Applied Mathematics

DATE : ....

ASSIGNMENT - 3

Syeda Reeha Ansar
1411480d719
4C7

Details

1) By the method of least square fit a parabola from the following data.

x	1	2	3	4	5
y	2	6	4	5	2

Ans: the eq^n for parabola is : $y = a + bx + cx^2$

So normal equations :

$$\sum y_i = na + b\sum x_i + c\sum x_i^2 \quad -(1)$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3 \quad -(2)$$

$$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \quad -(3)$$

where $n = 5$

x	y	x_i^2	x_i^3	x_i^4	$x_i y$	$x_i^2 y$
1	2	1	1	1	2	2
2	6	4	8	16	72	24
3	4	9	27	81	12	36
4	5	16	64	256	20	80
5	2	25	125	625	10	50
$\sum x_i = 15$	$\sum y_i = 19$	$\sum x_i^2 = 55$	$\sum x_i^3 = 225$	$\sum x_i^4 = 956$	$\sum x_i y = 56$	$\sum x_i^2 y = 192$

DATE

putting these values in ①, ② and ③

$19 = 5a + 15b + 55c$ — ④

$56 = 15a + 55b + 225c$ — ⑤

$192 = 55a + 225b + 979c$ — ⑥

solving these eqn

① × 3 ②

$$57 = 15a + 45b + 165c$$
$$56 = 15a + 55b + 225c$$
$$1 = -10b - 60c$$

$\Rightarrow 10b + 60c = 1$ — ⑦

① × 11 ③

$$192 = 55a + 225b + 979c$$
$$209 = 55a + 165b + 605c$$
$$-17 = 60b + 374c$$

$-17 = 60b + 274c$ — ⑧

solving ⑦ and ⑧

$$60b + 360c = -6$$
$$60b + 374c = -17$$
$$-14c = 11$$

$C = -11/14 = -0.79$

solving further $a = -1.4$, $b = 4.6$

$C = -0.79$

therefore the parabola that will fit in

$y = a + bx + Cx^2$

$\Rightarrow \boxed{y = -1.4 + 4.61x - 0.79x^2}$

2) Two random variables have the regression lines with eqn $3x + 2y = 26$ and $6x + y = 31$. Find the mean values of x and y

DATE : ....

Also find correlation coefficient between x and y.

$$3x + 2y = 26 \quad —(1)$$
$$6x + y = 31 \quad —(2)$$

Solving (1) and (2)

$$\cancel{6x} + 4y = 52$$
$$\cancel{6x} + y = 31$$
$$- \quad - \quad -$$

$$3y = 21$$

$$\boxed{y = 7}$$

Substituting value of y in (1)

$$3x - 2(7) = 26$$
$$3x = 12$$

$$\boxed{x = 4}$$

since, the point of regression of lines is intersection point of are two line

$$\therefore \boxed{\bar{x} = 4 \;,\; \bar{y} = 7} \quad \text{mean of } x \text{ and } y$$

Correlation coefficients

Let $3x + 2y = 26$ be the regression eq^n of y on x

$\therefore$ The eq^n becomes $2y = -3x + 26$

$$\Rightarrow y = -\frac{3}{2} x + \frac{26}{2}$$

Comparing eq^n $y = b_{yx} x + a$

$$b_{yx} = -3/2$$

Now let $6x + y = 31$ be the regression equation of x on y

$\therefore$ The eq^n becomes : $6x = -y + 31$

$$x = -\frac{1}{6} y + \frac{31}{6}$$

DATE

Comparing this with $x = b_{xy} y + a$

$b_{xy} = -1/6$

$\therefore r = \pm\sqrt{b_{xy} b_{yx}} = \pm\sqrt{(-1/6)(-3/2)}$

$= \pm\sqrt{1/4} = \pm 1/2$

Since the value of b_{yx} and b_{xy} are $(-ve)$ r will be also $-ve$

$$\boxed{r = -1/2 = -0.5}$$

3) If θ is the acute angle between two regression lines. Show that

$$\tan\theta = \frac{1-r^2}{r}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right) \text{ where}$$

σ_x, σ_y are the SDs of x and y series respectively and r is the correlation coefficient

Sol^n Lines of regression y on x:

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

Now, to find slope $= y = r\frac{\sigma_y}{\sigma_x}\left(x + \frac{\bar{y}\sigma_x}{\sigma_y}\right)$

Slope of $\boxed{m = r\frac{\sigma_y}{\sigma_x}}$

Line of regression x on y

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

Slope $m = \frac{1}{r}\left(\frac{\sigma_y}{\sigma_x}\right)$

Now θ is the angle b/w regression lines

$$\tan\theta = \frac{\left(\frac{1}{r} - r\right) \cdot (\sigma y / \sigma x)}{1 + \frac{\sigma^2 y}{\sigma x^2}}$$

$$\tan\theta = \frac{(1 - r^2 / r) \cdot \sigma y / \sigma x \, (\sigma^2 x)}{(\sigma x^2 + \sigma^2 y)}$$

$$\boxed{\tan\theta = \left(\frac{1 - r^2}{r}\right)\left(\frac{\sigma y \, \sigma x}{\sigma x^2 + \sigma y^2}\right)}$$

4) The equations of two lines of regression are $4x + 3y + 7 = 0$ & $3x + 4y + 8 = 0$. Find the regression coefficients b_{yx}, b_{xy} & the correlation coefficient r. Also, find the standard deviation of y, if variance of x is 4.

Soln

$4x + 3y + 7 = 0$ — ①

$3x + 4y + 8 = 0$ — ②

Let ① be regression eqn. of y on x

Let ② be regression eqn. of x on y.

from ① $y = -4/3 x - 7/3$

comparing with $y = b_{yx} x + a$ $\boxed{b_{yx} = -4/3}$

from ② $x = -4/3 y - 8/3$

comparing with $x = b_{xy} y + a'$ $\Rightarrow$ $\boxed{b_{xy} = -4/3}$

$\Rightarrow$ But $r^2 = b_{xy} \cdot b_{yx} \Rightarrow (-4/3)(-4/3)$ which is > 1

This is impossible.

Hence eqn. ① is regression line of x on y

& eqn ② is regression line of y on x.

so, now from ①.

$$x = -\frac{3}{4}y - \frac{7}{4} \quad \Rightarrow \quad b_{xy} = -3/4$$

from ② $\quad y = -3/4\, x - 8/4 \quad \Rightarrow b_{yx} = -3/4$

Also $r^2 = b_{xy} \cdot b_{yx}$

$$r = \pm\sqrt{\left(-\frac{3}{4}\right)\left(-\frac{3}{4}\right)} = \pm\frac{3}{4}$$

As both b_{xy} & b_{yx} are -ve, so r will also be (-ve)

$$r = -3/4$$

standard deviation of Y

$b_{xy} = r\frac{\sigma_x}{\sigma_y}$ (given the $\sigma^2_x = 4$, $\sigma_x = 2$)

$$\Rightarrow -\frac{3}{4} = \left(-\frac{3}{4}\right)\left(\frac{2}{\sigma_y}\right) \Rightarrow \boxed{\sigma_y = 2}$$

5) calculate the rank correlation coefficient from foll. data showing results ranks of 5 students in 2 subjects

Maths	3	2	4	1	5
Chem.	5	4	3	2	1

Soln

Rank correlation coefficient: $\frac{6\Sigma d^2}{N(N^2-1)}$

Maths	(Mm) Rank	Chem.	(Ch) Rank	|Mm-Ch| d	d^2
3	3	5	1	2	4
2	4	4	2	2	4
4	2	3	3	1	1
1	5	2	4	1	1
5	1	1	5	4	16

$$\Sigma d^2 = 26$$

These ranks are not tied, so

$$r = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 26}{5(25-1)} = 1 - \left(\frac{6 \times 26}{5 \times 24}\right) \Rightarrow 1 - \frac{13}{10}$$

$$= 1 - 1.3$$

$$r = -0.3$$

6) The mean of two large samples of 1000 & 2000 members are 168.75 cm & 170 cm. respectively. can the samples be regarded as drawn from the same population of standard deviation 6.25 cm?

Soln. To test the Null hypothesis:

$H_0 \Rightarrow \mu_1 - \mu_2 = 0$ (means are equal)

$H_1 \Rightarrow \mu_1 - \mu_2 \neq 0$ at significance level α

If $\alpha = 0.05 \Rightarrow Z_{\alpha/2} = 1.96$

$\therefore -1.96 < z < 1.96 \Rightarrow$ the acceptance region

$\Rightarrow$ If $\alpha = 0.01 \Rightarrow Z_{\alpha/2} = 2.575$

$\therefore -2.575 < z < 2.575 \Rightarrow$ the acceptance region

Here $n_1 = 1000$, $\bar{x}_1 = 168.75$, $\sigma_1 = 6.25$

$n_2 = 2000$, $\bar{x}_2 = 170$, $\sigma_2 = 6.25$

The test statistic,

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(168.75 - 170) - 0}{\sqrt{\frac{(6.25)^2}{1000} + \frac{(6.25)^2}{2000}}}$$

$z \Rightarrow -5.165$

$\Rightarrow$ Both cases are rejected.

Therefore, samples are not from same population

7) A sample of 9 items has the foll. items values → 45, 47, 50, 52, 48, 49, 47, 53, 51. Does the mean of sample differ significantly from the population mean 47.5?

Sol. ⇒ let $\mu_0 : \mu_0 = 47.5$.

⇒ let $\mu_1 : \mu \neq 47.5$.

Given: $n = 9$, $\mu = 47.5$

x	45	47	50	52	48	47	49	53	51	$\Sigma x = 442$
$x - \bar{x}$	-4.1	-2.1	0.9	2.9	-1.1	-2.1	-0.1	3.9	1.9	
$(x-\bar{x})^2$	16.81	4.41	0.81	8.41	1.21	4.41	0.01	15.21	3.61	$\Sigma = 54.8$

$\bar{x} = \frac{\Sigma x}{n} = \frac{442}{9} = 49.11$

⇒ $\Sigma(x-\bar{x})^2 = 54.89$

⇒ $s^2 = \frac{\Sigma(x-x^2)}{n-1}$ ⇒ $\frac{54.89}{9-1}$ ⇒ $\boxed{s^2 = 6.86}$

Applying the t-test

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{49.1 - 47.5}{2.61} \times \sqrt{8} = \frac{1.6\sqrt{8}}{2.61}$

$t = 1.7279$

$t_{0.05} = 2.31$ for $r = 8$

Since $|t| < t_{0.05}$ the hypothesis is accepted. so

⇒ there is no significance difference in their mean.

8) The theory predicts population of beans in four groups A, B, C & D should be 9:3:3:1. In an experiment among 1600 beans, the no. in four groups were 882, 313, 287 & 118. Does the experimental result support the theory?

Soln-

To set up the null hypothesis:

H_0: The result supports the theory.

Total no. of beans ⇒ 882 + 313 + 287 + 118 = 1600

Now, to calculate the expected frequencies.

These are divided into the ratio ⇒ 9:3:3:1

Thus,

$$E(882) = \frac{9}{16} \times 1600 = 900$$

$$E(313) = \frac{3}{16} \times 1600 = 300$$

$$E(287) = \frac{3}{16} \times 1600 = 300$$

$$E(118) = \frac{1}{16} \times 1600 = 100$$

Hence,

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(882-900)^2}{900} + \frac{(313-300)^2}{300} + \frac{(287-300)^2}{300} + \frac{(118-100)^2}{100}$$

$$= 0.3600 + 0.5633 + 0.5633 + 3.2400$$

$$= 4.7266$$

The degree of freedom: 3 ⇒ $\chi^2_{0.05} = 7.815$

Now, as the calculated value of χ^2 (4.7266) is less than the tabulated χ^2, the NULL hypothesis is accepted.

Therefore, YES, experimental result supports the theory.

9) A survey of 800 families with four children, eac recorded the foll. data.

No. of boys	0	1	2	3	4
No. of girls	4	3	2	1	0
No. of families	32	178	290	236	64

Test the hypothesis that male & female births are equally likely.

Soln. $P(\text{all boys}) = \frac{1}{2^4} = \frac{1}{16}$

$P(3\text{ boys}, 1\text{ girl}) = {}^4C_3 \frac{1}{2^3} \times \frac{1}{2^1} = \frac{1}{4}$

$P(2\text{ boys}, 2\text{ girls}) = {}^4C_2 = \frac{1}{2^2} \times \frac{1}{2^2} = \frac{3}{8}$

$P(1\text{ boy}, 3\text{ girls}) = {}^4C_1 \frac{1}{2} \times \frac{1}{2^3} = \frac{1}{4}$

$P(\text{all girls}) = {}^4C_0 \frac{1}{2^4} = \frac{1}{16}$

Let:
H_0 : Male & female births are equally likely.

No. of families with

all boys $= \frac{1}{16} \times 800 = 50$

3 boys $= \frac{1}{4} \times 800 = 200$

2 boys $= \frac{3}{8} \times 800 = 300$

1 boy $= \frac{1}{4} \times 800 = 200$

no boy $= \frac{1}{16} \times 800 = 50$

Observed frequency (O_i)	Expected frequency (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
32	50	324	6.48
178	200	484	2.42
290	300	100	0.333
236	200	1296	6.48
64	50	196	3.92

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 19.633$$

The value of χ^2 at 5% level of significance for $5-1=4 \Rightarrow$ difference is 9.49.

$\Rightarrow$ Since the calculated value of χ^2 is greater than the tabulated value, H_0 is rejected.

Hence, the hypothesis that male & female births are equally likely is rejected.

10) Two random samples from the normal populations are given below? Do the estimates of population variance differ significantly?

Sample 1	16	26	27	23	24	22
Sample 2	33	42	35	32	28	31

$F_{0.05}(5,5) = 5.05$ $\quad F_{0.05}(5,6) = 4.39$

$F_{0.05}(6,5) = 4.95$

DATE

Sol: Let

H_0: The samples are of equal means & hence the population variance doesn't differ significantly.

H_0, $\mu_1 = \mu_2$

H_1 : $\mu_1 \neq \mu_2$ (Two tailed test)

$n_1 = 6$ $\qquad$ $n_2 = 6$

$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{6} x_i \Rightarrow \frac{1}{6} \times 138 \Rightarrow \bar{x}_1 = 23$

$\bar{x}_2 = \frac{1}{n_2} \sum_{i=1}^{6} x_i \Rightarrow \frac{1}{6} \times 201 \Rightarrow \bar{x}_2 = 33.5$

x_1	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	x_2	$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$
16	-7	49	33	-0.5	0.25
26	3	9	42	8.5	72.25
27	4	16	35	1.5	2.25
23	0	0	32	-1.5	2.25
24	1	1	28	-5.5	30.25
22	-1	1	31	-2.5	6.25
$\sum x_1 = 138$					$\sum(x_2 - \bar{x}_2)^2 = 113.5$

$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{23 - 33.5}{S\sqrt{\frac{1}{6} + \frac{1}{6}}} = \frac{-10.5}{S\sqrt{1/3}}$

$S^2 = \frac{1}{n_1 + n_2 - 2} \left[\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2 \right]$

$S^2 = \frac{1}{6+6-2} [76 + 113.5] = \frac{1}{10} \times 189.5$

$S^2 = 18.95$

$S = 4.35$

DATE : ....

from ①

$$t = \frac{10.5}{5\sqrt{6.33}} = \frac{-10.5}{4.35(0.57)}$$

$$\boxed{t = 4.23}$$

given at significant level 5% at dof 5 5

$F_{0.05}(5, 5) = 5.05$ and etc

As the value of $|t|$ $[0.01]$ of tabulated value, the hypothesis is accepted so no the population variance doesn't differ significantly.