İSTATİSTİK VE İHTİMALLER TEORİSİ

VERİ DÜZENLENMESİ VE ANALİZİ

Dr. Öğr. Üyesi Fatma Zehra Göğüş

GİRİŞ VE ÖN BİLGİLER

Veri: Bir araştırmacı tarafından gözlemlerden elde edilen sayısal olan ya da olmayan sonuçlara bilimsel çalışmalarda **veri** denir.

Veriler toplandıktan sonra, istatistiksel çalışmadaki sonraki adım anlamlı yollarla veriyi organize etmektir. Bunun için tablolar, grafikler ya da çizgileri kullanırız.

Sayısal (Nicel) Veri: Ağırlık, uzaklık, zaman veya bireylerin veya kusurların sayısı gibi değerlerden oluşan nümerik ölçümlerdir. Sayımları ya da ölçümleri sayılarla ifade edilebilen verilerdir.

Örnek: Tüketilen su miktarı, kişilerin ağırlığı ve boyu vb.

Kategorik (Nitel) Veri: Sayısal olarak ifade edilemeyen, karakteristiklere göre birbirinden farklı kategorilere ayrılmış verilerdir.

Örnek: Cinsiyet, Renk, vb.

GİRİŞ VE ÖN BİLGİLER

Birimlerin farklı değerler alabildikleri yani gözlemden gözleme farklılık gösteren niteliksel veya niceliksel özelliklere **değişken** denir.

Değişkenler aşağıdaki gibi sınıflandırılır.

1. **Ölçme ve sayma değişkenleri: (Nümerik Değişkenler)**
   a) Sürekli Değişkenler: Tanımlı olduğu aralıkta tüm değerleri (sonsuz sayıda) alabilen değişkenlerdir. Örnek: Ağırlık, yükseklik vb.
   b) Kesikli Değişkenler: Tanımlı olduğu aralıkta sadece tam sayı değerleri alabilen değişkenlerdir. Örnek : Banka şubesinde gün içerisinde vadeli hesap açtıran müşteri sayısı.

2. **Sıralama Değişkenleri:** Bazı değişkenlere ait değerler ölçümle belirlenmemekle beraber büyüklükleri itibariyle sıralanabilirler (1., 2., …, n. vs sıralama işlemi yapılır)

3. **Özellik Belirten Değişkenler:** Özellikle kalite ve durum gibi nitel özellik belirten değişkenler bu sınıfa koyulabilir. Örneğin siyah-beyaz, ölü-canlı, erkek-dişi, evli-bekar vb.

FREKANS (SIKLIK) DAĞILIMI

Toplanan veri üzerinde herhangi bir işlem yapılmamışsa, bunlara "**ham veri**" ya da " **sınıflandırılmamış (gruplandırılmamış)**" veri denir.

- Birim sayısı az olan yığınların çeşitli özellikleri ham veriye dayanarak kolaylıkla belirtilir.
- Yığın çok sayıda birimden oluşuyorsa, bunları sınıflandırmakla yığının çeşitli özelliklerini belirlemek kolaylaşacaktır.

Sınıflandırmanın en doğru yolu **frekans tablosudur**. Burada, gözlenen veriler sınıflara ayrılır. Sonuçlanan tablo her bir sınıftaki gözlem sayısını verecektir.

Bir değerin gözlenme sıklığına, tekrar sayısına **Frekans (frequency)** denir.

FREKANS (SIKLIK) DAĞILIMI

Frekans tablosunu hazırlamak için aşağıdaki adımlar izlenir.

1.Adım: Gözlemlerin sayısı belirlerin. n = Gözlem sayısını gösterir ($50 \leq n$ tercih edilir).

2.Adım: En büyük değer (maksimum) ve en küçük değer (minimum) bulunur.

Aradaki fark hesaplandığında ise *değişim aralığı (range)* elde edilir.

$$\boxed{range\,(R) = max - min}$$

3.Adım: *Sınıf sayısı* bulunur.

$$\boxed{\begin{array}{c} \sqrt{n} \leq k,\ (k = tam\ sayı) \\ sınıf\ sayısı = k \end{array}}$$

Kesin bir yaklaşım olmamakla birlikte sınıf sayısı için k = 1 + 3.3 * log(n) formülü de kullanılabilir.
Genel olarak $5 \leq k \leq 20$ seçilmesi beklenir.

4.Adım: **Sınıf genişliği:** Ardışık iki sınıfın alt ya da üst sınıf limitleri arasındaki farka dağılım için sınıf genişliği denir.

Başka bir sınıflamaya gerek duyulmadıkça sınıf genişliklerini eşit alınır.

Sınıf genişliği h olmak üzere, $range$ = R, Sınıf sayısı = k ise

$$\boxed{\frac{R}{k} = \frac{max - min}{k} \leq h}$$

FREKANS (SIKLIK) DAĞILIMI

Frekans tablosunu hazırlamak için aşağıdaki adımlar izlenir.

5.Adım: *Sınıf limitleri*, frekans dağılımında sınıfları belirlemek için kullanılan sayılardır. En küçük gözlem değerine eşit ya da daha küçük olarak ilk sınıfın alt limiti seçilir. Bu değere ardışık olarak sınıf genişliğinin eklenmesiyle diğer sınıfların alt limitleri bulunur. Verinin sürekli ya da kesikli olmasına göre ilk sınıftan başlayarak üst limitler de aynı yolla bulunur.

6.Adım: *Sınıf sınırları* belirlenir.

$$\boxed{\frac{i.\ \text{sınıfın üst limiti} + \text{bir sonraki sınıfın alt limiti}}{2} = i.\ \text{sınıfın üst sınıf sınırı}}$$

[i. sınıfın üst sınıf sınırı = (i+1). sınıfın alt sınıf sınırıdır]. Sınıf sınırları arasındaki (üst sınır − alt sınır) fark bir sınıfın sınıf genişliğini verir.

7.Adım: *Sınıf frekansı*, her bir sınıf için -sınıf limitleri dahil -o sınıfa düşen gözlem sayısı (sınıf frekansı) saptanır.

8.Adım: *Sınıf orta noktası*, sınıf limitlerinin ya da sınıf sınırlarının ortalaması alınarak her sınıf için sınıf orta noktaları bulunur.

9.Adım: Frekanslar toplanarak kontrol edilir.

10.Adım: *Kümülatif frekans*, ardışık olarak frekanslar toplanarak kümülatif (eklemeli) frekans sütunu oluşturulur.

FREKANS (SIKLIK) DAĞILIMI

ÖRNEK: Büyük bir şirketin 100 satış elemanı vardır. Her bir satış elamanı tarafından gerçekleştirilen aylık satışlar aşağıda verilmiştir. Bu veriler için frekans tablosunu düzenleyiniz.

1. adım: Gözlem sayısı (n)= 100 ‘dür.

2. adım: En büyük değer: 29,

En küçük değer: 0, R= 29-0=29 dur.

3. adım: $\sqrt{n} \leq k, \sqrt{100} \leq k'dan$

$$k = 10 \; alınabilir.$$

4. adım: $\frac{R}{k} \leq h \,, \frac{29}{10} \leq h, \;\; 2.9 \leq h$

h’ı $tam\ almak\ için\ h = 3\ diyebiliriz.$

5., 6., 7., 8., 9.,10. adımlarda sırasıyla uygulanır

ve frekans tablosu elde edilir.

(Milyon Lira)				
23	16	14	20	27
19	17	17	16	17
26	14	9	11	14
11	17	13	19	17
20	17	20	16	16
11	24	21	27	5
17	20	8	16	17
16	16	14	22	13
14	27	19	16	20
16	15	9	17	8
19	14	8	19	27
22	21	0	9	3
20	14	6	11	12
7	20	9	13	20
10	16	10	19	13
15	15	14	13	25
14	9	16	8	16
7	8	13	5	13
9	16	19	14	29
18	14	18	13	10

FREKANS (SIKLIK) DAĞILIMI

Frekans (Sıklık) Tablosu

Sınıf No	Sınıf Limitleri	Sınıf Sınırları	Gözlem Sayısı (f_i)	Sınıf Orta Noktası (y_i)	Eklemeli frekanslar
1	0-2	-0,5 ~ 2,5	1	1	1
2	3-5	2,5 ~ 5,5	3	4	4
3	6-8	5,5 ~ 8,5	8	7	12
4	9-11	8,5 ~ 11,5	13	10	25
5	12-14	11,5 ~ 14,5	20	13	45
6	15-17	14,5 ~ 17,5	25	16	70
7	18-20	17,5 ~ 20,5	17	19	87
8	21-23	20,5 ~ 23,5	5	22	92
9	24-26	23,5 ~ 26,5	3	25	95
10	27-29	26,5 ~ 29,5	5	28	100

+

100

GRAFİKSEL GÖSTERİMLER

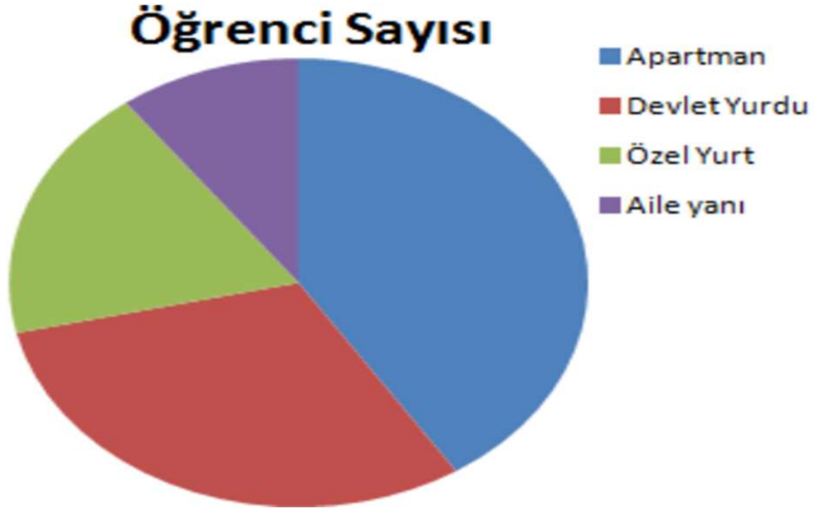
Grafikler Nasıl Olmalıdır?

- Verinin orijinali olmadan tek başına bir şeyler anlatabilmelidir.
- Başlığı ve etiketleri olmalıdır.
- İhtiyaç olduğunda açıklama bölümü konmalı ve verinin kaynağı tarihle birlikte belirtilmelidir.

➢ PASTA GRAFİKLERİ

- Her bir kategorinin bütün içindeki büyüklüğünü gösterir.
- Verinin kategorileri bütünün parçalarını oluşturduğunda kullanılır.

Barınma Türü	Öğrenci Sayısı
Apartman	20
Devlet Yurdu	15
Özel Yurt	9
Aile yanı	5
Total	49

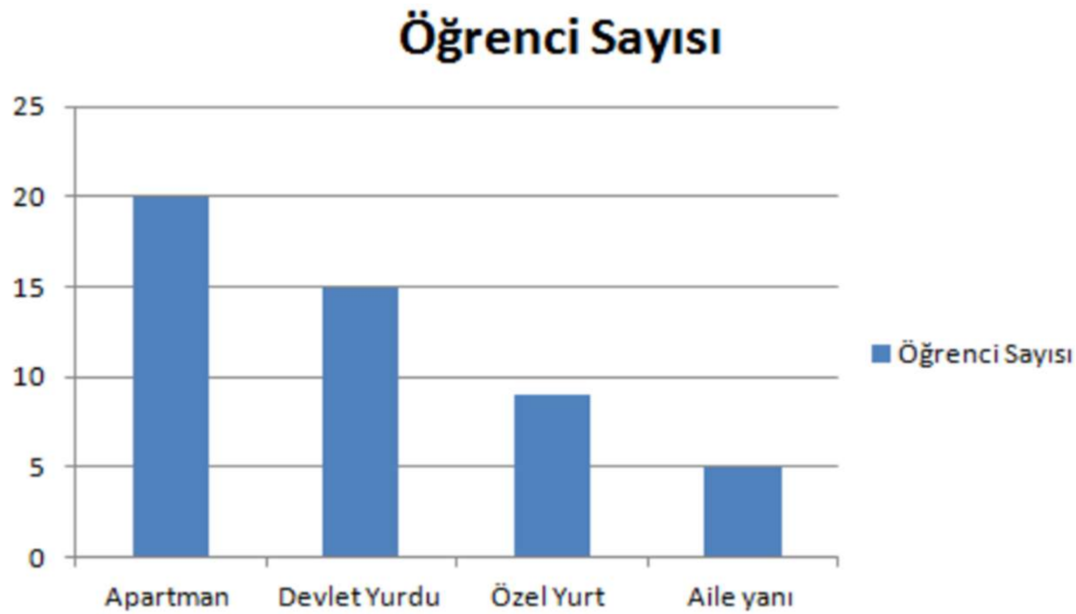


GRAFİKSEL GÖSTERİMLER

➢ BAR GRAFİKLERİ

- Barın yüksekliği ilgili kategorideki veri miktarını göstermektedir.
- Yatay eksen nitel kategorileri içermektedir.
- Dikey eksen her bir kategorinin frekansını göstermektedir.
- Genişliklerin bir anlamı yoktur.
- Barların yanlarındaki barlarla bir teması yoktur.

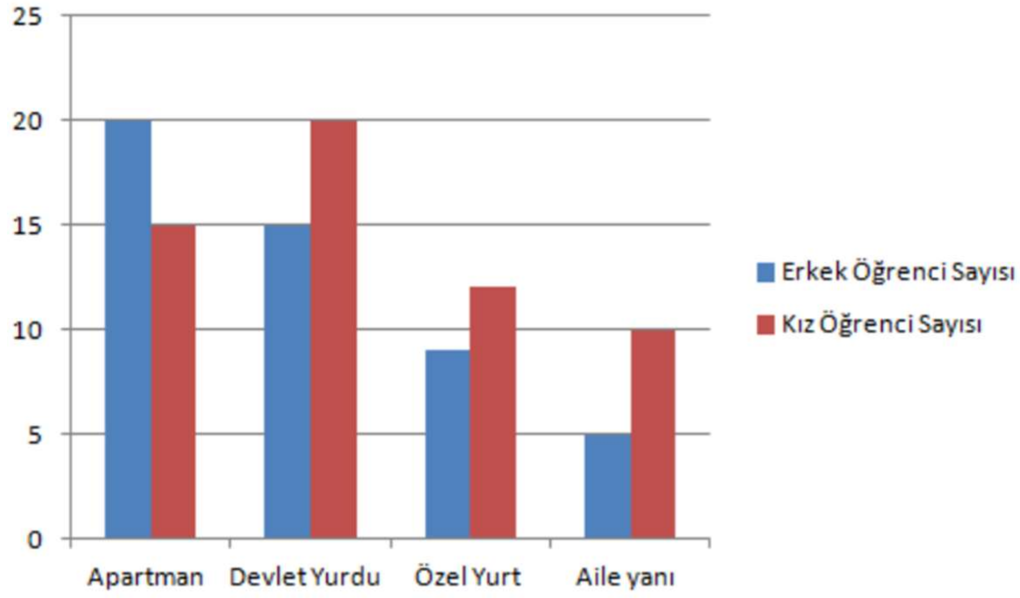
Barınma Türü	Öğrenci Sayısı
Apartman	20
Devlet Yurdu	15
Özel Yurt	9
Aile yanı	5
Total	49



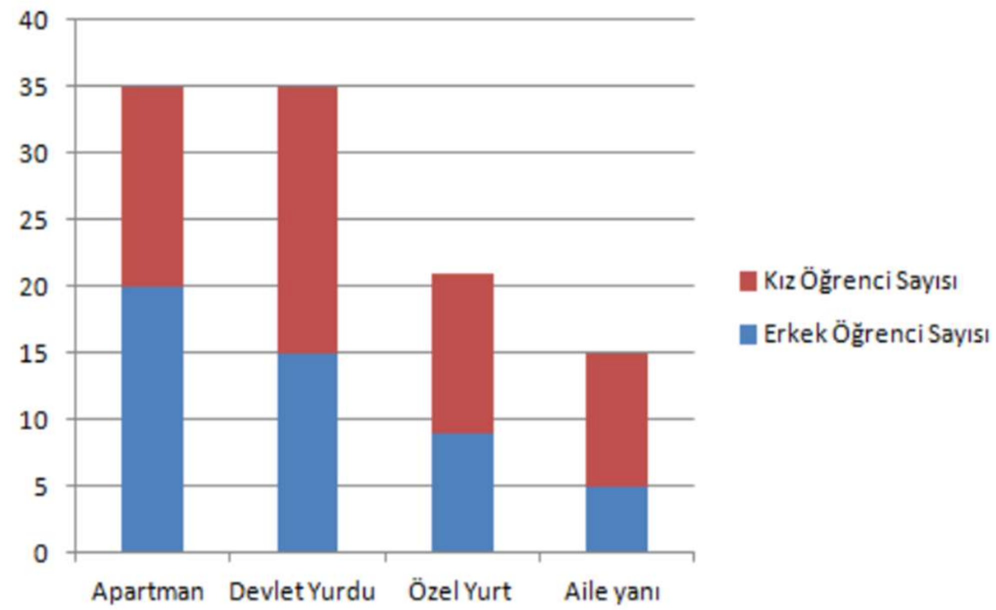
GRAFİKSEL GÖSTERİMLER

- **YanYana Bar Grafikleri:**

Barınma Türü	Öğrenci Sayısı
Apartman	20E, 15K
Devlet Yurdu	15E, 20K
Özel Yurt	9E, 12K
Aile yanı	5E, 10K
Total	49E, 57K



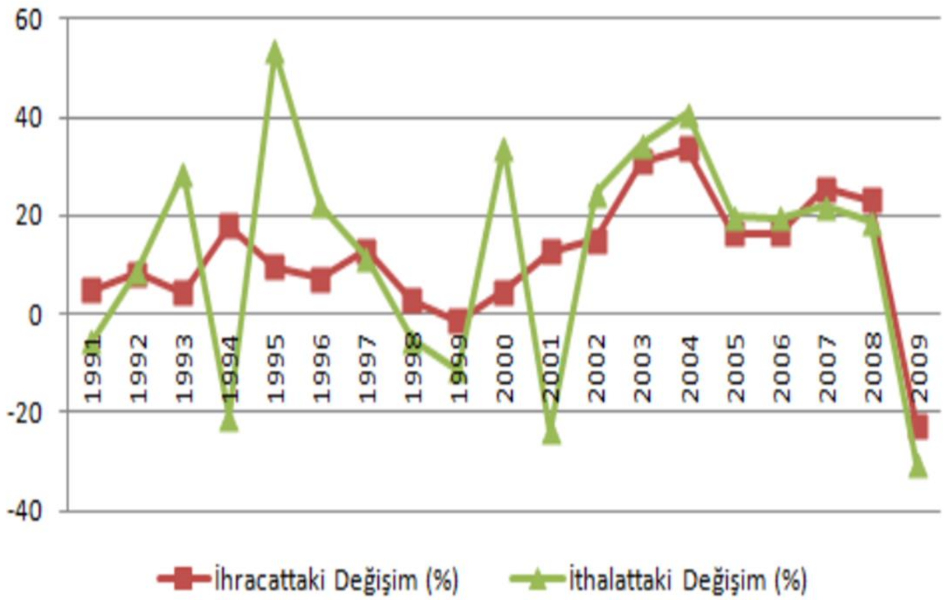
- **Parçalı Bar Grafikleri:**



GRAFİKSEL GÖSTERİMLER

➢ ## ÇİZGİ GRAFİKLERİ

Yıllar	İhracattaki Değişim (%)	İthalattaki Değişim (%)
1991	4,9	-5,6
1992	8,2	8,7
1993	4,3	28,7
1994	18	-20,9
1995	9,5	53,5
1996	7,3	22,2
1997	13,1	11,3
1998	2,7	-5,4
1999	-1,4	-11,4
2000	4,5	34
2001	12,8	-24
2002	15,1	24,5
2003	31	34,5
2004	33,7	40,7
2005	16,3	19,7
2006	16,4	19,5
2007	25,4	21,8
2008	23,1	18,8
2009	-22,6	-30,3



GRAFİKSEL GÖSTERİMLER

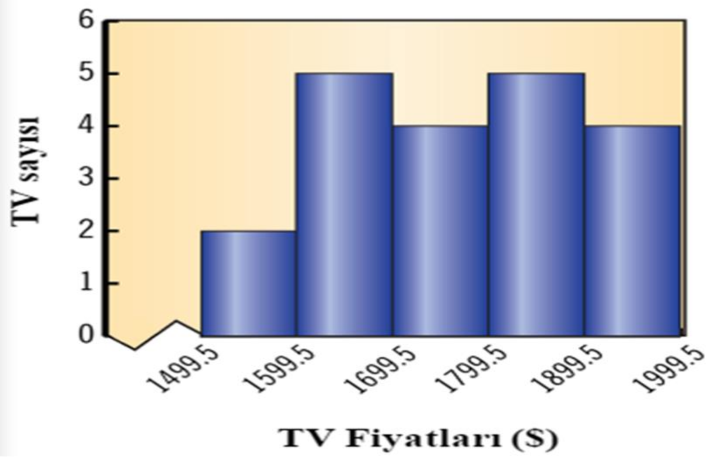
➢ HİSTOGRAM

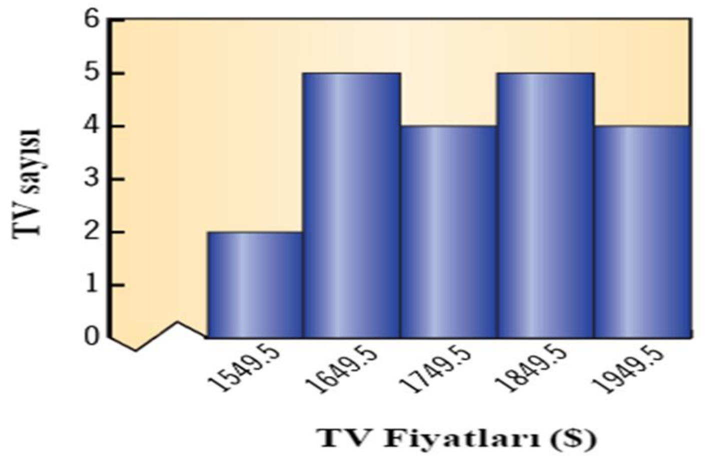
- Bar grafiklerine benzer.
- Barların yanlarındaki barlarla teması vardır.
- Yatay eksen, her bir sınıfın sınırını gösteren reel sayı eksenidir.
- Barların genişliği sınıf genişliğini temsil eder.
- Barların yüksekliği sınıfların frekanslarını gösterir.

Plazma TV Fiyatları

Sınıf	*F*	Orta Nokta	Sınıf Sınırları
$1500 – $1599	2	1549.5	1499.5 – 1599.5
$1600 – $1699	5	1649.5	1599.5 – 1699.5
$1700 – $1799	4	1749.5	1699.5 – 1799.5
$1800 – $1899	5	1849.5	1799.5 – 1899.5
$1900 – $1999	4	1949.5	1899.5 – 1999.5

Histogramların çizimi için sınıf sınırlarının kullanılmasına rağmen, yatay eksenin etiketlenmesinde ya sınıf sınırları (soldaki) yada orta noktalar (sağdaki) kullanılmalıdır.





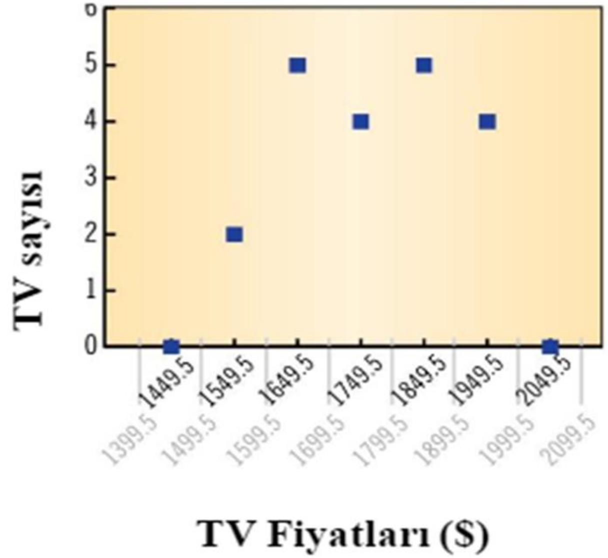
GRAFİKSEL GÖSTERİMLER

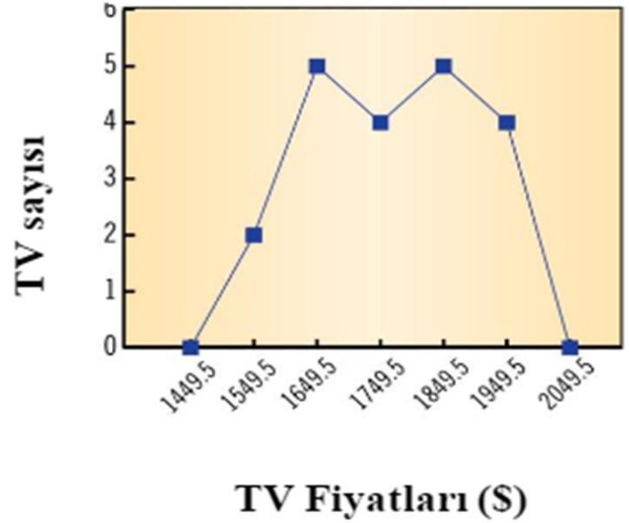
- **FREKANS POLİGONU**

Yatay eksende sınıf sınırları, dikey eksende ise frekanslar işaretlenir.

Orta noktalar yatay eksene eklenir, her bir sınıf için orta noktaların hizasında gelecek şekilde frekanslar işaretlenir.

Noktalar birleştirilir.





GRAFİKSEL GÖSTERİMLER

➢ **DAL-YAPRAK (GÖVDE-YAPRAK) GÖSTERİMİ**

Kursa Katılanların Yaşları

18	23	24	31	19
27	26	22	32	18
35	27	29	24	20
18	17	21	25	26

Gövde	Yapraklar
1	8 9 8 8 7
2	3 4 7 6 2 7 9 4 0 1 5 6
3	1 5 2

MERKEZİ EĞİLİM VE DAĞILIM ÖLÇÜMLERİ

Merkezi Eğilim ve Dağılım Ölçümleri

Merkezi Eğilim (Yığılma) Ölçüleri

- Ortalama
  - Aritmetik ortalama
  - Ağırlıklı ortalama
  - Geometrik ortalama
  - Harmonik ortalama
- Ortanca (medyan)
- Tepe değer (mod)

Merkezi Dağılım (Yayılma) Ölçüleri

- Değişim Aralığı (Genişliği)
- Ortalama Mutlak Sapma
- Varyans
- Standart sapma
- Değişim (Varyasyon) Katsayısı
- Alt çeyrek, üst çeyrek, çeyrekler açıklığı

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Merkezi Eğilim (Yığılma) Ölçüleri

- Ortalama
  - * Aritmetik ortalama
  - * Ağırlıklı ortalama
  - * Geometrik ortalama
  - * Harmonik ortalama
- Ortanca (medyan)
- Tepe değer (mod)

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Aritmetik Ortalama: Sınıflandırılmamış (gruplanmamış) yani ham veriler için aritmetik ortalama, bir veri setindeki tüm veri değerleri toplamının veri sayısına oranıdır.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

x_i : Veri Değeri

$\overline{x}$: Aritmetik Ortalama

n : Veri Sayısı

Örnek:

Ölçüm numarası n	Ölçüm değerleri x_i °C
1	5,6
2	5,8
3	4,9
4	6,8
5	6,4
6	7,4
7	6,5
8	7,7
9	5,5
10	5,7
11	6,7
12	6,9
13	6,2
14	6,0
15	5,9
	Σ 94,00

Ölçüm değerlerinin Aritmetik Ortalaması:

$$\frac{5.6+5.8+4.9+\cdots+5.9}{15}$$

$$\frac{94}{15} = 6.266\,^{\circ}\mathrm{C}$$

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Tekrarlı Gözlemlerin Aritmetik Ortalaması: Bir veri setinde sınıflandırılmamış değerler birden fazla kez tekrarlanabilir. Böyle durumda;

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} fi * xi$$

f_i = Bir örnekteki x_i'nin frekansı

k= örnekteki farklı gözlem sayısı,

n= örnekteki toplam veri sayısı

xi= i. Veri değeri

Örnek: 20 kişilik bir sınıfta istatistik dersinden geçen öğrencilerin notları şu şekildedir: 5,5,5,5,5,6,6,6,7,7,8,8,8,9,9,9,10,10,10,10

x1=5 f1=5, x2=6 f2=3, x3=7 f3=2, x4=8 f4=3, x5=9 f5=3, x6=10 f6=4 k=6, n=20

$$\bar{x} = \frac{(5*5)+(3*6)+(2*7)+(3*8)+(3*9)+(4*10)}{20} = 7.4$$

Veya yine basit aritmetik ortalama formülünden de hesaplanabilir. $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

$$= \frac{5+5+5+5+6+6+6+7+7+8+8+8+9+9+9+10+10+10+10}{20} = 7.4$$

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Ağırlıklı Aritmetik Ortalama: Aritmetik ortalamada, veri setindeki her bir veri değerinin öneminin eşit olduğu varsayılmaktadır. Fakat bazı durumlarda bazı değerlerin önemleri diğerlerinden farklı olabilir. Bu durumda ağırlıklı ortalama kullanılır.

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \qquad \textbf{yani} \qquad \bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}$$

Örnek: Bir öğrencinin sınava girdiği derslerin önem dereceleri (ağırlıkları) farklı ise not ortalaması hesabında derslerin önem derecelerinin göz önüne alınması gerekir. Kabul edelim ki bir öğrenci için aşağıdaki veriler vardır.

DERS	AĞIRLIK	PUAN
A	3	30
B	2	45
C	2	25
D	1	40
E	4	15
F	2	6
G	2	55

$$\bar{x} = \frac{3*30+2*45+2*25+1*40+4*15+2*6+2*55}{3+2+2+1+4+2+2}$$

$$\bar{x} = \frac{452}{16} = 28.8$$

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Aritmetik Ortalamanın Dezavantajı

Veri setindeki aşırı (sapan/aykırı) değerlerden kolay etkilenmesidir. Bir veri setindeki verilerden birkaçı çok yüksek veya çok düşük değerler içeriyor ise, aritmetik ortalama, veri setinin merkezi eğilim ölçümünü temsil etmek için uygun olmayabilir.

Örnek: 5.30, 0.10, 5.60, 5.80, 5.50

Verilen sayıların aritmetik ortalaması 4.46'dır. Bu veri dizisi içerisindeki sapan değer 0.10'dur. Bu değer ortalamayı aşağı çekmektedir. Bir normalizayon/standartlaştırma belirlenmeli ya da diziden sapan değer çıkarıldıktan sonra aritmetik ortalama hesaplanmalıdır. Böylece hesaplanan ortalama daha anlamlı olacaktır.

Sapan değerin çıkarılması ile dizinin aritmetik ortalaması yeniden hesaplanırsa 5.55 elde edilir.

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Geometrik Ortalama: Sınıflandırılmamış bir veri setindeki n tane x_1, x_2, …, x_n değerinin çarpımının n. Kökü geometrik ortalama olarak tanımlanır.

$$\boxed{G.O.=\sqrt[n]{x1 * x2 * \cdots * xn}}$$

Veri sayısı çok fazla olduğu durumlarda hesaplamaları kolaylaştırmak amacıyla logaritmalardan yararlanılır.

$$\boxed{G.O.=\frac{1}{n}\sum_{i=1}^{n}\log xi}$$

Tekrarlanan Verilerde Geometrik Ortalama: Sınıflandırılmamış bir veri setindeki n tane veriden bazıları birden fazla kez tekrarlanabilir. Böyle durumda;

$$\boxed{G.O. =(x1^{f1}+x2^{f2}+\cdots+xk^{fk})^{\frac{1}{n}}} \text{ veya } \boxed{\log(G.O.)=\frac{1}{n}\sum_{i=1}^{k} fi * \log xi}$$

Harmonik Ortalama: $$\boxed{H.O. = \frac{1}{\frac{1}{n}\left(\frac{1}{x1}+\frac{1}{x2}+\cdots+\frac{1}{xn}\right)} = \frac{n}{\sum_{i=1}^{n}\frac{1}{xi}}}$$

Aritmetik, Geometrik ve Harmonik ortalama arasındaki ilişki;

H.O < G.O < A.O

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Ortanca(Medyan): Bir veri setinde büyüklüklerine göre sıralanmış verilerin **merkezi değerine** medyan denir. Çift sayıda gözlemlenmiş veri varsa ortanca değer (medyan) iki merkezi değerin aritmetik ortalamasıdır. Her bir verinin tek bir medyanı vardır. Veri setindeki aşırı (sapan) değerlerden etkilenmediği için verilerin merkezi eğiliminin belirlenmesinde aritmetik ortalamaya nazaran daha doğru bir bilgi sunar.

Matematiksel olarak x1, x2, …, xn gözlem değerleri büyüklüklerine göre artan sırada düzenlenmişse, M ile gösterilen medyan aşağıdaki gibi hesaplanır.

n tek ise → $M=x_{\left(\frac{n+1}{2}\right)}$, n çift ise → $\frac{1}{2}\left(x_{\frac{n}{2}}+x_{\left(\frac{n}{2}\right)+1}\right)$

Örnek: 2, 3, 2, 4, 4, 6, 6, 5, 8, 8, 9 sayıları için ortanca (medyan) değeri bulunuz.

Çözüm: Önce veriler küçükten büyüğe doğru sıralanır.

2, 2, 3, 4, 4, 5, 6, 6, 8, 8, 9 n=11 olduğu için $M= x_{\left(\frac{n+1}{2}\right)} = x_{\frac{11+1}{2}} = x_6 =5$

Örnek: 1,2,3,3,5,5,5,6,7,7,7,8,9,9 biçiminde sıralanmış veriler için ortanca değeri bulunuz.

Çözüm: n=14 çift sayıdır. O halde $\frac{n}{2}$ =7, $\left(\frac{n}{2}\right)+1$ =8, M(x_7 + x_8)=(5+6)/2=5.5'dir.

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Tepe Değer (Mod): Bir veri setinde en çok tekrarlanan, en sık görülen veri değeridir. Her değer yalnız bir kez ya da tüm değerler eşit miktarda elde edilmişse mod yoktur.

Örnek: 3,5,6,9,9,9,10,12 değerleri için mod 9'dur.

Örnek: 1, 3, 4, 4, 4, 5, 6, 8, 8, 8, 10, 12 değerleri için 4 ve 8 olmak üzere iki mod değeri vardır.

SINIFLANDIRILMIŞ (GRUPLANMIŞ) VERİLERDE MERKEZİ EĞİLİM ÖLÇÜMLERİ

Bazı durumlarda veri değerleri sınıflandırılıp (gruplandırılıp), frekans dağılımları oluşturulmuş olabilir ve ham veriler mevcut olmayabilir. Bu gibi durumlarda aritmetik ortalama, medyan ve mod değerleri frekans dağılım tablolardan hesaplanabilir.

Bu hesaplanan değerler gerçek ham verilerden hesaplanandan farklı olabilir.

SINIFLANDIRILMIŞ (GRUPLANDIRILMIŞ) VERİLERDE MERKEZİ EĞİLİM ÖLÇÜMLERİ

Sınıflandırılmış (Gruplandırılmış) Verilerde Aritmetik Ortalama

$$\bar{x} = \frac{f1*y1+f2*y2+\cdots+fk*yk}{n} = \frac{\sum_{i=1}^{k} \text{fi}*\, yi}{n}$$

yi= i. Sınıfın orta değeri, k=sınıf sayısı, fi= i. sınıfın frekansı, n= toplam gözlem sayısı

Örnek: 100 satış elamanı tarafından gerçekleştirilen aylık satışlar için frekans dağılım tablosu

Sınıf No	Sınıf Limitleri	Sınıf Sınırları	Gözlem Sayısı (f_i)	Sınıf Orta Noktası (y_i)	Eklemeli frekanslar
1	0-2	-0,5 ~ 2,5	1	1	1
2	3-5	2,5 ~ 5,5	3	4	4
3	6-8	5,5 ~ 8,5	8	7	12
4	9-11	8,5 ~ 11,5	13	10	25
5	12-14	11,5 ~ 14,5	20	13	45
6	15-17	14,5 ~ 17,5	25	16	70
7	18-20	17,5 ~ 20,5	17	19	87
8	21-23	20,5 ~ 23,5	5	22	92
9	24-26	23,5 ~ 26,5	3	25	95
10	27-29	26,5 ~ 29,5	5	28	100

Çözüm: $\frac{\sum_{i=1}^{10} \text{fi}*\, yi}{100}$

$$= \frac{(1*1)+(3*4)+(8*7)+\cdots+(5*28)}{100}$$

$$= \frac{1+12+56+\cdots+75+140}{100} = 15{,}07$$

SINIFLANDIRILMIŞ (GRUPLANDIRILMIŞ) VERİLERDE MERKEZİ EĞİLİM ÖLÇÜMLERİ

Sınıflandırılmış (Gruplandırılmış) Verilerde Ortanca (Medyan)

n= gözlem sayısı, L_m=Medyan sınıfın alt sınıf sınırı, h= Sınıf genişliği, f_m= Medyan sınıfın frekansı, n_m= Medyan sınıfından önceki sınıfların frekansları toplamı olmak üzere;

$M=L_m + \frac{\left(\frac{n}{2} - nm\right)}{fm}$*h dir ve f1 + f2 +…+fi ≥ $\frac{n}{2}$ olan ilk sınıfa medyan sınıfı denir.

Örnek:

Sınıf No	Sınıf Limitleri	Sınıf Sınırları	Gözlem Sayısı (f_i)	Sınıf Orta Noktası (y_i)	Eklemeli frekanslar
1	0-2	-0,5 ~ 2,5	1	1	1
2	3-5	2,5 ~ 5,5	3	4	4
3	6-8	5,5 ~ 8,5	8	7	12
4	9-11	8,5 ~ 11,5	13	10	25
5	12-14	11,5 ~ 14,5	20	13	45
6	15-17	14,5 ~ 17,5	25	16	70
7	18-20	17,5 ~ 20,5	17	19	87
8	21-23	20,5 ~ 23,5	5	22	92
9	24-26	23,5 ~ 26,5	3	25	95
10	27-29	26,5 ~ 29,5	5	28	100

M=?

n=100 olduğu için $\frac{n}{2}$ =50 dir.

f1+f2+f3+f4+f5=45<50 ve f1+f2+f3+f4+f5+f6=70>50 olduğu göz önüne alınırsa 6. sınıfın medyan sınıf olduğu görülür.

$M=14.5 + \frac{(50-45)}{25}$*(3)=

14.5+0.6=15.1

SINIFLANDIRILMIŞ (GRUPLANDIRILMIŞ) VERİLERDE MERKEZİ EĞİLİM ÖLÇÜMLERİ

Sınıflandırılmış (Gruplandırılmış) Verilerde Tepe Değeri (Mod)

Frekans tablosunda eşit sınıf genişliklerinin olması durumunda mod'u inceleyeceğiz. Öncelikle modal sınıf bulunmalıdır. Modal sınıf (modu içinde bulunduran sınıf) en çok gözlem sayısına (frekansa) sahip olan sınıftır.

L_{td}= Modal sınıfın alt sınıf sınırı

$\Delta 1$= Modal sınıfın frekansı ile bir önceki sınıfın frekansı farkı

$\Delta 2$= Modal sınıfın frekansı ile bir sonraki sınıfın frekansı farkı

h= Sınıf genişliği

olmak üzere

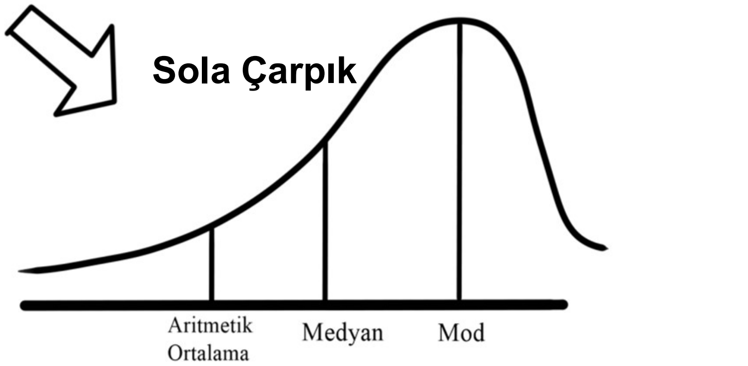
$$\text{Mod} = L_{td} + \frac{\Delta 1}{\Delta 1 + \Delta 2} * h$$

Örnek: Bir önceki örnekteki tablodaki mod nedir?

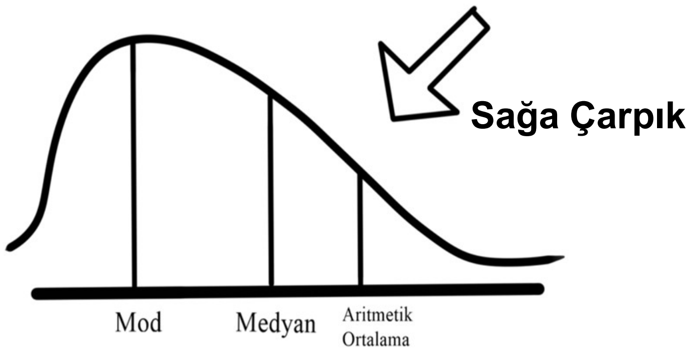
Çözüm: Altıncı sınıf modal sınıftır. Mod=$14{,}5 + \frac{(25-2\ \ )}{(25-20)\ \ (25-17)} * 3 = \frac{(5)}{(5)+(8)} * 3$

Mod=$14{,}5 + \frac{15}{13}$ =15,65

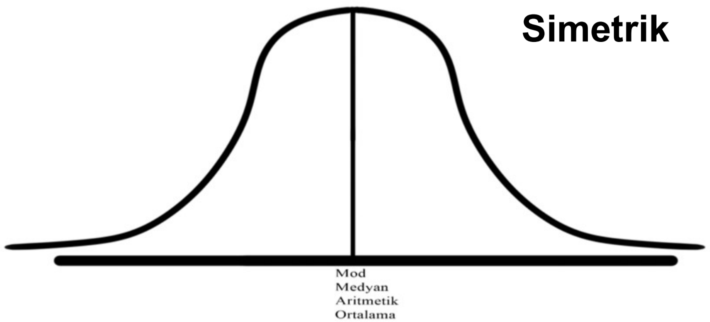
MERKEZİ EĞİLİM ÖLÇÜMLERİ



A.O < M< Mod



Mod < M < A.O



A.O=M=Mod

MERKEZİ EĞİLİM ÖLÇÜMLERİ

Merkezi eğilim ölçülerinin karşılaştırılması

- Aşırı (Sapan, Anormal) değerler aritmetik ortalamayı büyük ölçüde etkiler.
- Ortanca değerler (Medyan), aşırı değerler tarafından etkilenmez ancak gözlem sayısından etkilenir.
- Büyüklüklerine göre sıralanmış değerlerin çoğunluğu merkeze yakın ise ortanca değer iyi bir merkezi eğilim ölçüsüdür.
- Tepe değeri (Mod) aşırı değerlerin etkisinden bağımsızdır.
- Gözlem sayısının az olması halinde değerler tekrarlanmayabilir. Bu nedenle tepe değer belirlenemeyebilir.
- Çarpık dağılımlar için, medyan, ortalamadan daha iyi bir merkezi eğilim ölçüsüdür. Çünkü aritmetik ortalama çarpıklık yönünde merkez bölgeden uzaklaşmaktadır. Açık uçlu frekans tablosundan sınıf ortalamalarının tümü hesaplanamayacağından aritmetik ortalama hesaplanamaz. Böyle durumlarda medyan aritmetik ortalamaya göre daha uygun bir merkezi eğilim ölçüsüdür.