DAĞILMA YADA DEĞİŞKENLİK ÖLÇÜTLERİ (MEASURE OF DISPERSION)

AMAÇLAR

- Mevcut veri seti için bulunan merkezi eğilim ölçüsünün yorumlamak
- Birden fazla veri seti için dağılımlar arası kıyaslama yapabilmek

amaçlarıyla değişkenlik ölçüleri analizi yapılır.

ÇEŞİTLERİ

DEĞİŞKENLİK ÖLÇÜTLERİ

Değişkenlik aralığı	**Ortalama sapma**	**Varyans**	**Standart sapma**
(range)	(mean deviation)	(variance)	(standart deviation)

a) DEĞİŞKENLİK ARALIĞI (Range)

Bir veri setinde mevcut olan minimum ve maksimum değerler arasındaki fark olarak tanımlanabilir.

O halde; değişkenlik aralığI(R)

R = Maksimum değer – Minimum değer

Örnek

- ***GRUPLANMAMIŞ DATA İÇİN:*** Aşağıda iki veri serisi ve bunlar için hesaplanan değişkenlik aralığı verilmiştir.

SERİ 1	SERİ 2
2	5
3	5
6	5
7	6
8	7
10	8
R=10-2=8	**R=8-5=3**

- ***GRUPLANMIŞ VERİLER İÇİN:*** Örnekte verilern frekans dağılım tablosunda değişkenlik aralığını belirleyelim.

SAATLİK ÜCRET(ytl)	FREKANS
5-10	10
10-15	21
15-20	9
20-25	5

R=25-5=20

b) ORTALAMA - MUTLAK ORTALAMA SAPMA (mean deviation or mean absolute deviation)

Bir popülasyon için tüm veri değerlerinin ortalamaları bulunur. Daha sonra her bir veri değeri için bu ortalamadan sapmaların mutlak değerleri toplanır ve data sayısına bölünür. Yani; her bir verinin ortalamadan sapmalarının mutlak değerinin aritmetik ortalamasıdır.

N= veri sayısı

$$\boxed{O.S. = \frac{\Sigma\left|X - \overline{X}\right|}{N}}$$

$\left|X - \overline{X}\right|$ = mutlak sapma

Örnek

- ***GRUPLANMAMIŞ DATALAR İÇİN:*** 15,16,18,21,25 data serisi için ortalama sapmayı hesaplayacak olursak;

Veriler	$\left|X - \overline{X}\right|$	$\left|X - \overline{X}\right|$
15	|15-19|	4
16	|16-19|	3
18	|18-19|	1
21	|21-19|	2
25	|25-19|	6
	TOPLAM	16

$\overline{X}$=(15+16+18+21+25)/5=19

O.S=16/5=3.2

◆ ***GRUPLANMIŞ DATALAR İÇİN:*** Bu tip veriler için artık fX ağırlıklandırmasına göre ortalama hesaplanıp diğer işlemler yapılmalıdır.

Sınıf	frekans (f)	Sınıf orta noktası (X)	fx	$\lvert X-\overline{X}\rvert$	$\lvert X-\overline{X}\rvert$	$f\lvert X-\overline{X}\rvert$
10-20	7	15	105	$\lvert 15-33.2\rvert$	18.2	127.4
20-30	14	25	350	$\lvert 25-33.2\rvert$	8.2	114.8
30-40	16	35	560	$\lvert 35-33.2\rvert$	1.8	28.8
40-50	9	45	405	$\lvert 45-33.2\rvert$	11.8	106.2
50-60	5	55	275	$\lvert 55-33.2\rvert$	21.8	109
Toplam	*51*		*1695*			*486.2*

$$\overline{X} = \frac{\sum fX}{N} = \frac{1695}{51} = 33.2$$

$$O.S. = \frac{\sum f\lvert X-\overline{X}\rvert}{N} = \frac{486.2}{51} = 9.53$$

c) VARYANS (variance)

- Bir değişkenin aldığı değerlerin ortalamadan sapmasını (yani merkeze ne kadar yakın olduğunu) gösteren ölçü birimine **varyans** denir. Yine bir aritmetik ortalama ile hesaplanır.
- **!!!** Varyans hiçbir zaman negatif olamaz.

Bir örnek için varyans;

$$\sigma^2 = \frac{\Sigma\left(X - \overline{X}\right)^2}{N}$$

Özellikleri

- Herhangi bir a sabitinin varyans değeri daima sıfırdır. Bu nedenle bir serideki değerlere herhangi bir sabitin eklenmesi ya da çıkarılması serinin varyansını etkilemez. Yani a bir sabit olmak üzere;

$$\sigma^2(X-a)=\sigma^2(X)$$
$$\sigma^2(X+a)=\sigma^2(X)$$

- Herhangi bir c sabiti ile veri setindeki tüm değerleri çarparak elde edilen setin varyansı ilk setin varyansı ile sabitin karesi çarpımına eşittir. Yani c bir sabit olmak üzere;

$$\sigma^2(cX)=c^2\sigma^2(X)$$

....

Tek bir veri setinin varyansını yorumlayabilmek ortada bir kıyas söz konusu olmadığı için çok zordur.

Bu sebeple ortalama sapma ve değişkenlik aralığı gibi varyans için de yorumlama en az 2 veri seti için yapılır.

Bu setlerin değişkenlik dereceleri varyans analizi ile karşılaştırılır.

d) STANDART SAPMA (standart deviation)

- Bir veri seti için varyans hesabı yapıldıktan sonra bunun karekökü alınarak o setin standart sapması bulunabilir.

$$\sigma = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{N}}$$

- Varyans için geçerli özelliklerden yola çıkarak farklı standart sapma durumları incelenebilir.

Örnek

- ***GRUPLANMAMIŞ DATALAR İÇİN:***22,25,28,30 ve 35 veri seti için standart sapma ve varyans değerlerini hesaplayalım.

X	$(X-\overline{X})$	$(X-\overline{X})^2$
22	-6	36
25	-3	9
28	0	0
30	2	4
35	7	49
$\Sigma X = 140$	$\sum(X-\overline{X}) = 0$	$\sum(X-\overline{X})^2 = 98$

$$\overline{X} = \frac{140}{5} = 28$$

$$\sigma^2 = \frac{98}{5} = 19.6$$

$$\sigma = \sqrt{19.6} = 4.43$$

◆ ***GRUPLANMIŞ DATALAR İÇİN:***Aşağıda frekans dağılım tablosu için ağırlıklandırmalardan faydalanarak varyans ve standart sapmayı belirleyelim.

sınıflar	frekans (f)	Sınıf orta noktası (X)	fX	$(X-\overline{X})$	$(X-\overline{X})^2$	$f(X-\overline{X})^2$
0-200	8	100	800	-270	72900	583200
200-400	11	300	3300	-70	4900	53900
400-600	7	500	3500	130	16900	118300
600-800	6	700	4200	330	108900	653400
Toplam	*32*		*11800*			*1408800*

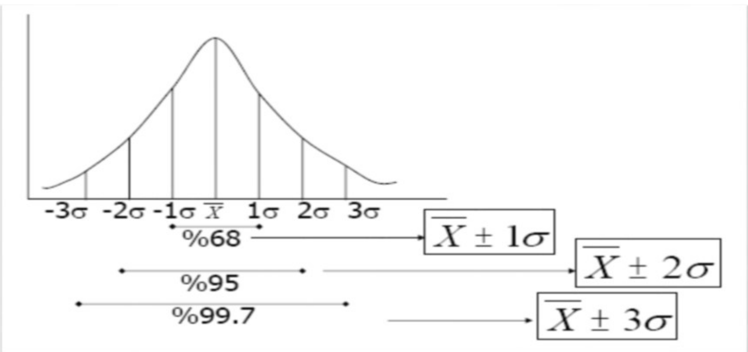
$$\overline{X} = \frac{11800}{32} = 370$$

$$\sigma^2 = \frac{1408800}{32} = 44025$$

$$\sigma = \sqrt{44025} = 209.82$$

STANDART SAPMA VE ARİTMETİK ORTALAMA İLİŞKİSİ

- Simetrik frekans dağılım grafikleri kullanılrak %68, %95, %99.7 olasılıklarına karşılık gelen bir analizi şu şekilde yaparız:



NİSBİ DAĞILMA (relative dispersion)

DEĞİŞİM KATSAYISI(coeffient of variation): Bir veri setinde standart sapmanın veri setinin ortalamasına oranı olarak ifade edilebilir. Fakat değişim katsayısı % bir değişken olduğundan 100 ile ağırlıklandırılmalıdır.

$$\text{Degisim Katsayisi}(\%) = \frac{\sigma}{\overline{X}} * 100$$

- Faklı birimlere sahip veri setleri kıyaslanırken
- Aynı birimli fakat çok faklı ortalamalara sahip veri setleri kıyaslanırken kullanılır.

ÇARPIKLIK (skewness)

- Frekans dağılım grafikleri için söz konusu olan bir ifadedir.
- Veri setlerinde çarpıklık standart sapma, ortalama ve medyan kullanılarak hesaplanır. Genellikle -3 ve 3 aralığında bir değerdir.
- Simetrik grafikler için çarpıklık=0 bulunur.

$$\text{Carpiklik} = \frac{3(\overline{X} - \text{medyan})}{\sigma}$$

Farklı Dağılımlara Göre Çarpıklık Çeşitleri

- Simetrik frekans dağılım grafiği için,

- Negatif asimetrili frekans dağılım grafiği için,

- Pozitif asimetrili frekans dağılım grafiği için,