sınıf değeri (x_i, sınıf orta noktası) ile çarpılır ($f_i x_i$).

Tüm sınıflarda belirlenen bu değerlerin toplamı ($\Sigma f_i x_i$) toplam frekansa (Σf_i, n, toplam birim sayısı) bölünür.

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{n} \qquad i = 1,2,\ldots,k$$

SINIFLANMAMIŞ VERİLERDE ORTALAMANIN HESAPLANMASI

Örnek : 10 hastanın Sistolik Kan Basınçları (SKB mm/Hg) aşağıdaki gibi ölçülmüştür.

134 145 127 150 165 154 142 140 130 125

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{134 + 145 + \ldots + 125}{10} = \frac{1412}{10}$$

$$\overline{X} = \frac{1412}{10} = 141.2 \quad mm / Hg$$

SINIFLANMIŞ VERİLERDE ORTALAMANIN HESAPLANMASI

Örnek: **Rasgele seçilen 260 hastanın Sistolik Kan Basınçları (SKB mm/Hg) Frekans Tablosu aşağıdaki gibi verilmiştir.**

SKB	x	f	$f_i x_i$
60-79.99	70	18	1260
80-99.99	90	37	3330
100-119.99	110	79	8690
120-139.99	130	65	8450
140-159.99	150	42	6300
160+	170	19	3230
Toplam	-	260	31260

$$\overline{X} = \frac{\sum f_i x_i}{n} = \frac{31260}{260}$$

$$\overline{X} = 120.23 \text{ mm / Hg}$$

GEOMETRİK ORTALAMA

Dizideki değerlerin artan ya da azalan aritmetik ya da geometrik bir diziyi izlediği durumlarda hesaplanan tipik değerdir.

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^{n} X_i} = \sqrt[n]{X_1 * X_2 * \dots X_n}$$

$$\ln(\overline{X}_G) = \frac{\sum_{i=1}^{n} \ln(X_i)}{n}, \overline{X}_G = e^{(\ln(\overline{X}_G))}$$

$$\overline{X}_G = \mathrm{Exp}\left(\frac{\sum_{i=1}^{n} \ln(X_i)}{n} \right)$$

$$\log_{10}(\overline{X}_G) = \frac{\sum_{i=1}^{n} \log_{10}(X_i)}{n} \qquad \overline{X}_G = 10^{(\log_{10}(\overline{X}_G))}$$

$$\overline{X}_G = \text{Anti}\log\left(\frac{\sum_{i=1}^{n} \log(X_i)}{n}\right)$$

Örnek: D hastanesine 01.01.2015 ile 10.1.2015 tarihleri arasında başvuran hasta sayısı aşağıdaki gibidir.

Günlük başvuran hasta sayısı (birey)

98, 67, 121, 141, 150, 156, 170, 181, 190, 200

$$\ln(\overline{X}_G) = \frac{\ln(98) + \ln(67) + \ldots + \ln(200)}{10}$$

$$= \frac{49.4743}{10} = 4.94743$$

$$\overline{X}_G = e^{(4.94743)} = 140.8$$

$$\cong 141 birey$$

HARMONİK ORTALAMA

Veri setindeki değerler bir zaman serisi ise, eşit şartlarda yapılmamış k sayıda deneyin sonuçlarının bir araya getirilmesi ile elde edilmiş bir veri seti ise ve birbirini izleyen sayılar bir dalgalanma gösteriyorsa (aylık, mevsimsel, yıllık dalgalanmalar) verinin merkezi eğilim ölçüsü harmonik ortalama ile hesaplanır.

HARMONİK ORTALAMANIN HESAPLANMASI

Örnek : Aşağıda 2013-2014 yılları arasında B hastanesi Dahiliye Kliniğinde yatarak tedavi gören hasta sayıları verilmiştir (örnek varsayımsal). On yıllık verilere göre tedavi edilen ortalama hasta sayısını bulalım. Veri seti bir zaman serisidir.

Lösemili hasta sayısı

14, 27, 41, 121, 36, 47, 105, 18, 19, 76

$$\overline{X}_H = \frac{10}{\sum_{i=1}^{n} (1/X_i)}$$

$$= \frac{10}{\frac{1}{14} + \frac{1}{27} + \ldots + \frac{1}{19} + \frac{1}{76}}$$

$$= \frac{10}{0.31148} = 31.15 \text{ hasta / yil}$$

AĞIRLIKLI ORTALAMA

Farklı yer ve zamanda yapılmış araştırmaların verilerinden yararlanılarak daha geniş hacimli bir örnekte araştırma yapılmış gibi merkezi eğilim ölçüsü hesaplamak istendiğinde uygulanan bir ortalama türüdür.

Nicel Verilerde

$$\overline{X}_W = \frac{\sum_{i=1}^{k} n_i \overline{X}_i}{\sum_{i=1}^{k} n_i}$$

Nitel Verilerde

$$P_W = \frac{\sum_{i=1}^{k} n_i p_i}{\sum_{i=1}^{k} n_i}$$

Tedavi Kliniği Adi	Hasta Sayısı n_i	Ortalama mayi miktarı (kg) $\overline{X}_i$	$n_i * \overline{X}_i$
A	88	4.250	374.0
B	112	3.500	392.0
C	56	4.000	224.0
D	12	6.250	75.0
Toplam	268	-	1065.0

$$\overline{X}_W = \frac{\sum_{i=1}^{k} n_i \overline{X}_i}{\sum_{i=1}^{k} n_i} = \frac{1065}{268} = 3.974 \text{ Kg.}$$

ORTANCA DEĞER (MEDYAN)

Küçükten büyüğe sıralanmış bir diziyi iki eşit parçaya ayıran değere ortanca değer denir.

Bir dizide ana eğilimden aşırı derecede küçük ya da büyük değerler varsa ortalamalar bu değerlerden etkilenerek gerçek eğilimi yansıtmaktan uzak kalırlar.

Dizinin sağ ve sol uçta aşırı sapan değerleri çoksa diziyi temsil eden istatistik OD'dir.

Frekans dağılımında diziyi birim sayısı cinsinden iki eşit parçaya ayıran değerdir.

Veri setlerinde ve frekans tablolarında farklı biçimde hesaplanır.

Veri Setlerinde Ortanca Değer

Dizideki değerler küçükten büyüğe dizilir. Dizideki birim sayısının (n) tek ya da çift olmasına göre farklı olarak hesaplanır.

n tek ise $$\mathbf{OD = X_{((n+1)/2)}}$$

n çift ise $$\mathbf{OD = (X_{(n/2)} + X_{((n+2)/2)})\ /\ 2}$$

Örnek:

15 öğrencinin Biyoistatistik puanları aşağıdaki gibidir.

55	100	24	60	67	14	78	86	10	60	50	80	40	90	95
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}

Veri setinin sıralanmış biçimi aşağıdaki gibidir.

10	14	24	40	50	55	60	60	78	80	80	86	90	95	100
$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$	$X_{(11)}$	$X_{(12)}$	$X_{(13)}$	$X_{(14)}$	$X_{(15)}$

OD = $X_{(8)}$ = 60 Puan

Örnek:

12 öğrencinin Boy uzunluğu cm olarak aşağıdaki gibidir.

165, 178, 170, 166, 171, 172, 177, 180, 173, 175, 175, 169

X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12}

Veri setinin sıralanmış biçimi aşağıdaki gibidir.

165, 166, 169, 170, 171, 172, 173, 175, 175, 177, 178, 180

$X_{(1)}$ $X_{(2)}$ $X_{(3)}$ $X_{(4)}$ $X_{(5)}$ $X_{(6)}$ $X_{(7)}$ $X_{(8)}$ $X_{(9)}$ $X_{(10)}$ $X_{(11)}$ $X_{(12)}$

$$OD = [X_{(12/2)} + X_{((12+2)/2)}]/2 = [X_{(6)}+X_{(7)}]/2)$$

$$OD = (172+173) / 2 = 172.5 \text{ cm.}$$

Frekans Tablolarında Ortanca Değerin Hesaplanması

Frekans tablolarında ortanca değer hesaplamak için yığılımlı (Kümülatif) frekans sütunu belirlenir (F(X)).

Yığılımlı frekans değerleri verilerin sıraya dizilmiş biçimidir.

Frekans tablolarında Ortanca, $X_{(n/2)}$ olarak belirlenir.

Frekans tablosunda her sınıftaki frekansın o sınıf içine, SBD ile SÜD değeri arasında, SA'na bağlı olarak eşit olasılıkla dağıldığı varsayılır.

(n/2)'nci değerin bulunduğu sınıf belirlenir, bu sınıfa ortanca sınıfı denir.

$$OD = B + \frac{(n/2) - F_A}{f_B} * SA$$

Burada;

B, Ortanca sınıfın SBD’ni

F_A, Ortanca sınıfından bir önceki sınıfın yığılımlı frekansını

f_B, Ortanca sınıfın frekansını

SA, Sınıf aralığını

belirtmektedir.

Ortanca Değer

Tablo: Rasgele seçilen 246 kişinin boy uzunluğu

Boy Uzunluğu	Frekans (f)	Yığılımlı Frekans (F)
160 - < 165	25	1.-25.
165 - < 170	48	26.-73.
170 - < 175	76	74.-149.
175 - < 180	56	150.-205.
180 - < 185	28	206.-233.
185 - < 190	9	234.-242.
190 +	4	243.-246.
TOPLAM	246	

$B=170$

$F_A=73$

$f_B=76$

$SA=5$

Ortanca=X(n/2)=123

$$OD = B + \frac{(n/2) - F_A}{f_B} * SA$$

$$OD = 170 + \frac{123 - 73}{76} * 5$$

$$OD = 170 + \frac{50}{76} * 5$$

$$OD = 170 + 3.29 = 173.29 \text{ Cm.}$$

Örnek: Rasgele seçilen 466 bireyin SKB değerleri saptanmış ve Tablo'da verilmiştir. Bu tabloda yığılımlı frekanslar ve her sınıfta birimlerin SKB değerlerinin büyüklük sırasına dizilmesi halinde hangi birimlerin (hangi nolu birimlerin?) hangi sınıflarda yer aldıkları gösterilmiştir.

Tablo :466 kişinin SKB değerleri frekans dağılımı

SKB mm/Hg)	Frekans (f)	Yığılımlı Frekans (F)	Her sınıftaki birimlerin sıra noları
70-79.9	15	15	1.-15.
80-89.9	26	41	16.-41.
90-99.9	39	80	42.-80.
100-109.9	62	142	81.-142
*110-119.9	98	240	143.-240.
120-129.9	88	328	241.-328.
130-139.9	65	393	329.-393.
140-149.9	43	436	394.-436.
150+	30	466	437.-466.
Toplam	466	-	-

Tablo’dan;

$n/2 = 466/2 = 233$

olarak hesaplanır. 233’ncü birimin SKB değeri OD olacaktır. 233’ncü birim 143’ncü ve 240’ncı birimlerin yer aldığı 110-119.9 sınıfı içinde bir bireydir ve 98 kişiden birisidir (f_B=98).

Tablo verilerinin Ortancası formülde değerler yerine konarak aşağıdaki gibi hesaplanır.

$$OD = 110 + \frac{233 - 142}{98} * 10$$
$$= 119.29 mm / Hg$$

TEPE DEĞERİ (TD)

Bir seride en sık tekrarlanan değerdir.

Bir seride benzer sayıda tekrarlanan birden fazla değer varsa o seriye çok tepeli seri denir.

Tablolaştırılmış verilerde Tepe değeri en çok frekansa sahip sınıfta yer alır.

Bu sınıfa Tepe sınıfı adı verilir.

Sınıflanmamış Dizilerde Tepe Değeri

Serideki değerlerin tekrarlama sayıları (frekanslar) belirlenir.

En yüksek frekansa sahip gözlenen değer Tepe Değeri (TD) olarak belirlenir.

Örnek : 12 Bireyin boy uzunlukları aşağıdaki gibidir.

165, 178, 170, 166, 171, 172, 177, 180, 173, 175, 175, 169

165 (1) 166 (1) 169 (1) 170 (1) 171 (1) 172 (1) 173 (1) 175 (2)

177 (1) 178 (1) 180 (1)

Frekans Tablolarında Tepe Değeri

Frekans tablolarında tepe değeri (TD), veri setinin Ortalama ve Ortanca değerinden yararlanılarak aşağıdaki gibi hesaplanır.

$$\mathbf{TD} = \overline{\mathbf{X}} - 3(\overline{\mathbf{X}} - \mathbf{OD})$$

Örnek: Hesaplanmış veri setinin ortalaması 170.6 cm Ortancası OD=170.5 cm. olarak hesaplanmış olsun. TD=?

TD = 170.6 - 3(170.6 - 170.5) = 170.3 cm. olarak hesaplanır.

Dörttebirlikler

Büyüklük sırasına dizilmiş bir veri setini dört eşit parçaya bölen istatistiklere dörttebirlik denir.

Dörttebirliklerin hesaplanması ortanca değer hesaplanmasına benzer.

Dörttebirlikler

Birinci dörttebirlik Q1 ile gösterilir ve veri setinin ilk çeyreğini yani %25. değerini belirtir. $Q1= X_{[n/4]}$

İkinci dörttebirlik Q2 ile gösterilir yani bu aynı zamanda ortanca değerdir ve veri setinin %50. değerini gösterir.

Üçüncü dörttebirilik Q3 ile gösterilir ve veri setinin %75. değerini belirtir. $Q3= X_{[3n/4]}$

Yüzdebirlikler

Büyüklük sırasına dizilmiş bir veri setini yüzdelik bölümlere ayıran istatistiklerdir. P(%) olarak gösterilirler.

Örneğin P(5) veri setinin %5. değerini gösterir.

Burada P(25)=Q1,

P(50)=Ortanca Değer=Q2

P(75)=Q3.

değerini gösterir. Örneğin $P(5)=X_{[5n/100]}$, $P(90)= X_{[90n/100]}$, $P(80)= X_{[80n/100]}$ olarak hesaplanır.

DAĞILIM ÖLÇÜLERİ

Bir değişkenin dağılımını, değerler aralığındaki serpilmesini ve ortalama etrafında yayılışlarını, belirli değerlerde yığılma eğilimlerini belirlemeye yarayan tanımlayıcı istatistiklere dağılım ölçüleri adı verilir.

Dağılım ölçüleri, merkezi eğilim ölçülerini destekleyen ve verilerin merkezi eğilim ölçüleri etrafında yayılışlarını, dağılımlarını ve serpilmelerini gösteren ölçülerdir.

DAĞILIM ARALIĞI

Dizideki en büyük ve en küçük değer arasındaki farka Dağılım Aralığı denir ve R ile gösterilir.

$$R = X_{max} - X_{min}$$

ya da

$$R = X(n) - X(1)$$

şeklinde hesaplanır.

Dizideki değerlerin kabaca kaç birimlik yayılış gösterdiğini belirtir.

DAĞILIM ARALIĞI

Örnek: 15 bireye ait hemoglobin değerleri;

11, 12, 14, 11, 15, 16, 15, 12, 11, 10, 9, 11, 13, 14, 15

ise bu veri setinin dağılım aralığı,

$$R = X_{max} – X_{min}$$

$$R = 16 – 9$$

$$R = 7$$

olarak hesaplanır.

VARYANS

Dizideki verileri ortalama etrafında nasıl bir dağılım gösterdiğini, yayılış ve serpilmenin durumunu değişkenin ölçü biriminin karesi olarak belirten dağılım ölçüdür.

S^2 ya da değişken adına göre V(X), V(Y)... ile gösterilir.

$$V(X) = S^2 = \frac{\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i \right)^2}{n} \right)}{(n-1)}$$

STANDART SAPMA

Dizideki verilerin ortalama etrafında nasıl bir dağılım gösterdiğini, yayılış ve serpilmenin durumunu belirtmekte kullanılan istatistiktir.

Ölçü birimi değişkenin ölçü birimi ile aynıdır.

S harfi ile gösterilir.

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

STANDART SAPMA

Örnek: 40 yaşında 165 cm. boyunda 10 sağlıklı kadına ait FEV1 (litre) değerleri ölçülmüş ve

2.82 2.78 2.88 2.70 2.91 2.76 2.80 2.86 2.80 2.87

değerleri elde edilmiştir. FEV1 ölçümlerine ait standart sapma değeri;

$$S = \sqrt{\frac{2.82^2 + 2.78^2 + \cdots + 2.87^2 - \frac{(2.82 + 2.78 + \cdots + 2.87)^2}{10}}{10 - 1}} = 0.063$$

olarak hesaplanır.

STANDART HATA

n birimlik örnekten elde edilen ortalamanın, toplum ortalaması μ'yü tahminde ne kadar tutarlı olduğunu, tahminin ne kadar hata taşıdığını belirtmekte yararlanılan bir dağılım ölçüsüdür.

Standart Hata;

$$SH = \frac{S}{\sqrt{n}}$$

olarak hesaplanır. S

Standart sapmasının örnek hacminin kare köküne bölünmesi ile elde edilir. SH ya da S ile gösterilir.

STANDART HATA

Örnek: 40 yaşında 165 cm. boyunda 10 sağlıklı kadına ait FEV1 (litre) değerleri ölçülmüş ve

2.82 2.78 2.88 2.70 2.91 2.76 2.80 2.86 2.80 2.87

değerleri elde edilmiştir. FEV1 ölçümlerine ait standart hata değeri;

$$S=\sqrt{\frac{2.82^2+2.78^2+\cdots+2.87^2-\frac{(2.82+2.78+\cdots+2.87)^2}{10}}{10-1}}=0.063$$

$$SH=\frac{0.063}{\sqrt{10}}=0.02$$

olarak hesaplanır.

STANDART SAPMA MI STANDART HATA MI?

Standart sapma, bir dizide değerlerin dağılımı hakkında bilgi verirken; **Standart hata**, n hacimli örneklerden elde edilebilecek ortalamaların toplumda dağılımı hakkında bilgi verir.

Bilimsel yayınlarda daha çok topluma ilişkin genellemelere yer verilmesi gerektiğinden örnek ortalamasının, standart hatası ile birlikte verilmesi gerekir.

DEĞİŞİM KATSAYISI

Birim sayıları ve ölçü birimleri birbirlerinden farklı olan değişkenlerin ortalamaya göre yayılışlarını karşılaştırmak için yararlanılan ve değişkenin ortalama ve standart sapmasından yararlanılarak hesaplanan bir dağılım ölçüsüdür.

DK% ile gösterilir ve

$$DK\,\% = \frac{S}{\overline{\overline{X}}} \times 100$$

olarak hesaplanır.

DEĞİŞİM KATSAYISI

Örnek: 15 bireye ait hemoglobin değerleri

11, 12, 14, 11, 15, 16, 15, 12, 11, 10, 9, 11, 13, 14, 15 olarak belirlenmiştir.

40 yaşında 165 cm. boyunda 10 sağlıklı kadına ait FEV1 (litre) değerleri ölçülmüş ve

2.82 2.78 2.88 2.70 2.91 2.76 2.80 2.86 2.80 2.87 değerleri elde edilmiştir.

Acaba bu iki değişkeninden hangisi daha fazla dağılım, serpilme göstermektedir?

DEĞİŞİM KATSAYISI

Bu soruyu cevaplamak için her iki değişkene ait değişim katsayısı hesaplanmalıdır.

Eğer değişkenlerin standart sapmalarını doğrudan karşılaştıracak olursak hata yapmış oluruz çünkü değişkenlerin ölçü birimleri birbirinden farklıdır.

Bu durumda her iki değişkene ait değişim katsayısını hesaplamak gerekir.

DEĞİŞİM KATSAYISI

FEV1 değişkenine ait değişim katsayısı

DK(FEV1) = (0.063/2.818) x 100 = % 2.235

Hemoglobin değişkenine ait değişim katsayısı

DK(HG) = (2.131/12.6) x 100 = %16.9

olarak hesaplanır.

Bu sonuçlara göre Hemoglobin değişkeni FEV1 değişkenine göre daha fazla dağılım, serpilme göstermektedir.