

TOBB-ETÜ, İktisat Bölümü

İstatistik (İKT 253)

1. Çalışma Soruları - Cevaplar

**Soru 1:** Bir hafta boyunca saat 2-3pm arasında bir mağazayı ziyaret eden insan sayısı aşağıdaki gibidir

	Pzt.	Sa.	Çar.	Per.	Cu.	Cmt.	Paz.
Mağaza 1:	10	25	20	25	35	55	50

a-) Bu data için frekans dağılım tablosunu (frequency distribution table), günlük ziyaretçi sayısının toplam ziyaretçi sayısı içindeki yüzdesiyle hazırlayınız

	Pzt.	Sa.	Çar.	Per.	Cu.	Cmt.	Paz.
	10	25	20	25	35	55	50
$\frac{10}{220} \cong$	4.5	%11.3	%9	%11.3	%16	%25	%22.5

**b-)** Ziyaretçi sayılarını çubuk grafik (bar chart), dairesel grafikte (pie chart), ve çizgisel grafikte (line chart) ile gösteriniz (grafiklerin çok düzgün gözükmelerine gerek duymadan)

c-) Datanın pareto diyagramını (kümülatif dağılımla beraber) çiziniz

Pzt.	Sa..	Çar.	Per.	Cu.	Paz.	Cmt.
10	25	20	25	35	50	55
%4.5	%11.3	%9	%11.3	%16	%25	%22.5
%4.5	%15.8	%24.8	%36	%52	%75	%100

d-) Mağazayı günde ortalama kaç kişinin ziyaret ettiğini, yani ortalamasını (mean), ayrıca medyan (median), ve modunu da (mode) hesaplayınız

	Pzt.	Sa.	Çar.	Per.	Cu.	Cmt.	Paz.
Mağaza 1:	10	25	20	25	35	55	50

$$Mean = \bar{x} = \left( \sum_{i=1}^n x_i \right) / n =$$

$$(10 + 25 + 20 + 25 + 35 + 55 + 50) / 7 = 31.4$$

$$Medyan = 25 \quad Mod : 25$$

e-) Bu datayı hangi merkezi eğilim ölçütü (measure of central tendency) en iyi tanımlayabilir, yorumlayınız

	Pzt.	Sa.	Çar.	Per.	Cu.	Cmt.	Paz.
Mağaza 1:	10	25	20	25	35	55	50

**f-)** Örneklem varyansını (variance), standart sapmasını (standard deviation), ve varyasyon katsayısını (coefficient of variation) bulunuz

$$\begin{aligned} \text{Variance} = s^2 &= \left( \sum_{i=1}^N (x_i - \bar{x})^2 \right) / (n - 1) = \\ &= [(10 - 31.4)^2 + (25 - 31.4)^2 + (20 - 31.4)^2 + (25 - 31.4)^2 \\ &+ (35 - 31.4)^2 + (55 - 31.4)^2 + (50 - 31.4)^2] / 6 = 264.2 \end{aligned}$$

$$\text{Standartsapma} = s = \sqrt{264.2} = 16.25$$

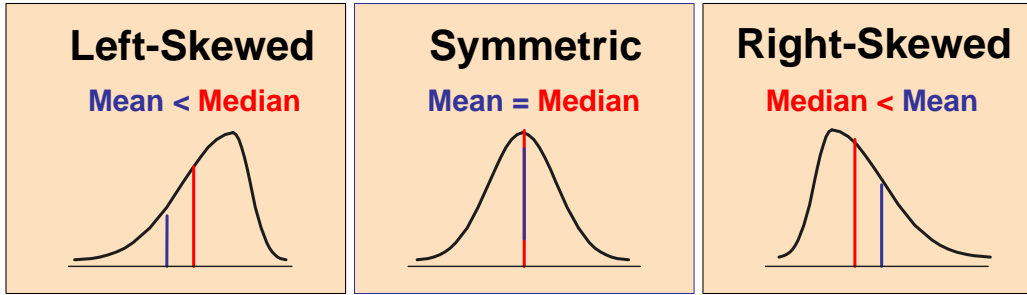
$$\text{Varyasyon kat.} = \frac{s}{\bar{x}} * 100 = \frac{16.25}{31.4} * 100 = \%50$$

**g-)** Datanın ortalamasına en fazla iki standart sapma uzaklıktaki sayıları tanımlayan aralığı bulunuz

$$\begin{aligned} [\bar{x} - 2s, \bar{x} + 2s] &= [31.4 - 2 * 16.25, 31.4 + 2 * 16.25] = \\ &= [-1, 63] \end{aligned}$$



**h-)** Bu datanın uzun vadede tekrarlandığını düşündüğünü ortaya çıkacak dağılımın simetrik olup olmadığı; simetrik değilse hangi tarafa doğru çarpık (skewed) olduğunu yorumlayınız



Sağa yatık

**i-)** Datanın çarpıklığını (skewness) hesaplayınız ve bir önceki seçeneği yorumlayınız

$$Skewness = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(n - 1)s^3} = 0.3$$

|

Mağazanın ziyaretçi başına ortalama kazancını haftaiçi 5TL, haftasonunda ise 3 TL kabul edelim

**j-**) Mağazanın günlük kazanç dadasını çubuk grafik (bar chart) ile gösteriniz

Pzt.	Sa.	Çar.	Per.	Cu.	Cmt.	Paz.
10	25	20	25	35	55	50
$10 * 5$	$25 * 5$	$20 * 5$	$25 * 5$	$35 * 5$	$55 * 3$	$50 * 3$

**k-)** Kazanç dadasının ortalamasını (ki bu ziyaretçi dadasının fiyat *ağırlıklı ortalaması* olacaktır) ve varyasyonunu hesaplayınız

$$\text{Mean} = \bar{x} = \left( \sum_{i=1}^n x_i \right) / n = (50 + 125 + 100 + 125 + 175 + 165 + 150) / 7 = 127.14$$

$$\text{Variance} = s^2 = \left( \sum_{i=1}^N (x_i - \bar{x})^2 \right) / (n - 1) = 1843$$

$$\text{Standart sapma} = s = 43$$

1-) Kazanç dadasının şekli (skewness) ziyaretçi sayısının şekline göre ne yöne değişti, yorumlayınız

$$Skewness = \left( \sum_{i=1}^N (x_i - \bar{x})^3 \right) / [(n - 1)s^3] = 0$$

|

Aynı yerleşkede başka bir giyim mağazasını ziyaret eden insan sayısı sayısı aşağıdaki gibi olsun

	Pzt.	Sa.	Çar.	Per.	Cu.	Cmt.	Paz.
Mağaza 2:	15	24	27	28	36	43	45

**m-)** İki datayı saçılım grafiđi (scatter plot), ve kenar kenara çubuklu grafik (side-by-side bar chart) ile çiziniz



**n-)** İkinci mağazanın günlük ortalama ziyaretçi sayısını, ve bunun standart sapmasını bulup, bunu birinci mağaza için bulduğunuz değerlerle karşılaştırınız

$$\text{Mean} = \bar{y} = \left( \sum_{i=1}^n y_i \right) / n = 31.14$$

$$\text{Variance} = s^2 = \left( \sum_{i=1}^N (y_i - \bar{y})^2 \right) / (n - 1) = 115.7$$

$$\text{Standartsapma} = s = 10.7$$

o-) İkinci mağazanın günlük ziyaretçi sayısı dağılımının varyasyon katsayısını (coefficient of variation) bulup, bunları daha önce bulduğunuz ilk mağazanınkiyle karşılaştırırız

$$\text{Varyasyon katsayısı} = (s/\bar{x}) * 100 = \%28$$

**p-)** **n** ve **o** şıklarında yaptığınız karşılaştırmaların birbiriyle ilgisini yazıp yorumlayınız

**r-)** İki data arasındaki örneklem kovaryansını (covariance), ve korelasyon katsayısını (correlation coefficient) hesaplayıp, yorumlayınız

$$Cov(x, y) = \left( \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \right) / (n - 1) = 168$$

$$r = \frac{Cov(x, y)}{s_x s_y} = \frac{168}{16.25 * 10.7} = 0.96$$

s-) Eđer 2. mađazanın ziyaretçi başına ortalama kazancı haftaıçi 3TL, haftasonunda 5TL ise, bu mađazanın günlük ortalama kazancını ve bu kazancın varyasyonunu hesaplayınız

$$\text{Mean} = \bar{y} = \left( \sum_{i=1}^n y_i \right) / n = 118.57$$

$$\text{Variance} = s^2 = \left( \sum_{i=1}^N (y_i - \bar{y})^2 \right) / (n - 1) = 5154$$

$$\text{Standartsapma} = s = 72$$

t-) İki mağazanın ortalama kazançlarını, ve bunların standart sapmalarını karşılaştırınız. Buradaki sonuçları  $n$  şikkındaki sonuçlarla karşılaştırınız

|

**u-**) Başka bir mağazanın aynı zamanlarda ortalama ziyaretçi sayısı 40, bu sayının varyasyonu'da 25 olsun. Chebychev teoremine kullanarak, bu dağılımın 30 ile 50 arasında olabilecek minimum gözlem yüzdesini hesaplayınız

|

- *Chebyshev's Theorem:* For *any distribution* (not necessarily normal) with mean  $\mu$  and standard deviation  $\sigma$ , and  $k > 1$ , the part of the observations that fall within the interval

$$\mu \mp k\sigma$$

(i.e.  $k$  standard deviations of the mean) includes at least this much of the data

$$100[1 - (1/k^2)]\%$$

|

$$[30, 50] = [40 - 2 * 5, 40 + 2 * 5]$$

$$= [\bar{x} - 2s, \bar{x} + 2s] \Rightarrow k = 2$$

and

$$(1 - 1/k^2) = (1 - 1/2^2) = 75\%$$