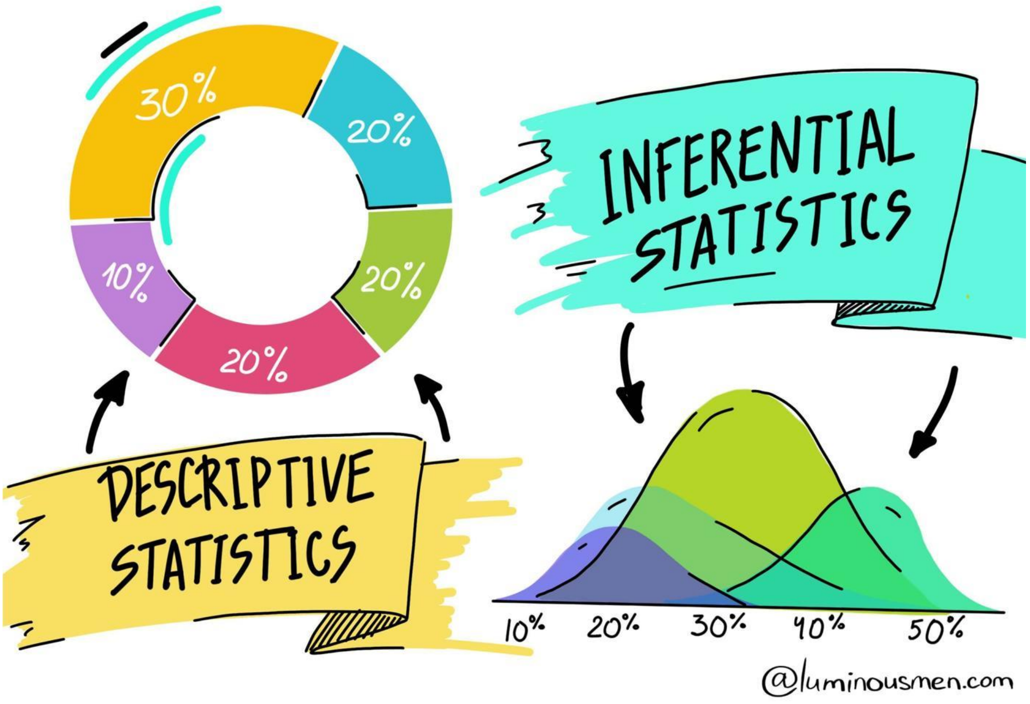
Tanımlayıcı İstatistikler



Hazırlayan: Dr. Gökhan AKSU

Tanımlayıcı İstatistikler

- Bir veri setini tanımak veya birden fazla veri setini karşılaştırmak için kullanılan ve ayrıca örnek verilerinden hareket ile frekans dağılışlarını sayısal olarak özetleyen değerlere **tanımlayıcı istatistikler** denir**.**

- Analizlerde kullanılan veri tiplerine (*basit*, *gruplanmış, sınıflanmış*) göre hesaplamalarda kullanılacak formüller değişmektedir.

Tanımlayıcı İstatistikler

Merkezi Eğilim

1)Aritmetik ort.
2)Geometrik ort.
3)Harmonik ort.
4)Mod
5)Medyan

Değişkenlik Ölçüleri

1) Range
   (Değişim Aralığı)
2) Ort. Mutlak sapma
3) Varyans
4) Standart Sapma
5) **Değişkenlik(Varyasyon) Katsayısı**
6)Kartiller (çeyrekler/yüzdelikler)

Çarpıklık Ölçüleri

Basıklık Ölçüleri

❖ Merkezi Eğilim Ölçüsü

Veri setinin orta noktası veya merkezinin değeridir.

1) Aritmetik Ortalama

- Üzerinde inceleme yapılan veri setindeki elemanların toplanıp incelenen eleman sayısına bölünmesiyle elde edilen yer ölçüsüne aritmetik ortalama denir.

- Örnek:
  - Sınav notlarının ortalaması,
  - Yaz aylarında m^2'ye düşen ortalama yağış miktarı

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

Örnek Ortalaması ve Anakütle Ortalaması

$\overline{x}$, x-bar şeklinde telaffuz edilir ve örneklemin ortalamasıdır.

$$\overline{x} = \frac{\sum x}{n}$$

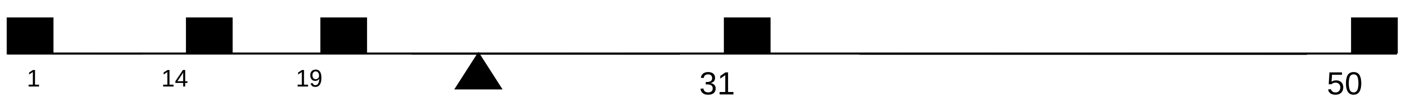
μ, “mü” şeklinde telaffuz edilir ve anakütle ortalamasıdır

$$\mu = \frac{\sum x}{N}$$

μ

Bir Denge Noktası Olarak Ortalama

- 1, 14, 19, 31, 50 sayılarının ortalaması μ=23 tür. Şekil sayıları bir çizgi üzerinde yerleştirilmiş eşit küçük ağırlıklar şeklinde gösterir.**1,14,19,31,50**
- Aritmetik ortalama denge noktasıdır.



Basit Veriler için Aritmetik Ortalama Örneği

Örnek: İzmir ilinde ilköğretim ikinci sınıfta okuyan öğrenciler üzerinde yapılan bir araştırmada rasgele 8 öğrenci seçilmiş ve ailenizde kaç çocuk vardır sorusuna aşağıdaki gibi cevap vermişlerdir.

Buna göre Ailelerin çocuk sayılarının ortalamasını hesaplayınız.

1,3,2,1,4,5,6,2

n = 8 i = 1,2,…,8

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+1+2+2+3+4+5+6}{8} = 3$$

Gruplanmış Veriler İçin Aritmetik Ortalama

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \, f_i}{\sum_{i=1}^{k} f_i}$$

$$\sum_{i=1}^{k} f_i = n$$

f : frekans

k: grup sayısı

i = 1,2,3,……….,k

Örnek: Bir otomobil bayisinde 80 gün boyunca yapılan inceleme sonucunda satılan arabaların adetlerine göre dağılımı yandaki tabloda verilmiştir.

Buna göre bir gün içinde satılan ortalama araba sayısını hesaplayınız.

Araba (x_i)	GÜN(f_i)	$x_i.f_i$
0	5	***0***
1	12	***12***
2	35	***70***
3	14	***42***
4	8	***32***
5	6	***30***
	$\sum f_i=80$	

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{0+12+70+42+32+30}{80} = \frac{186}{80} = 2,33$$

- Ör: Öğrencilerin okuma hızları ortalaması kaçtır?

Ölçüm (x)	**Frekans (f)**
7	2
6	3
5	5
4	7
3	4
2	3
1	1

Sınıflanmış Veriler İçin Aritmetik Ortalama

$$\overline{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i}$$

f : frekans

k : sınıf sayısı

i = 1,2,3,………,k

m : sınıf orta noktası

$$\sum_{i=1}^{k} f_i = n$$

- Sınıflanmış verilerde her bir sınıf içindeki değerlerin neler olduğu bilinmediğinden dolayı ve yalnızca her bir sınıfın frekans değerleri bilindiğinden dolayı sınıfı temsil etmek üzere sınıf orta noktaları hesaplamada kullanılır.

- Kullanılan formül gruplanmış veriler için kullanılan formüle benzerdir.

Örnek: Bir sınıftaki öğrencilerin boyları hakkında bir araştırma yapılmaktadır. Bu amaçla 50 öğrencinin boyları ölçülerek kaydedilmiştir.

Öğrencilerin boylarının aritmetik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$m_i f_i$
150-157'den az	5		
157-164'den az	7		
164-171'den az	14		
171-178'den az	9		
178-185'den az	8		
185-192'den az	4		
192-199'dan az	3		
Toplam	50		

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i} = \frac{153,5(5) + 160,5(7) + \ldots + 195,5(3)}{50} = \frac{8599}{50} = 171,98\ cm.$$

Ağırlıklı Ortalama

Veri setindeki gözlemlerin belirli bir kritere göre ağırlıklandırılması durumunda veri setinin ortalamasının hesaplanması için kullanılan ortalamadır.

$$\overline{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

Örnek: Aşağıdaki tabloda şipariş büyüklüklerine göre elde edilen kar miktarları ve sipariş sayıları verilmiştir.

Buna göre bir siparişden elde edilecek ortalama kar miktarı kaç $'dır?

Sipariş büyüklüğü	**Sipariş başına kar x_i**	**Sipariş sayısı w_i**	**$x_i w_i$**
Küçük (1-3)	$1	120	
Orta (4-6)	$3	60	
Büyük (7-üstü)	$6	20	
		$\Sigma w_i =$	$\Sigma x_i w_i =$

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{420}{200} = \$2{,}1$$

2) Geometrik Ortalama

- Bir veri setinde bulunan n adet elemanın çarpımının n nci dereceden kökünün alınmasıyla elde edilen yer ölçüsüdür.

$$G = \sqrt[n]{x_1 \times x_2 \times .... \times x_n}$$

- Geometrik ortalamanın formülüne bakıldığında hesaplama zorluğu olduğundan dolayı logaritma ifadesi kullanılır. Genellikle basit veriler için kullanışlı olup negatif sayılar için kullanışlı değildir.

$$Log\ G = \frac{\sum_{i=1}^{n} \log x_i}{n} \qquad G = anti\log \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

Geometrik Ortalama’nın Kullanım Alanları

- Ortalama oranları,
- Değişim Oranları,
- Logaritmik dağılış gösteren veri setleri,

için kullanışlıdır.

Örnek: fiyat indeksleri, faiz formülleri.

Örnek: Abac şirketinin yıldan-yıla olan fuel deki tüketim harcamalarının değişimi yüzde -5, 10, 20, 40, ve 60. büyüme faktörlerinin geometrik ortalamasını kullanarak harcamalardaki ortalama yıllık yüzde değişim belirlenir. Büyüme faktörleri için yüzde değişim dönüştürme ile elde edilenler;

0.95 1.10 1.20 1.40 1.60

$$G = \sqrt[n]{x_1 \times x_2 \times \ldots . \times x_n} = \sqrt[5]{(0,95)(1,10)(1,20)(1,40)(1,60)}$$

$$= \sqrt[5]{2.80896} \approx 1,229$$

$$Log\ G = \frac{\sum_{i=1}^{n} \log x_i}{n} = \frac{-0,022276 + 0,041393 + 0,079181 + 0,146128 + 0,204120}{5}$$

$$Log\ G = \frac{0,448546}{5} \approx 0,08971$$

G = anti log 0,27045 = 10^{0,08971} ≈ 1,229 ≈ 1,23

3) Harmonik Ortalama

- Bir veri setinde bulunan n adet elemanın çarpma işlemine göre terslerinin ortalamasının tersinin alınmasıyla elde edilen yer ölçüsüdür. Genellikle basit veriler için kullanışlıdır.

$$H = \frac{1}{\left(\dfrac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n} \right)} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

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

Harmonik Ortalama'nın Kullanım Alanları

Zaman verileri için kullanışlıdır.

Örnek: Zaman birimi başına hız, para birimi başına satın alınan birim sayısı.

Belirli koşullar ve fiyat tipleri için zaman verilerinin ortalamalarının hesaplanmasında kullanılan bir yer ölçüsüdür.

Zamana bağlı hız, fiyat verimlilik gibi oransal olarak ifade edilebilen verilerin ortalamasın alınmasında da kullanılabilir.

NOT: ARİTMETİK ORT. > GEOMETRİK ORT. > HARMONİK ORT.

Örnek: Bir tekstil fabrikasında çalışan dört kişinin bir pantolonu ütüleme süreleri aşağıda verilmiştir. Buna göre bu fabrikada bir pantolon ortalama kaç dakikada ütülenir?

İşçi 1: 10 dk. İşçi 2: 6 dk. İşçi 3: 4 dk. İşçi 4 : 5 dk.

$$\frac{1}{H} = \frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n} = \frac{\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{10}}{4} = \frac{43}{240}$$

$$H = \frac{240}{43} \approx 5{,}58\,dk.$$

4) Mod

- Bir veri setinde en çok gözlenen (en çok tekrar eden) değere veya frekansı en fazla olan şans değişkeni değerine **mod** adı verilir.

- Veri setinin *modu* olmayacağı gibi birden fazla da *modu* olabilir.

- Mod genellikle kesikli şans değişkenleri için oluşturulan gruplanmış verilerde aritmetik ortalama yerine kullanılabilir.

Mod

- Mod, büyük veri setlerinde verinin daha çok nerede toplandığını bulmak için kullanılır. Örneğin erkek kıyafetleri satan bir perakendeci, potansiyel müşterilerini belirlemek için gömlek kol uzunluğu ve gömlek yaka ölçüsüyle ilgilenebilir.
- Nicel veri seti çok büyük olmadığı zaman mod anlamlı olmayabilir.
- Sınıflandırılmış veriler için kullanılabilecek tek merkezi eğilim ölçüsüdür.

Örnekler

1) 5,40 1,10 0,42 0,73 0,48 1,10 ⇦ Modu 1,10

2) 27 27 27 55 55 55 88 88 99 ⇦ 1 den fazla moda sahip , 27 ve 55

3) 1 2 3 6 7 8 9 10 ⇦ Modu yok

Gruplanmış Veriler İçin Mod

Basit verilerde bulunduğu gibi hesaplanır.

Örnek: Bir otomobil bayisinde 80 gün boyunca yapılan inceleme sonucunda satılan arabaların adetlerine göre dağılımı yandaki tabloda verilmiştir. Buna göre araba satışları için mod değeri nedir?

Araba(x_i)	GÜN(f_i)
0	5
1	12
2	35
3	14
4	8
5	6

En yüksek frekansa sahip olan gözlem değeri 2 olduğundan dolayı araba satışları için mod değeri 2'dir.

Sınıflanmış Veriler İçin Mod

• Sınıflanmış verilerde mod değeri hesaplanırken ilk olarak mod sınıfı belirlenir.

• Mod sınıfı frekansı en yüksek olan sınıftır.

• Mod sınıfı belirlendikten sonra bu sınıf içerisinde yer alan modun tam değeri sınıf frekansı ve kendine komşu olan sınıf frekansları dikkate alınarak hesaplanır.

$$\textbf{Mod} = L_{\text{mod}} + \frac{\Delta_1}{\Delta_1 + \Delta_2} . i$$

L_{Mod} = Mod Sınıfı Aralığının Alt Sınırı

Δ_1 = Mod Sınıfı Frekansı - Kendinden Bir Önceki Sınıf Frekansı

Δ_2 = Mod Sınıfı Frekansı – Kendinden Bir Sonraki Sınıf Frekansı

i = Mod Sınıfının Sınıf Aralığı

Örnek: Bir sınıftaki öğrencilerin boyları hakkında bir araştırma yapılmaktadır. Bu amaçla 50 öğrencinin boyları ölçülerek kaydedilmiştir.Öğrencilerin boylarının mod değerini hesaplayınız.

Sınıflar	f_i
150-157'den az	5
157-164'den az	7
164-171'den az	14
171-178'den az	9
178-185'den az	8
185-192'den az	4
192-199'dan az	3
Toplam	50

Mod sınıfı

$$L_{\text{mod}} + \frac{\Delta_1}{\Delta_1 + \Delta_2}.i$$

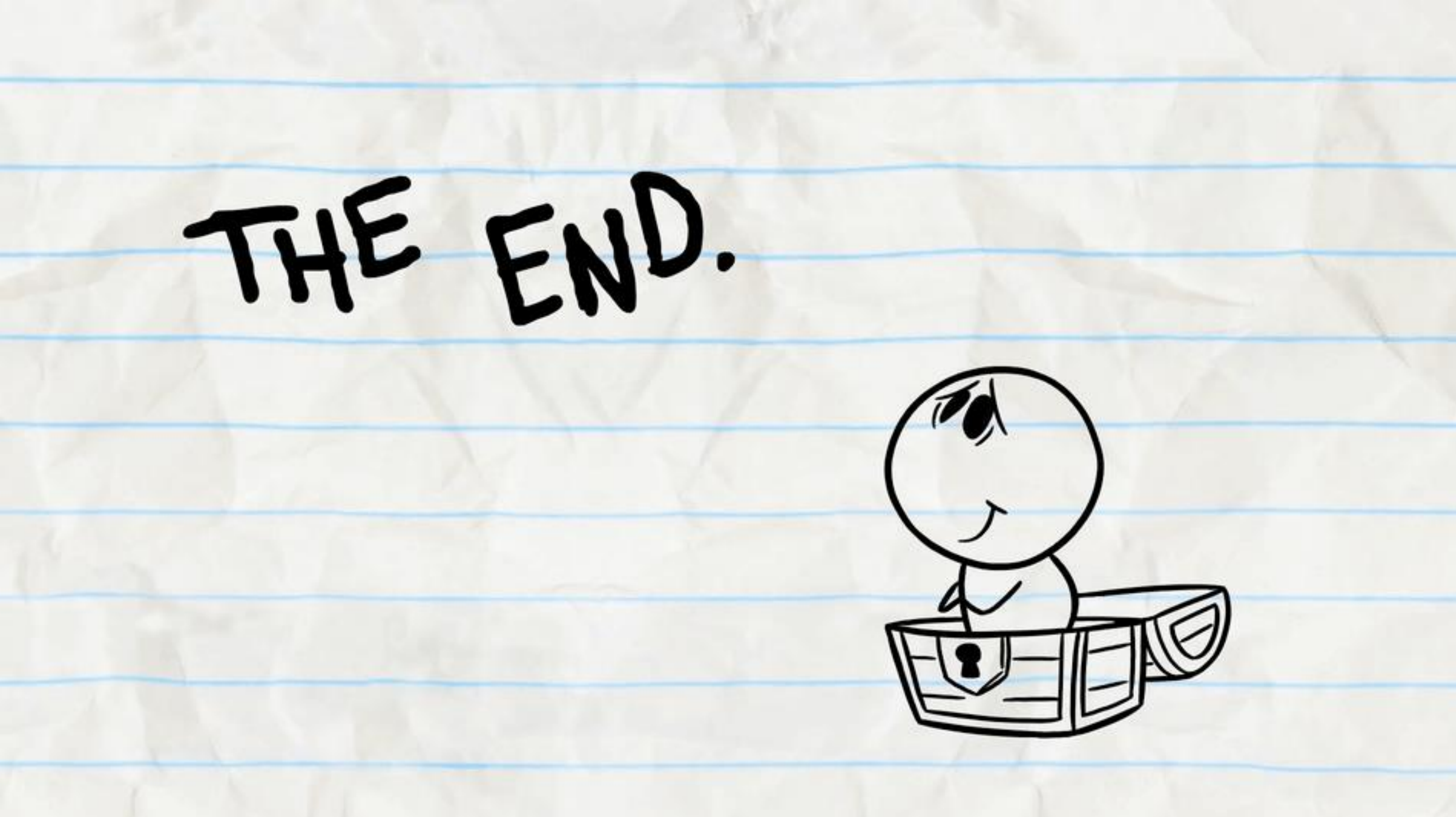
164+(7/12)*7

Frekansı en yüksek olan sınıf mod sınıfı olarak belirlenir.

Mod sınıfı belirlendikten sonra formülde ilgili değerler yerine koyularak **mod** değeri hesaplanır.

$$Mod = L_{\mathrm{mod}} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$$= 164 + \frac{(14-7)}{(14-7)+(14-9)} \times 7 = 168,08\ cm.$$



Dinlediğiniz İçin Teşekkür Ederim…

Dr. Gökhan AKSU