ASSIGNMENT-1

Course Name: Probability & Statistics
Course Code: MAT335
Section- 02
Date- 20-10-2025

Submitted by

Md. Redoan Rahman Rifat

Student ID: 222 226 038

Semester: 11th(running)

Submitted to

Muddachsir Rahman

Lecturer

Department of Computer Science & Engineering

Presidency University

CamScanner

Md. Fedoan Rahman Rifat
ID: 222 226 038

Question: 1|

Twenty adult males between the ages 30 and 40 participated in a study to evaluate the effect of a specific health regimen involving diet and exercise on blood Cholesterol.

Ten were randomly selected to be a control group and the other ten were assigned to take part in the regimen as the treatment group for a period of 6 months.

The following data show the reduction in Choresterol(in mg/dL) experiences by 20 subjects during the study period:

Control group: 7, 3, -4, 14, 2, 5, 22, -9, 9, 5

Treatment group: -6, 5, 9, 4, 4, 12, 37, 5, 3, 3

a) Draw a dot plot of the data for both groups on the same graph.

b) Compute the mean, median, mode and 10% trimmed mean for both groups.

c) Explain why the difference in means suggests one conclusion about the effect of the regimen, while the difference in medians or trimmed means suggest a different conclusion.

d) calculate the sample variance as well as the standard deviation in tensile strength for both samples.

Md. Redoan Rahman Rifat
ID: 222 226 038

Answer to the Question No- 1

Given data after sorting in ascending order,

Control: -7, -4, 2, 3, 5, 5, 7, 9, 14, 22

Treatment: -6, 3, 3, 4, 4, 5, 5, 9, 12, 37

a)

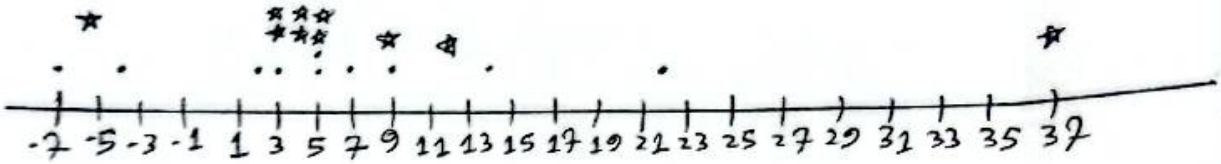


Figure: Dot plot for both groups where (•) indicates control group and (★) indicates treatment group.

b) Means:

∴ Control group; $\bar{x}_1 = \frac{-7-4+2+3+5+5+7+9+14+22}{10}$

$= 5.6$

∴ Treatment group, $\bar{x}_2 = \frac{-6+3+3+4+4+5+5+9+12+37}{10}$

$= 7.6$

Median:

$n = 10$ for both groups

$\frac{n}{2} = 5$

∴ median $= \frac{5th + 6th}{2}$ [since n is even]

Md. Redoan Rahman Rifat
ID: 222 226 038

$\therefore$ Control $= \frac{5+5}{2}$

$= 5$

$\therefore$ treatment $= \frac{4+5}{2}$

$= 4.5$

Mode:

$\therefore$ control $= 5$ [5 appeared the most]

$\therefore$ treatment $= 3, 4, 5$ [3, 4, 5 appeared the most]

10% trimmed mean:

10% of 10 $= 1$

After removing first and last values, we get,

control $= -4, 2, 3, 5, 5, 7, 9, 14$

$\therefore$ 10% trimmed mean $= \frac{-4+2+3+5+5+7+9+14}{8}$

$= 5.125$

treatment $= 3, 3, 4, 4, 5, 5, 9, 12$

$\therefore$ 10% trimmed mean $= \frac{3+3+4+4+5+5+9+12}{8}$

$= 5.625$

M2. Redoan Rahman Rifat
ID: 222 126 038

c) Difference in mean between the groups
$= (7.6 - 5.6)$
$= 2$

Difference in median between the groups
$= (5 - 4.5)$
$= 0.5$

Difference in 10% trimmed mean between the groups $= (5.625 - 5.125)$
$= 0.5$

Mean suggests treatment group has higher reduction and median and 10% trimmed mean are close between the groups. Treatment mean is inflated due to outlier which is 37.
∴ outlier makes the treatment group seem better.

d) control:

sample variance, $s_1^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_1)^2}{n-1}$

$= \frac{1}{10-1}\{(-7-5.6)^2 + (-4-5.6)^2 + (2-5.6)^2 + (3-5.6)^2 + (5-5.6)^2 + (5-5.6)^2 + (7-5.6)^2 + (9-5.6)^2 + (14-5.6)^2 + (22-5.6)^2\}$

$= \frac{1}{9}(624.4)$
$= 69.38$

∴ standard deviation, $s_1 = \sqrt{69.38}$
$= 8.33$

Md. Redoan Rahman Rifat
ID: 222 226 038

Treatment:

Sample variance, $s_2^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_2)^2}{n-1}$

$= \frac{1}{10-1}\{(6-7.6)^2 + (3-7.6)^2 + (3-7.6)^2 + (4-7.6)^2 + (4-7.6)^2 + (5-7.6)^2 + (5-7.6)^2 + (9-7.6)^2 + (12-7.6)^2 + (37-7.6)^2\}$

$= \frac{1}{9}(1152.4)$

$= 128.05$

$\therefore$ Standard deviation, $s_2 = \sqrt{128.05}$

$= 11.32$

Md. Fedoan Rahman Rifat
20.222.226.038

Question 2|

A certain polymer is used for evacuation systems in aircraft. It is important that the polymer be resistant to the aging process.

Twenty specimens of the polymer were used in an experiment. Ten were randomly assigned to be exposed to an accelerated aging process that involved exposures to high temperatures for 10 days.

measurement of tensile strength (in psi) were recorded for all specimens, as shown below:

no aging: 227, 222, 218, 217, 225, 218, 216, 229, 228, 221

aging: 229, 214, 215, 211, 209, 218, 203, 204, 201, 205

a) Draw a dot plot of the data.

b) From your plot, it appears as if the aging process has had an effect on the tensile strength of this polymer? Explain.

c) Calculate the sample mean tensile strength for both groups.

d) Calculate the median for both ~~groups~~ samples. Discuss the similarity or lack of similarity between the mean and median of each group.

e) Calculate the sample variance as well as the standard deviation in tensile strength for both samples.

Md. Rezoan Rahman Rifat
ID: 222 226 038

f) Does there appear to be any evidence that aging affects the variability in tensile strength?

Answer to the question No-2

Given data after sorting in ascending order,

No aging: 216, 219, 218, 218, 221, 222, 225, 227, 228, 229

Aging: 201, 203, 204, 205, 209, 211, 214, 215, 218, 219

a)

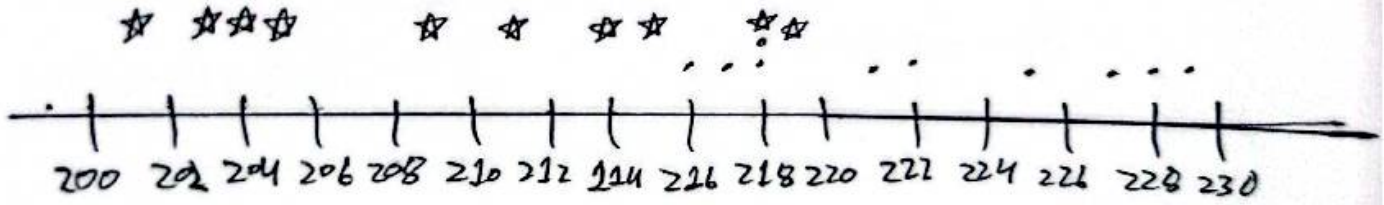


Figure: Dot plot for both groups where (·) indicates no aging group and (☆) indicates aging group.

Md. Redoan Rahman Rifat
ID: 222 226 038

b) The dot plot shows that the no aging sample values lie mostly between 216 - 229 and aging sample values are between 201 - 209.

Thus, aging group has lower tensile strength indicating that aging reduces the polymer's strength.

c) Means:

$$\text{No aging}, \bar{x}_1 = \frac{216+217+218+218+221+222+225+228+229}{10}$$

$$= 222.1$$

$$\text{Aging}, \bar{x}_2 = \frac{201+203+204+205+209+211+214+215+218+219}{10}$$

$$= 209.9$$

d) Medians:

$n = 10$; which is an even number

$\frac{n}{2} = 5$

$\therefore$ median $= \frac{5\text{th}+6\text{th}}{2}$

$\therefore$ No aging $= \frac{221+222}{2}$

$= 221.5$

$\therefore$ Aging $= \frac{209+211}{2}$

$= 210$

P.T.O

Md. Redoan Rahman Rifat
ID: 222 226 038

Both the mean and median shows the same pattern. The aging group has lower central tendency which indicates aging reduces tensile strength.

e) No aging:

sample variance, $s_1^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_1)^2}{n-1}$

$= \frac{1}{10-1}\{(216-222.1)^2+(217-222.1)^2+$
$(218-222.1)^2+(218-222.1)^2+(221-222.1)^2+(222-222.1)^2+$
$(225-222.1)^2+(227-222.1)^2+(228-222.1)^2+(229-222.1)^2\}$

$= \frac{1}{9}(212.9)$

$= 23.66$

∴ standard deviation, $s_1 = \sqrt{23.66}$

$= 4.86$

Aging:

sample variance, $s_2^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_2)^2}{n-1}$

$= \frac{1}{(10-1)}\{(201-209.9)^2+(203-209.9)^2$
$+(204-209.9)^2+(205-209.9)^2+(209-209.9)^2+(214-209.9)^2+$
$(215-209.9)^2+(218-209.9)^2+(219-209.9)^2+(211-209.9)^2\}$

$= \frac{1}{9}(378.9)$

$= 42.1$

∴ standard deviation, $s_2 = \sqrt{42.1}$

$= 6.48$

Md. Redoan Rahman Rifat
ID: 222 226 038

f) The no aging group has smaller sample variance and standard deviation while aging group has the larger ones. which means the aging samples are more variable.
Thus, aging lowers the average streagth and increases avariability.

CamScanner

Md. Redoan Rahman Rifat
ID: 222 226 038

Question: 3)

The following dataset gives the percentages of the families that are in the upper income level for the same individual school.

Data: 72.2, 31.9, 26.5, 29.1, 27.3, 8.6, 22.3, 26.5, 20.4, 12.8, 25.1, 19.2, 24.1, 58.2, 68.1, 89.2, 55.1, 9.4, 14.5, 13.9, 20.9, 17.9, 8.5, 55.4, 38.1, 54.2, 21.5, 26.2, 59.1, 43.3

a) Calcualate the sample mean.

b) Calculate the sample median.

c) Construct a relative frequency histogram of data

d) Compute the 10% trimmed mean. Compare with the results in (a) and (b) and comment.

CamScanner

Md. Fedoan Rahman Rifat
ID: 222 226 038

Answer to the Question No-3

Given data after sorting in ascending order,

8.5, 8.6, 9.4, 12.8, 13.9, 14.5, 17.9, 19.2, 20.4, 20.7, 21.5, 22.3, 24.1, 25.1, 26.2, 26.5, 26.5, 27.3, 29.1, 31.9, 38.1, 43.3, 54.2, 55.1, 55.4, 58.2, 59.1, 68.1, 72.2, 89.2

a) Mean, $\bar{x} = \frac{8.5+8.6+9.4+12.8+13.9+14.5+17.9+19.2+20.4+20.7+21.5+22.3+24.1+25.1+26.2+26.5+27.3+29.1+31.9+38.1+43.3+54.2+55.1+55.4+58.2+59.1+68.1+72.2+89.2+26.5}{30}$

$= 33.31$

b) $n = 30$, which is an even number

$\frac{n}{2} = 15$

$\therefore$ median $= \frac{15\text{th} + 16\text{th}}{2}$

$= \frac{26.2+26.5}{2}$

$= 26.35$

CamScanner

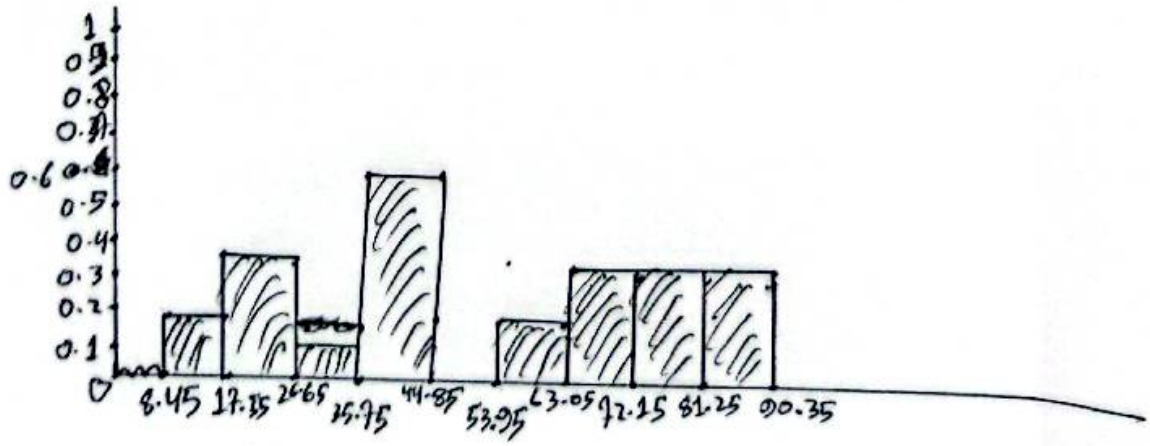
Md. Redoan Rahman Rifat
ID: 222 226 038

Q) let the class size = 9

$\therefore$ total class $= \frac{90.2 - 8.5}{9}$

$= 8.97 \approx 7$

Class	Continuous Class	Mid Point	Tally	Frequency (F)	Relative Frequency $(\frac{F}{50})$
8.5-17.5	8.45-17.55	13	𝍸 I	6	0.2
17.6-26.6	17.55-26.65	22.1	𝍸 𝍸 I	11	0.367
26.7-35.7	26.65-35.75	31.2	III	3	0.1
35.8-44.8	35.75-44.85	40.3	II	2	0.67
44.9-53.9	44.85-53.95	49.4		0	0
54.0-63.0	53.95-63.05	58.5	𝍸	5	0.167
63.1-72.1	63.05-72.15	67.6	I	1	0.33
72.2-81.2	72.15-81.25	76.7	I	1	0.33
81.3-90.3	81.25-90.35	90.6	I	1	0.33



Md. Redoan Rahman Rifat

ID: 222 226 038

d) 10% of 30 = 3

∴ After removing first 3 and last 3 values, we get,
12.8, 13.9, 14.5, 17.9, 19.2, 20.4, 20.7, 21.5, 22.3, 24.1,
25.1, 26.2, 26.5, 26.5, 27.3, 29.1, 31.9, 38.1, 43.3,
54.2, 55.1, 55.4, 58.2, 59.1

∴ 10% trimmed mean =

$$\frac{12.8+13.9+14.5+17.9+19.2+20.4+20.7+21.5+22.3+24.1+25.1+26.2+26.5+26.5+27.3+29.1+31.9+38.1+43.3+54.2+55.1+55.4+58.2+59.1}{24}$$

$$= 30.07$$

median < 10% trimmed mean < mean

Trimmed mean lies between median and mean, showing a more balanced central value because extreme values were removed.

Md. Redoan Rahman Rifat
ID: 222 226 03B

Question: 4)

Following data were obtained on the age and blood glucose (BGL) collected from 6 individuals.

AGE: 43 21 25 42 57 59
BGL: 99 65 79 75 87 81

Identify the dependent and independent variables. Fit a simple linear regression model. Interpret the results, also,

a) Predict the value of BGL when the age of an individual is 45 years.

b) Compare the BGL between individuals of ages

(i) 35 years and 30 years.

(ii) 40 years and 48 years.

Answer to the Question No-4

In the given problem,

Independent variable, X = AGE

Dependent variable, Y = BGL

Independent and dependent variable are continuous here.

∴ predicted value of $\hat{y} = \hat{\alpha} + \hat{\beta}X$

CamScanner

Md. Redoun Rahman Rifat
ID: 222226034

where,
$\hat{\alpha}$ = intercept
$\hat{\beta}$ = slope.

Calculation of regression parameters:

x	y	x^2	y^2	xy
43	99	1849	9801	4257
21	65	441	4225	1365
25	79	625	6241	1075
42	75	1764	5625	3150
57	87	3249	7529	4959
59	81	3481	6561	4779
Σ 247	486	11409	40022	20485

Number of observations, $n = 6$

we know,

$$\bar{x} = \frac{\Sigma x}{6}$$
$$= \frac{247}{6}$$
$$= 41.17$$

$$\bar{y} = \frac{\Sigma y}{6}$$
$$= \frac{486}{6}$$
$$= 81$$

$$\hat{\beta} = \frac{\Sigma xy - n \cdot \bar{x} \cdot \bar{y}}{\Sigma x^2 - n \cdot \bar{x}^2}$$
$$= \frac{20485 - 6 \cdot (41.17) \cdot 81}{11409 - 6 \cdot (41.17)^2}$$
$$= 0.384$$

Md. Redoan Rahman Rifat
ID: 222226038

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
$= 81 - (0.384) \cdot (41.17)$
$= 65.19$

∴ A fitted simple linear regression model,
$\hat{y} = 65.19 + 0.384x$

In the absence of independent variable, the mean BCrC of individuals is approximately 65.19. For one unit increase in age of an individual, the mean BCrC increased by 0.384 units approximately. which indicates a positive correlation between BCrC and Age.

a) Predicted value of BCrC when age of an individual is 45 years, $\hat{y} = 65.19 + (0.384 \times 45)$
$= 82.47$ units.

b) i) The mean BCrC of an individuals of age 35 years is approximately $(0.384 \times 35 - 30) = 1.92$ units higher than that of individual of age 30 years.

ii) The mean BCrC of an individual of age 40 years is approximately $0.384 \times (48 - 40) = 3.072$ units lower than that of individual of age 48 years.

Md. Redoan Rahman Rifat
ID: 222 226 038

Question: 5)

12 customer satisfaction scales are given below:

14, 27, 17, 19, 32, 42, 49, 23, 60, 54, 50, 71

a) Find the 5-number summary.

b) Find P_{20} and P_{90}

CamScanner

Md. Redoan Rahman Rifat
ID: 222 226 038

Answer to the Question No-5

Given data after sorting in ascending order,

14, 19, 19, 23, 27, 32, 42, 49, 54, 59, 71, 80.

here, $n = 12$, which is an even number.

a) $Q_1 = P_{25} = \frac{25}{100} \times 12$

$= 3$

$\therefore Q_1 = \frac{3rd + 4th}{2}$

$= \frac{19 + 23}{2}$

$= 21$

$Q_2 = P_{50} = \frac{50}{100} \times 12$

$= 6$

$\therefore Q_2 = \frac{6th + 7th}{2}$

$= \frac{32 + 42}{2}$

$= 37$

$Q_3 = P_{75} = \frac{75}{100} \times 12$

$= 9$

$\therefore Q_3 = \frac{9th + 10th}{2}$

$= \frac{54 + 59}{2}$

$= 56.5$

Md. Redoan Rahman Rifat
ID: 222 226 038

∴ Five Number Summary,

Minimum = 12

First quartile, $Q_1 = 21$

Second quartile, $Q_{2\&} = 37$

Third quartile, $Q_3 = 56.5$

Maximum = 80

b) $P_{20} = u = \frac{20}{100} \times 12$

$= 2.4 \approx 3$

∴ $P_{20} = 3rd = 19$

$P_{90} = u = \frac{90}{100} \times 12$

$= 10.8 \approx 11$

∴ $P_{90} = 11th = 71$.

Md. Redoan Rahman Rifat
ID: 222 226 036

Question: 6

a) What do you mean by outlier, skewness and kurtosis of a dataset.

b) Write down the classification with conditions of skewness and kurtosis.

c) For a distribution Karl Pearson's coefficient of skewness is 0.64, standard deviation is 13 and mean is 59.2. Find mode and median.

d) The first four moments about mean of distribution are 0, 2.5, 0.7, 18.75. Find coefficient of skewness and kurtosis.

Answer to the question No-6

a) Outlier: A data value that lies far away from most other observation in a dataset.

skewness: A measure that shows whether data are spread more to one side. Right-skewed if tail is on the right side, left-skewed if tail is on the left side.

kurtosis: A measure of how peaked or flat a distribution is compared to a normal curve.

CamScanner

Md. Raiboon Rahman Rifat
ID: 222 226 036

b) Skewness = Sk

if Sk = 0, then curve = symmetric skewness
if Sk > 0, then curve = positive skewness
if Sk < 0, then curve = Negative skewness

Kurtosis = β_2
if $\beta_2 = 3$, then curve = Mesokurtic
if $\beta_2 > 3$, then curve = Lepokurtic
if $\beta_2 < 3$, then curve = platykurtic

c) Given that,
skewness, Sk = 0.64
standard deviation, $\sigma = 13$
mean = 59.2

we know,
$$Sk = \frac{\text{mean} - \text{mode}}{\sigma}$$

$$\Rightarrow \text{mode} = \text{mean} - Sk \cdot \sigma$$
$$= 59.2 - (13 \times 0.64)$$
$$= 50.88$$

Again,
$$Sk = \frac{3(\text{mean} - \text{median})}{\sigma}$$

$$\Rightarrow \text{median} = \text{mean} - \left(\frac{Sk \cdot \sigma}{3}\right)$$
$$= 59.2 - \left(\frac{13 \times 0.64}{3}\right)$$
$$= 56.42$$

Md. Rezoan Rahman Rifat
ID: 222 226 038

d) Given that,

4 moments = 0, 2.5, 0.7, 18.75

Skewness, $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

$= \frac{(0.7)^2}{(2.5)^3}$

$= 0.03$

Kurtosis, $\beta_2 = \frac{\mu_4}{\mu_2^2}$

$= \frac{18.75}{(2.5)^2}$

$= 3$

CamScanner