Statistics
Unit-3

* ~~[illegible]~~ practice questions.

Q1) First four moments of a distribution about the value 5 are, 2, 20, 40, 50, From the given information obtain first 4 Central moments, Coefficient of Skewness and kurtosis

→ Given: first four moments about point (5):

$\mu_1' = 2, \; \mu_2' = 20, \; \mu_3' = 40, \; \mu_4' = 50$

To find:

i) First four Central moments ($\mu_1, \mu_2, \mu_3, \mu_4$)

ii) Coefficient of Skewness & Kurtosis

Soln:

Origin $A = 5$, and the mean $= \bar{x} = A + \mu_1'$

$\bar{x} = 5 + 2 = 7$

$\mu_1 = 0$ (by definition)

$\mu_2 = \mu_2' - (\mu_1')^2 = 20 - (2)^2 = 16$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$

$= 40 - 3(20)(2) + 2(2)^3$

$= 40 - 120 + 16$

$= -64$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$
$$= 50 - 4(40)(2) + 6(20)(4) - 3(2)^4$$
$$= 162$$

Now,
Coefficient of skewness (β_1):

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-64)^2}{(16)^3} = \frac{4096}{4096} =$$

Coefficient of kurtosis (β_2):

$$\beta_2 : \frac{\mu_4}{\mu_2^2} = \frac{162}{(16)^2} = 0.6328$$

$\therefore$

$\mu_1 = 0$, $\mu_2 = 16$, $\mu_3 = -64$, $\mu_4 = 162$
$\beta_1 = 1$, $\beta_2 = 0.6328$

Q2> Obtain the regression lines y on x and x on y for the data.

x	5	1	10	3	9
y	10	11	15	10	6

→

x	y	x^2	y^2	xy
5	10	25	100	50
1	11	1	121	11
10	15	100	225	150
3	10	9	100	30
9	6	81	36	54

$\Sigma x = 28$ $\Sigma y = 52$
$\Sigma x^2 = 216$ $\Sigma y^2 = 582$
$\Sigma xy = 275$
$N = 5$

i) Line of regression y on x

$\Sigma y = aN + b\Sigma x$
$\Sigma xy = a\Sigma x + b\Sigma x^2$

$\therefore$ $52 = 5a + 28b$ —— ①
$275 = 28a + 216b$ —— ②
Solving eq① & eq②

$a = 11.9324$ $b = -0.2736$
$\therefore$ y on x
$\therefore$ $y = a + bx$
$\boxed{y = 11.9324 - 0.2736x}$

ii) Line of regression x on y

$\Sigma x = aN + b\Sigma y$
$\Sigma xy = a\Sigma y + b\Sigma y^2$

$\therefore$ $28 = 5a + 52b$ —— ①
$275 = 52a + 582b$ —— ②
Solving eq① & eq②
$a = 9.6893$ $b = -0.3932$

$\therefore$ $x = a + by$
$\therefore$ $\boxed{x = 9.6893 - 0.3932y}$

Q.3) Calculate Standard deviation for the following frequency distribution. Decide whether arithmetic mean is good or not.

Wages (per day)	0-10	10-20	20-30	30-40	40-50	50-60
No. of labor	5	9	15	12	10	3

Class	x	f	fx	$f(x-\bar{x})^2/N$
0-10	5	5	25	54897.8629
10-20	15	9	135	91266.17333
20-30	25	15	375	140026.9889
30-40	35	12	420	102754.9511
40-50	45	10	450	78240 25926
50-60	55	3	165	21355.41778

$$\text{mean } (\bar{x}) = \frac{\Sigma fx}{N} = \frac{1570}{6} = 261.666$$

$$\text{S.D } (\sigma) = \sqrt{\sum_{i=1}^{n} \frac{f(x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{488541.631}{6}}$$

$$= 285.3482$$

Class	x	f	fx	$f(x-\bar{x})^2/N$	
0-10	5	5	25	~~182.9695~~	53.662
10-20	15	9	135	~~297.1162~~	33.012
20-30	25	15	375	~~41.84936~~	4.610
30-40	35	12	420	~~70.2349~~	7.803
40-50	45	10	450	~~422.7291~~	46.969
50-60	55	3	165	~~336.0787~~	37.342

$$\text{mean } (\bar{x}) = \frac{\Sigma fx}{N} = \frac{1570}{54} = 29.074$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\sum_{i=1}^{n} \frac{f(x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{183.398}{54}} = 1.8408$$

Class	x	f	fx	fx^2	
0-10	5	5	25	~~625~~	125
10-20	15	9	135	~~18225~~	2025
20-30	25	15	375	~~140625~~	9375
30-40	35	12	420	~~176400~~	14700
40-50	45	10	450	~~202500~~	20250
50-60	55	3	165	~~27225~~	9075

$$\text{mean}(\bar{x}) = \frac{\Sigma fx}{N} = \frac{1570}{54} = 29.074$$

$$\text{S.D } (\sigma) = \sqrt{\frac{\Sigma fx^2}{N} - (\bar{x})^2}$$

$$= \sqrt{\frac{\cancel{55610}\ 55550}{54} - (29.074)^2}$$

$$\boxed{= 13.5427}$$

Since the Standard deviation is still high Compared to the mean, Arithmetic mean is not a good representative, as there's significant spread in the data.

Q4) following are the values of import of raw material and export of finished products in Suitable units.

Export	10	11	14	14	20	22	16	12	15	13
import	12	14	15	16	21	26	21	15	16	14

Calculate the Coefficient of Correlation between the import values & export values.

→

x	y	x^2	y^2	xy
10	12	100	144	120
11	14	121	196	154
14	15	196	225	210
14	16	196	256	224
20	21	400	441	420
22	26	484	676	572
16	21	256	441	336
12	15	144	225	180
15	16	225	256	240
13	14	169	196	182

$\Sigma x = 147 \quad \Sigma y = 170 \quad n = 10$

$\bar{x} = \dfrac{\Sigma x}{n} = 14.7 \quad \bar{y} = \dfrac{\Sigma y}{n} = 17$

$\Sigma xy = 2638 \quad \Sigma x^2 = 2291 \quad \Sigma y^2 = 3056$

$$\mathrm{Cov}(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\cdot\bar{y}$$

$$= \frac{1}{10} \times 2638 - 14.7 \times 17$$

$$= 13.9$$

Coefficient of Correlation $(r) = \dfrac{\mathrm{Cov}(x,y)}{\sigma x \cdot \sigma y}$

$$\sigma x = \sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2} = \sqrt{\frac{1}{10} \times 2291^2 - (14.7)^2}$$

$$= 724.3286$$

$$\sigma y = \sqrt{\frac{1}{10} \times (3056)^2 - (17)^2}$$

$$= 966.2425$$

$$r = \frac{13.9}{724.3286 \times 966.2425}$$

$=$

$$\sigma x = \sqrt{\frac{1}{n} \times \Sigma x^2 - \bar{x}^2}$$

$$= \sqrt{\frac{1}{10} \times 2291 - 14.7^2} = 3.6069$$

$$\sigma y = \sqrt{\frac{1}{10} \times 3056 - 17^2} = 4.0743$$

Coeff. of Correlation $(r) = \frac{Cov(x,y)}{\sigma x \cdot \sigma y}$

$$= \frac{13.9}{(3.6069)(4.743)}$$

$$\boxed{= 0.81250}$$

Q5) If the two lines of regression are $9x + y - \lambda = 0$ and $4x + y = \mu$ and the means of x and y are 2 & -3 respe find the values of λ, μ and Coefficient of correlation between x & y

→ Given:

$$9x + y - \lambda = 0 \quad \text{--- (1)}$$
$$4x + y = \mu \quad \text{--- (2)}$$
$$\bar{x} = 2, \; \bar{y} = -3$$

$9(2) + (-3) - \lambda = 0$

$18 - 3 = \lambda$

$\boxed{\lambda = 15}$

$4(2) + (-3) = \mu$

$\boxed{\mu = 5}$

$9x + y - 15 = 0$

$y = -9x + 15$

$b_{yx} = -9$

$4x + y = 5$

$4x = -y + 5$

$x = -\frac{y}{4} + \frac{5}{4}$

$b_{xy} = -\frac{1}{4} = -0.25$

$r = \sqrt{b_{yx} \cdot b_{xy}}$

$= \sqrt{(-9) \times (-0.25)}$

$= 1.5$

This means our assumption violet the rule that: $|r| \leq 1$

$\therefore$ such regression coefficients are not possible because the product $b_{yx} \cdot b_{xy}$ must be ≤ 1

Q6> Compute the first four moments about arbitary mean A=25 for the following frequencies. Q7>

No. of jobs	0-10	10-20	20-30	30-40	40-50
No. of workers	6	26	47	15	6

→ A=25

Class	f	x	d=x-A	fd	fd^2	fd^3	fd^4
0-10	6	5	-20	-120	2400	-48000	960000
10-20	26	15	-10	-260	2600	-26000	260000
20-30	47	25	0	0	0	0	0
30-40	15	35	10	150	1500	15000	150000
40-50	6	45	20	120	2400	48000	960000
Total	100			-110	8900	-11000	2330000

N = 100

$$\mu_1' = \frac{\Sigma fd}{N} = \frac{-110}{100} = -1.1$$

$$\mu_2' = \frac{\Sigma fd^2}{N} = \frac{8900}{100} = 89$$

$$\mu_3' = \frac{\Sigma fd^3}{N} = \frac{-11000}{100} = 110$$

$$\mu_4' = \frac{\Sigma fd^4}{N} = \frac{2330000}{100} = 23300$$

may-june-22

Q7) Calculate mean and standard deviation for the following table giving the age distribution of 542 members.

→

Age	f	x	fx	fx^2
20-30	5	25		
30-40	61	35		
40-50	132	45		
50-60	153	55		
60-70	140	65		
70-80	51	75		
80-90	2	85		
Total:	542		29710	1700800

$$\therefore \text{Mean}(\bar{x}) = \frac{\Sigma fx}{N} = \frac{\cancel{542}\ 29710}{542} = 54.81$$

$$S.D.(\sigma) = \sqrt{\frac{\Sigma fx^2}{N} - (\bar{x})^2}$$

$$= \sqrt{\frac{1700800}{542} - (54.81)^2}$$

$$= \underline{\underline{11.57}}$$

30

Q.8) In a partially destroyed laboratory, record of an analysis of Correlation data, the following results are legible. Variance of $x = 9$. Regression equations $8x - 10y + 66 = 0$, $40x - 18y = 214$. What are

i) the means of values x and y

ii) the Correlation Coefficient betn x & y

iii) The standard deviation of y.

$8x - 10y + 66 = 0$ —— ①

$40x - 8y - 214 = 0$ —— ②

eq① × 5

$40x - 50y = -330$ —— ①

$40x - 8y = 214$ —— ②

solving eq① & eq②

$\boxed{\bar{x} = 7.9404}$ $\boxed{\bar{y} = 12.95}$

variance of $x = 9$

$\therefore$ S.D. of x $(\sigma_x) = \sqrt{9} = 3$

$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2}$

$3 = \sqrt{}$

$8x - 10y + 66 = 0$ —— Eq①
$40x - 18y = 214$ —— Eq②

multiply eq① by 5

$40x - 50y = -330$ —— Eq①
$40x - 18y = 214$ —— Eq②

Solving eq① & eq②

$\boxed{\bar{x} = 13}$, $\boxed{\bar{y} = 17}$ ..... (means of x & y)

$8x - 10y + 66 = 0$

$8x = 10y - 66$

$x = \frac{10}{8}y - \frac{66}{8}$

$\underline{b_{xy}} = \frac{10}{8} = 1.25$

$40x - 18y = 214$

$40x - 214 = 18y$

$y = \frac{40}{18}x - \frac{214}{18}$ = 2.222

Coeff. of Correlation (r) $= \sqrt{b_{xy} \cdot b_{yx}}$
$= \sqrt{1.25 \times 2.222}$
$= 1.666$

But this is wrong, r must lie betn -1 & 1
This implies wrong assignment of regression directions, lets reverse them

$8x - 10y + 66 = 0$

$10y = 8x + 66$

$y = \frac{8}{10}x + \frac{66}{10}$

$\underline{byx} = \frac{8}{10} = 0.8$

$40x - 18y = 214$

$40x = 214 + 18y$

$x = \frac{18}{40}y + \frac{214}{40}$

$\underline{bxy} = \frac{18}{40} = 0.45$

$\therefore$ Coeff. of Correlation $(r) = \sqrt{0.8 \times 0.45}$

$r = \boxed{0.6}$

S.D. of y

we have

$r = 0.6$

$\sigma x = 3$

$byx = r \cdot \frac{\sigma y}{\sigma x} = 0.8$

$\therefore 0.8 = 0.6 \times \frac{\sigma y}{3}$

$\therefore \sigma y = \frac{0.8 \times 3}{0.6}$

$\boxed{\sigma y = 4}$

Q9) For
the

→ x | 1
y | 2

Deti
and
non

x	y
1	2
1	7
2	7
2	10
3	8
3	12
4	10
5	14
6	11
7	14
Σ 34	95

fo
T

S

Q9) For 10 randomly selected observations the Following data were recorded.

x	1	1	2	2	3	3	4	5	6	7
y	2	7	7	10	8	12	10	14	11	14

Determine the Coefficient of regression and regression equation using the non-linear from $y = a + b_1x + b_2x^2$.

	x	y	x^2	x^3	x^4	xy	x^2y
	1	2	1	1	1	2	2
	1	7	1	1	1	7	7
	2	7	4	8	16	14	28
	2	10	4	8	16	20	40
	3	8	9	27	81	24	72
	3	12	9	27	81	36	108
	4	10	16	64	256	40	160
	5	14	25	125	625	70	350
	6	11	36	216	1296	66	396
	7	14	49	313	2401	98	686
Σ	34	95	154	820	4774	377	1849

for the equation: $y = a + b_1x + b_2x^2$

The normal equations are:

$$\Sigma y = na + b_1\Sigma x + b_2\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b_1\Sigma x^2 + b_2\Sigma x^3$$
$$\Sigma x^2y = a\Sigma x^2 + b_1\Sigma x^3 + b_2\Sigma x^4$$

Subtitute known values:

$$95 = 10a + 34b_1 + 154b_2 \quad \text{--- (1)}$$
$$377 = 34a + 154b_1 + 820b_2 \quad \text{--- (2)}$$
$$1849 = 154a + 820b_1 + 4774b_2 \quad \text{--- (3)}$$

Solving eq ① ② & ③

$a = \cancel{\ldots}$ 1.8022

$b1 = \cancel{\ldots}$ 3.4822

$b_2 = \cancel{\ldots}$ -0.2689

The regression equation using non-linear form $y = a + b_1 x + b_2 x^2$ is:

$y = \cancel{\ldots}$

$y = 1.802 + 3.482x - 0.268x$

the Coefficients of regression in the non-linear model are:

a (intercept) ≈ ~~...~~ 1.8022

b1 (linear regression coefficient) ≈ ~~...~~ 3.4...

b2 (quadratic regression coefficient) ≈ -0...

May june 2023

Q.10)

paper I	45	55	56	58	60	65	68	70	75	80	85
paper II	56	50	48	60	62	64	65	70	74	82	90

Calculate the Coefficient of Correlation

X	Y	x^2	y^2
45	56		
55	50		
56	48		
58	60		
60	62		
65	64		
68	65		
70	70		
75	74		
80	82		
85	90		

$\Sigma x = 717$ N=

$\Sigma y = 721$

$\Sigma x^2 = 48149$

$\Sigma y^2 = 48905$

$\Sigma xy = 48398$

$\bar{x} = \frac{\Sigma x}{n} = 65.18$

$\bar{y} = \frac{\Sigma y}{n} = 65.54$

$$\sigma x = \sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2} = \sqrt{\frac{1}{11}(48149) - (65.18)^2}$$
$$= 11.34$$

$$\sigma y = \sqrt{\frac{1}{n}\Sigma y^2 - \bar{y}^2} = \sqrt{\frac{48905}{11} - (65.54)^2}$$
$$= 12.26$$

$$\cancel{\sigma xy(x,y) = \frac{1}{n}\Sigma xy - \bar{x}.\bar{y}}$$
$$\cancel{= \sqrt{\frac{48398}{11} - (65.18)^2 \times (65.54)^2}}$$
$$\cancel{= \sqrt{-182447605.87}}$$

$$Cov(x,y) = \frac{1}{n}\Sigma xy - \bar{x}.\bar{y}$$
$$= \frac{1}{11}(48398) - (65.18)\times(65.54)$$
$$= 127.92$$

Coefficient of Correlation (r):

$$\frac{Cov(x,y)}{\sigma x . \sigma y}$$
$$= \frac{127.92}{11.34 \times 12.26}$$
$$= \boxed{0.92}$$

Q.11) find quartile deviation and coefficient of quartile deviation

marks	<10	10-20	20-30	30-40	40-50	50-60
students	10	20	30	50	40	30

→

marks	Students	C.F.
<10	10	10
10-20	20	30
20-30	30	60
30-40	50	110
40-50	40	150
50-60	30	180

$N = 180$

$Q_1 = \left(\frac{N}{4}\right)^{th}$ obs

$= \left(\frac{180}{4}\right) = 45^{th}$

$\therefore$ 20-30 is Q_1 cl

$\therefore Q_1 = l + \frac{w}{f}\left(\frac{N/4 - C.F}{f}\right)$

$= 20 + \frac{10}{30}\left(\frac{45-30}{30}\right)$

$= 20.166$

$Q_3 = \left(\frac{3N}{4}\right)^{th}$ obs

$= \left(\frac{180\times3}{4}\right) = 135^{th}$ obs $\therefore$ 40-50 Q_3 c

$\therefore Q_3 = 40 + \frac{10}{40}\left(\frac{\frac{3\times180}{4} - C.F}{40}\right) =$

$= 40.156$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{40.156 - 20.166}{2} = \boxed{9.995}$

~~Coeff. of Q.D =~~

$\frac{Q_3 - Q_1}{Q_3 + Q_1} = \boxed{0.33138}$

Q.12) Determine eqns of regression line. Also find (i) y for $x = 4.5$ (ii) x when $y = 13$

x	2	3	5	7	9	10	12	15
y	2	5	8	10	12	14	15	16

x	y	x^2	y^2	xy
2	2	4	4	4
3	5	9	25	15
5	8	25	64	40
7	10	49	100	70
9	12	81	144	108
10	14	100	196	140
12	15	144	225	180
15	16	225	256	240

$\Sigma x = 63$
$\Sigma y = 85$
$\Sigma x^2 = 637$
$\Sigma y^2 = 1014$
$\Sigma xy = 797$
$n = 8$

i) Line of regression y on x

$\Sigma y = aN + b\Sigma x$
$\Sigma xy = a\Sigma x + b\Sigma x^2$

$85 = 8a + 63b$ —— (1)
$797 = 63a + 637b$ —— (2)

Solve eq (1) & (2)

$a = 3.49 \quad b = 0.905$

$\therefore$ y on x

$y = a + bx$

$\boxed{y = 3.49 + 0.905x}$

ii) Line of regression x on y

$\Sigma x = aN + b\Sigma y$
$\Sigma xy = a\Sigma y + b\Sigma y^2$

$63 = 8a + 85b$ —— ①
$797 = 85a + 1014b$ —— ②

Solving eq① & eq②

$a = -4.355 \quad b = 1.151$

$x = a + by$

$\boxed{x = -4.355 + 1.51b}$

i) y for $x = 4.5$

$y = 3.49 + 0.905(4.5)$

$\boxed{y = 7.5625}$

ii) x when $y = 13$

$x = -4.355 + 1.151(13)$
$= 10.608$

Q.13) The first four moments of dist. about the value 4 are 2, 20, 40, 100 respectively
i) obtain first Central moments
ii) Find mean, S.D
iii) Coeffi. of Skewness & Kurtosis.

→ Given: first four moments about point (4):
$\mu_1' = 2,\ \mu_2' = 20,\ \mu_3' = 40,\ \mu_4' = 100$

Solution:

Origin $A = 4$, and the mean $\bar{x} = A + \mu_1'$
$\bar{x} = 4 + 2$
$= \boxed{6}$

first four Central moments

$\mu_1 = 0$ (by definition)

$\mu_2 = \mu_2' - (\mu_1')^2$
$= 20 - (2)^2$
$= 16$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$
$= 40 - 3(20)(2) + 2(2)^3$
$= -64$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$
$= 100 - 4(40)(2) + 6(20)(2)^2 - 3(2)^4$
$= \cancel{452}\ 212$

Now Coeff of Skewness $(\beta_1) = \frac{\mu_3^2}{\mu_2^3}$

$= \frac{(-64)^2}{(16)^3} = \boxed{-1}$

Coeff. of Kurtosis $(\beta_2) = \frac{\mu_4}{\mu_2^2}$

$= \frac{212}{16^2} = \boxed{0.8281}$

Now,
Second Central moment is variance
$\therefore$ S.D $\sigma = \sqrt{\mu_2}$
$= \sqrt{16}$
$\boxed{\sigma = 4}$

may june 2024

Q.14) The first four moments about work mean 30.2 of dist. 0.255, 6.22, 30.211, ... Calculate first four moments about the mean, Also β_1, β_2

$\mu_1' = 0.255$, $\mu_2' = 6.22$, $\mu_3' = 30.211$, $\mu_4' = ...$

$\mu_1 = 0$ ... by definition

$\mu_2 = \mu_2' - (\mu_1')^2$
$= 6^2 \quad 6.15$

Q.15) Ob... fo... x ... y ...

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$
$= 25.48$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$
$= 371.8488$

$\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 2.79$

$\beta_2 = \dfrac{\mu_4}{\mu_2^2} = 9.8$

Q.15) obtain the regression lines y on x & x on y for the data :

x	5	1	10	3	9
y	10	11	5	10	6

→ $\Sigma x = 28$ $\Sigma x^2 = 216$ $\Sigma xy = 1\cancel{77}95$
$\Sigma y = 42$ $\Sigma y^2 = 382$

i) Regression of line y on x
$\Sigma y = aN + b\Sigma x$
$\Sigma xy = a\Sigma x + b\Sigma x^2$

$42 = 5a + 28b$ —— ①
$177 = 28a + 216b$ —— ②
Solve eq ① & ②
$a = 13.905$ $b = -0.983$

$\boxed{y = 13.905 - 0.983x}$

$y = \cancel{0.67}9\ 12.2 - 0.679x$

ii) Regression of line x on y

$\Sigma x = aN + b\Sigma y$

$\Sigma xy = a\Sigma y + b\Sigma y^2$

$28 = 5a + 42b$ —— (1)

$177 = 42a + 382b$ —— (2)

Solving eq (1) & (2)

$a =$ $b =$

$a = 22.342$ $b = -1.993$

$x = \cancel{22.342 - 1.933y}$

$\boxed{x = 17.16 - 1.377y}$

Q.10) For the following dist. find (i) first 4 moments about any point (ii) four Central moments

X	2	2.5	3	3.5	4	4.5	5
F	5	38	65	92	70	40	10

Lets take $A = 5$

f	x	$d = x - A$	fd	fd^2	fd^3	fd^4
5	2	−3				
38	2.5	−2.5				
65	3	−2				
92	3.5	−1.5				
70	4	−1				
40	4.5	−0.5				
10	5	0				
Σ 320						

Σ ~~−192~~ ~~[illegible]~~ −1634.25 ~~[illegible]~~

Σ −408 829.5 3467.625

$\therefore \mu_1' = \frac{\Sigma fd}{N} = \frac{-468}{320} = -1.4625$

$\mu_2' = \frac{\Sigma fd^2}{N} = \frac{829.5}{320} = 2.5922$

$\mu_3' = \frac{\Sigma fd^3}{N} = \frac{-1634.25}{320} = -5.10$

$\mu_4' = \frac{\Sigma fd^4}{N} = \frac{3467.625}{320} = 10.83$

four central moments:

$\mu_1 = 0$ ...... by definition

$\mu_2 = \mu_2' - (\mu_1')^2 = 2.59 - (-1.46)^2$

$= \cancel{0.4584}$

$= 0.4584$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$

$= -5.10 - 3(2.59)(-1.46) + 2(-1.46)^3$

$= \cancel{4}\ 0.0199$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$

$= 10.83 - 4(-5.10)(-1.46) + 6(2.59)(-1.46)^2 - 3(-1.46)^4$

$= \cancel{\ldots}\ 0.5399$

Q.17) Regression equations are $8x - 10y + 66 = 0$ and $40x - 18y = 214$, variance of $x = 9$
i) The mean values of x and y
ii) The Correlation x & y
iii) S.D. of y.

$8x - 10y \cancel{+66} = \cancel{0}\ -66$ —— ①
$40x - 18y \cancel{-214} = \cancel{0}\ 214$ —— ②
Solving eq① & ②

$x = 13$, $y = 17$
∴ The mean values of x & y are 13 & 17 respectively.

$8x - 10y + 66 = 0$ $\quad$ $40x - 18 = 214$
$x = \frac{10y}{8} + \frac{-66}{8}$ $\quad$ $y = \frac{40x}{18} - \frac{214}{18}$
$bxy = \frac{10}{8} = 1.25$ $\quad$ $byx = \frac{40}{8} = 2$

~~(r)~~ Coeff. of Correlation $(r) = \sqrt{1.25 \times 2.22}$
$= 1.66$

$byx = \frac{8}{10} = 0.8$ $\quad$ $bxy = \frac{18}{40} = 0$
$(r) = \sqrt{0.8 \times 0.45} = \boxed{0.6}$

iii) S.D of y
Variance of $(x) = 9$ ∴ $\sigma x = \sqrt{9} = 3$

$byx = r \cdot \frac{\sigma y}{\sigma x}$ ∴ $0.8 = 0.6 \times \frac{\sigma y}{3}$

$\sigma y = \frac{0.8 \times 3}{0.6}$
$= 1.2$

SDS Page No. 14

Nov-Dec 2022

Q.18) Calculate

i) Q.D

ii) M.D. from mean

Class	(f) Students	C.F.	x	fx
0-10	6	6	5	30
10-20	5	11	15	75
20-30	8	19	25	200
30-40	15	34	35	525
40-50	7	41	45	315
50-60	6	47	55	330
60-70	8	55	65	520

$N = \Sigma f = 55$

$\Sigma fx = 1995$

$(\bar{x}) = \frac{\Sigma fx}{N} = \frac{1995}{55}$

$= 36.277$

$Q_1 = \left(\frac{N}{4}\right)^{th}$ obs. $\left(\frac{55}{4}\right) = 13.75$

$\therefore Q_1$ class is 20-30

$Q_1 = \cancel{20}\ l + \frac{w}{f}\left(\frac{\frac{N}{4} - C.F.}{f}\right)$

$= 20 + \frac{10}{8}\left(\frac{13.75 - 11}{8}\right)$

$= 20.4296$

$Q_3 = \left(\frac{3N}{4}\right)^{th}$ obs $\left(\frac{3 \times 55}{4}\right) = 41.25$

$\therefore$ 50-60 is the Q_3 class

$Q_3 = 50 + \frac{10}{6}\left(\frac{41.25 - 41}{6}\right)$

$= 50.0694$

Quartile deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{50.0694 - 20.4295}{2}$

$= 14.8199$

ii) Mean Deviation from mean

M.D about mean $= \frac{\Sigma|x-\bar{x}|}{n}$

$= \cancel{(5-36.277) + (15-36.277)} + (2$

$\Sigma|x-\bar{x}| = -8.939$

$M.D. = \frac{\Sigma|x-\bar{x}|}{n} = \frac{-8.939}{55}$

$= \boxed{-0.1625}$

Q19) The variables x and y are connected by equation $ax+by+c=0$. Show that the correlation between them is -1 if the of a & b are alike and $+1$ if they are diff.

→ we are given linear relation betn two variabs:

$ax+by+c=0$ — ①

from ①

$by = -ax - c$

$y = -\frac{ax}{b} - \frac{c}{b}$

This is a linear eqn of the form
$y = mx + k$ where
$m = -\frac{a}{b}$, $k = -\frac{c}{b}$

The Coeff. of Correlation (r) for a perfect linear eqn is either
1) $r = +1$ if slope $m > 0$
$r = -1$ if slope $m < 0$

If a & b have same sign (both +ve or -ve) then $\frac{a}{b} > 0$, and later by applying
the $-$ sign in front it becomes negative
$\therefore m = -\frac{a}{b} < 0 \Rightarrow$ negative slope $\Rightarrow r = -1$

&

if the a & b have opposite signs then
$\frac{a}{b} < 0$, so and later by appying the $-$ sign in front it becomes positive
$\therefore m = -\frac{a}{b} > 0 \Rightarrow$ positive slope $\Rightarrow r = +1$

$\therefore$ final conclusion
if a & b have same sign $\Rightarrow r = -1$
& if a & b have opposite sign $\Rightarrow r = +1$

SDS Page No. Date

Q.20) Given : workers Avg. daily wage Variance (may june 2019)

	workers	Avg. daily wage	Variance
firm A	500	Rs. 186	81
firm B	600	Rs. 175	100

i) larger wage bill :
firm A : $500 \times 186 = 93000$
firm B : $600 \times 175 = 105000$
$\therefore$ firm B has a larger wage bill

ii) greater variability in wages
variability is determined by variance of wages :
firm A variance = 81
firm B variance = 100
$\therefore$ firm B has greater variability.

iii)(a) Average daily wage of all workers combined
Combined mean formula

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2} = \frac{93000 + 105000}{1100} =$$

(b) variance of dist. of wages of all wo[rkers] combined

$$\sigma^2 = \frac{n_1(\sigma_1^2 + (\bar{x}_1 - \bar{x})^2) + n_2(\sigma_2^2 + (\bar{x}_2 \ldots}{n_1 + n_2}$$

$$\frac{500(81 + 36) + 600(100 + 25)}{1100} = \boxed{121.36}$$

A odd pro lea

may june 2024

Unit - 4
Q3 & Q4

Ques A Dice is thrown 10 times. If getting odd number is a success. What is the prob. of getting (i) 8 success (ii) At least 6 successes

→ $n = 10$
$p = 0.5$ (Bionomial Dist.)
$q = 0.5$

(i) $r = 8$
$P(x=r) = {}^nC_r \, p^r q^{n-r}$
$P(x=8) = {}^{10}C_8 (0.5)^8 (0.5)^2$
$= 0.04394$

(ii) at least 6
This means $P(x \geq 6) = P(6) + P(7) + P(8) + P(9) + P(10)$

r	${}^{10}C_r$
6	210
7	120
8	45
9	10
10	1

$\left({}^{10}C_6 \times (0.5)^6 \times (0.5)^4\right) + \left({}^{10}C_7 \times 0.5^7 \times 0.5^3\right) +$
$\left({}^{10}C_8 \times 0.5^8 \times 0.5^2\right) + \left({}^{10}C_9 \times 0.5^9 \times 0.5^1\right) +$
$\left({}^{10}C_{10} \times 0.5^{10} \times 0.5^0\right)$

$= 0.3769$

Q22) If the probablity than an individual suffers a bad reaction from a certain injection is 0.001. determine the prob. that out of 2000 individuals (i) Exactly 3 (ii) more than 1 suffers a bad reaction

→ we are given $n = 2000$, $p = 0.0001$

$m = np = 2000 \times 0.0001 = 2$

for poissons distribution

$$p(x) = \frac{e^{-m} \cdot m^x}{x!}$$

$$p(x) = \frac{e^{-2} \cdot 2^x}{x!}$$

(i) P(exactly 3)

$= P(X=3)$

$$= \frac{e^{-2} \cdot 2^3}{3!} = \boxed{0.1804}$$

(ii) probablity that more than 1 suffer:

$P(x>1) = 1 - P(0) - P(1)$

$$P(0) = \frac{e^{-2} \cdot 2^0}{0!} = 0.1353$$

$$P(1) = \frac{e^{-2} \cdot 2^1}{1!} = 0.2706$$

$P(x>1) = 1 - 0.1353 - 0.2706$

$= 0.5941$

Q.23

Q23) for a normal dist. when mean = 2, SD = 4, find the probablities of the following intervals:
i) $4.43 < x < 7.29$
ii) $-0.43 < x < 5.39$
[Given: $A(z=0.61) = 0.2291$, $A(z=1.32) = 0.4066$, $A(z=0.85) = 0.3023$]

→ Given: mean $\mu = 2$
S.D. $\sigma = 4$

(i) $4.43 < x < 7.29$
Convert both values to z-score

$z = \frac{x-\mu}{\sigma}$

$z_1 = \frac{4.43-2}{4} = 0.61$

$z_2 = \frac{7.29-2}{4} = 1.32$

$A(z=0.61) = 0.2291$
$A(z=1.32) = 0.4066$

$4.43 < x < 7.29 = A(1.32) - A(0.61)$
$= 0.4066 - 0.2291$
$= 0.1775$

(ii) $-0.43 < x < 5.39$

$z_1 = \frac{-0.43-2}{4} = -0.61$

$z_2 = \frac{5.39-2}{4} = 0.85$

$A(z=0.61) = 0.2291$, $A(-0.61) = -0.$

$A(z=0.85) = 0.3023$

$$P(-0.43 < x < 5.39) = A(0.85) - A(-0.61)$$
$$= 0.3023 - (-0.2291)$$
$$= 0.3023 + 0.2291$$
$$= \boxed{0.5314}$$

Q.24) A Random variable X with the following prob. dist.

X	0	1	2	3	4
$P(x)$	0.1	k	2k	2k	k

Find :

i) k

ii) $P(x<2)$

iii) $P(x \geq 3)$

iv) $P(1 \leq x \leq 3)$

→ We know the total probablity must be 1

$$p(0) + p(1) + p(2) + p(3) + p(4) = 1$$
$$0.1 + k + 2k + 2k + k = 1$$
$$0.1 + 6k = 1$$
$$6k = 0.9$$
$$k = \frac{0.9}{6}$$
$$= 0.15$$

ii) $P(x<2)$

$$p(0) + p(1) = 0.1 + k$$
$$= 0.1 + 0.15$$
$$= \cancel{2.5}\ \boxed{0.25}$$

iii) $P(x \geq 3)$

$p(3) + p(4) = 2k + k$

$= 3k$

$= 3(0.15)$

$= \boxed{0.45}$

iv) $p(1 \leq x \leq 3)$

$p(1) + p(2) + p(3) = k + 2k + 2k$

$= 5k$

$= 5(0.15)$

$= \boxed{0.75}$

Q25) MNC Company Conducted 1000 candidates aptitude test. Avg score = 45, SD = 25. Assume normal dist.

find i) Candidates whose score exceed 60

ii) candidates whose score lies between 30 & 60.

→ $n = 1000$

mean $\mu = 45$

S.D $\sigma = 25$

Normal dist

$A(z = 0.6) = 0.2257$

i) No. of Candidates score exceeds 60

$$z = \frac{x - \mu}{\sigma} = \frac{60 - 45}{25} = 0.6$$

find Area to the right of z i.e. $8P$.

$P(x > 60) = 0.5 - A(z = 0.6)$

$= 0.5 - 0.2257 = 0.2743$

Convert to number of Candidates

Required Candidates $= 1000 \times 0.2743$

$= 274.3 \approx \boxed{274}$

(ii) Candidates Score lie between 30 & 60

$z_1 = \frac{30 - 45}{25} = -0.6$

$z_1 = 0.6$

Total area between -0.6 and 0.6

$p(30 < x < 60) = 2 \times A(z = 0.6)$

$= 2 \times 0.2257$

$= 0.4514$

Convert to no. of Candidates

$= 0.4514 \times 1000$

$= 451.4 \approx \boxed{451}$

May june 2023

Q.26) In a Certain Company install 2000 LED bulbs on each floor. If LED bulbs have Average life of 1000 burning hours with Standard deviation of 200 hours using normal dist find what number of LED bulbs might ~~get~~ be expection ed to fail in 700 hours

(Given: $P(0 < z < 1.5) = 0.4332$)

$n = 2000$
mean $\mu = 1000$
SD $\sigma = 200$

$$z = \frac{700 - 1000}{200} = \frac{-300}{200} = -1.5$$

we are given $P(0 < z < 1.5) = 0.4332$

Since normal distribution is symmetric
$P(z < -1.5) = 0.5 - P(0 < z < 1.5)$
$= 0.5 - 0.4332 = 0.0668$

No. of bulbs that fail in 700 hours
$2000 \times 0.0668 = 133.6 \approx \boxed{134 \text{ bulbs}}$

Q27) Between 2pm to 4pm the average no. of phone calls per minute is 2.5. find prob. that during a particular minute there will be.
i) no phone call
ii) exactly 3 phone calls.

$$P(X = x) = \frac{e^{-m} \cdot m^x}{x!}$$

$m = 2.5$
x = no. of phone calls

i) probability of no phone calls $(x = 0)$

$$\frac{e^{-2.5} \cdot 2.5^0}{0!} = \boxed{0.0820}$$

ii) Exactly 3 $(x=3)$

$$\frac{e^{-2.5} \cdot 2.5^3}{3!} = 0.2137$$

Q.28) weights of 4000 students found to be normally distributed with mean 50 kg and SD 5kg. find the no. of students with weights i) less than 45kgs ii) between 45 to 50 kgs. (~~for SD~~ area under the curve between $z=0$ to $z=1$ is 0.3413 & that betn $z=0$ to $z=2$ is 0.4772)

→ $n = 4000$

$\mu = 50$ $\quad \frac{x-\mu}{\sigma}$

$\sigma = 5$

$p(x < 45) = z = \frac{45-50}{5} = -1$

Use symmetry of normal dist.

Area between 0 and $z=1$ is 0.3413

So, area to the left of $z=-1$ is

$p(x<45) = 0.5 - 0.3413 = 0.1587$

total students with weight < 45

$= 4000 \times 0.1587$

$= 634.8 \approx 635$

Stud

ii) Students between 45 to 60 kg

$z_1 = \frac{45-50}{5} = -1$

$z_2 = \frac{60-50}{5} = 2$

Area between $z = -1$ and $0 = 0.3413$
Area between $z = 0$ and $2 = 0.4772$

So total Area betwee -1 & 2
$0.3413 + 0.4772 = 0.8185$

multiply by total students
$4000 \times 0.8185 = 3274$ Students

29) If 10% bolts produced by a machine are defective. Determine the probablity that out of 10 bolts chosen at random
i) two will be defective
ii) at most two will be defective

→

$n = 10$
$p = 10\% = 10/100 = 0.10$
$q = 1 - p = 0.90$

binomial probablity formula is
$p(X = r) = {}^{n}C_{r} \cdot p^{r} \cdot q^{n-r}$

i) probability that exactly 2 bolts are defe[ctive]
$p(X=2) = {}^{10}C_{2} \times 0.10^{2} \times 0.9^{8}$

$\boxed{= 0.1937}$

ii) At most 2 defective bolts.

$p(x \le 2) = p(0) + p(1) + p(2)$

$\therefore \left({}^{10}C_0 \times 0.1^0 \times 0.9^{10}\right) + \left({}^{10}C_1 \times 0.1^1 \times 0.9^9\right) + \left({}^{10}C_2 \times 0.1^2 \times 0.9^8\right)$

$= \boxed{0.9298}$

Nov Dec 2023

Q.30) 20% of bolts produced by a machine are defective. Determine the prob. out of 4 bolts chosen at a random

i) 1 is defective

ii) zero are defective

iii) At most 2 bolts are defective.

→ $n = 4$

$p = 0.2$

$q = 1 - p = 0.8$

i) $p(1) = {}^4C_1 \times p^1 \times q^3$

$= {}^4C_1 \times 0.2^1 \times 0.8^3$

$= \boxed{0.4096}$

ii) $P(0) = {}^4C_0 \times 0.2^0 \times 0.8^4$

$= \boxed{0.4096}$

iii) At most 2 : $p(0)+p(1)+p(2)$

$p(2) = {}^4C_2 \times 0.2^2 \times 0.8^2$

$= 0.1536$

$p(0)+p(1)+p(2)$

$= 0.4096 + 0.4096 + 0.1536$

$= \boxed{0.9728}$

31) The average no. of misprints per page of a book is 1.5. Assuming the dist. of no. of misprints to be poisson. find

i) The prob. that a particular book is free from misprint

ii) No. of pages containing more than one misprint if the book contains 900 pages

→ $m = 1.5$

$x =$ no. of misprints

i)

$$p(x=0) = \frac{e^{-1.5} \times 1.5^0}{0!}$$

$= 0.2231$

ii) $p(x>1) = 1 - p(0) - p(1)$

$$p(1) = \frac{e^{-1.5} \times 1.5^1}{1!} = 0.3346$$

$p(x > 1) = 1 - 0.2231 - 0.3346$
$= 0.4423$

Number of pages with more than one misprint when book contains 900 pages $= 0.4423 \times 900$
$= 398.07 \approx 398.$

Q32) Fit a poissons dist. to the following data and calculate theoretical frequencies.

x	0	1	2	3	4	Total
f	109	65	22	3	1	200

→ i) Calculate the mean $(m) = \frac{\Sigma fx}{N}$

$\Sigma fx = 122$ $\frac{122}{200} = 0.61$

ii) using poissons formula for theoretical frequencies

$$p(x) = \frac{e^{-m} \cdot m^x}{x!}$$

Expected freq

$\frac{e^{-0.61} \times 0.61^0}{0!} = 0.5433,\ 0.5433 \times 200 = 108$ Q33
≈ 109

$\frac{e^{-0.61} \times 0.61^1}{1!} = 0.3314,\ 0.3314 \times 200 = 66.28$
≈ 66

$\frac{e^{-0.61} \cdot 0.61^2}{2!} = 0.1010, \; 0.1010 \times 200 = 20.21$

≈ 20

$\frac{e^{-0.61} \cdot 0.61^3}{3!} = 0.0205, \; 0.0205 \times 200 = 4.1$

≈ 4

$\frac{e^{-0.61} \cdot 0.61^4}{4!} = 0.00313, \; 0.00313 \times 200 = 0.62$

≈ 1

final theoretical frequencies

x	observed	theoretical
0	109	109
1	65	66
2	22	20
3	3	4
4	1	1

These match closely with the given frequencies - so poissons distribution is a good fit here.

The lifetime of a article has a normal distribution with mean 400 hours & S.D. 50 hours. find expected no. of articles out of 2000 whose lifetime lies between 335 hours to 465 hours.
$(A(z=1.3) = 0.4032)$

→ $\mu = 400$
$\sigma = 50$
$P(\cancel{x}\ 335 < x < 465)$

$$z_1 = \frac{335-400}{50} = -1.3$$

$$z_2 = \frac{465-400}{50} = 1.3$$

$A(z=1.3) = 0.4032$

The total probability from -1.3 to 1.3 is:

$P(-1.3 < z < 1.3) = 2 \times A(1.3)$
$= 2 \times 0.4032$
$= 0.8064$

multiply by total Articles
Expected no. of articles whose lifetime lies between 335 and 465 hours
$= 2000 \times 0.8064$
$= 1612.8 \approx \boxed{1613}$

Nov. Dec. 2022

Q.34> fit poissons dist. to following data & Calculate theoretical frequencies.

x	0	1	2	3	4
f	122	60	15	2	1

$\Sigma fx = 100$

mean $m = \frac{100}{200} = 0.5$ $N = 200$

for $x = 0$

$\frac{e^{-0.5} \cdot 0.5^0}{0!} = 0.6065, 0.6065 \times 200 = 121.3 \approx 121$

for $x = 1$

$\frac{e^{-0.5} \cdot 0.5^1}{1!} = 0.3032, 0.3032 \times 200 = 60.65 \approx 61$

for $x = 2$

$\frac{e^{-0.5} \cdot 0.5^2}{2!} = 0.0758, 0.758 \times 200 = 15.16 \approx 15$

for $x = 3$

$\frac{e^{-0.5} \cdot 0.5^3}{3!} = 0.0126, 0.0126 \times 200 = 2.52 \approx 3$

for $x = 4$

$\frac{e^{-0.5} \cdot 0.5^4}{4!} = 0.0015, 0.0015 \times 200 = 0.3 \approx 1$

x	Observed (f)	theoretical (f)
0	122	121.3
1	60	60.65
2	15	15.16
3	2	2.52
4	1	0.32

Q35) In a Sample 1000 Case the means of Certain test is 14 and S.D = 2.5 Assume dist. is normal find
i) How many students scored betn 12 to 15
ii) How many scored below 8.
$(z=0.8)=0.2881$, $(z=0.4)=0.1554$, $(z=2.4)=0.4918$

→ $N=1000$
$\mu=14$
$\sigma=2.5$

i) students scored betn 12 to 15
$p(12<z<15)$

$$z_1 = \frac{12-14}{2.5} = -0.8$$

$$z_2 = \frac{15-14}{2.5} = 0.4$$

$p(\cancel{12}\ -0.8<z<0.4)$
$= (0.2881+0.1554)$
$= 0.5568$
~~Students scored betn 12 to 15 = 1000 × 0.556~~
~~= 556.5 ≈ 557 students~~

$p(-0.8<z<0.4)$
$= (0.2881+0.1554)$
$= 0.4435$
No. of students betn 12 to 15
$= 0.4435 \times 1000$
$= 443.5 \approx \boxed{444 \text{ students}}$

ii) Students scored below 8
$p(x<8)$

$$z = \frac{8-14}{2.5} = -2.4$$

$p(z<2.4) = 0.5 - 0.4918 = 0.0082$

Number of students scored below 8 $= 0.0082 \times 1000$
$\boxed{= 8}$

30) A Random variable X with the following prob. dist.

X	1	2	3	4	5	6	7
f	k	2k	3k	k^2	k^2+k	$2k^2$	$4k^2$

→ $p(1)+p(2)+p(3)+p(4)+p(5)+p(6)+p(7) = 1$
$\therefore k+2k+3k+k^2+k^2+k+2k^2+4k^2 = 1$
$\therefore 7k + 8k^2 = 1$
$\cancel{8k^2} = 1-7k$
$\cancel{k^2 = \frac{1-7k}{8}}$
$\cancel{7k+k^2 = \frac{1}{8}}$
$8k^2+7k-1=0$

$$\cancel{\frac{-b \pm \sqrt{b^2-4ac}}{2a}}$$

$x_1 = 0.125$ (valid) $\quad x_2 = -1$ (invalid)

$\therefore k = 0.125$

i) $k = 0.125$

ii) $p(x > 5) : p(6) + p(7)$

$= 2(0.125)^2 + 4(0.125)^2$

$= 0.9375$

(iii) $p(1 \le x \le 5) : p(1) + p(2) + p(3) + p(4) + p(5)$

$k + 2k + 3k + k^2 + k^2 + k$

$7k + 2k^2$

$= 0.9062$

Q37) for a distribution when mean = 1, Standard deviation = 4, find the probablities of the following intervals:

i) $3.43 < x < 6.19$ ii) $-1.43 \le x < 6.19$

$(A(z = 0.81) = 0.2910, A(z = 1.73) = 0.4582]$

→ $\mu = 1$, $\sigma = 4$

i) $p(3.43 \le x < 6.19)$

Q.38)

1	2	3	4	5
k	$2k$	$2k$	k	$7k^2$

$7k^2 + 6k - 1 = 0$

$x_1 = \cancel{0.85}$ $1/7$ (valid)

$x_2 = \cancel{0}$ -1 (invalid)

i) $k = 0.1428$ ii) $P(x \ge 2) = 5k + 7k^2 = 0.85714$

iii) $P(x < 3) = 3k = 0.4285$ iv) $P(2 \le x \le 3)$

$4k = 0.5714$

v) $P(x \ge 3) = 3k + 7k^2 = 0.5714$

Unit 5

May-June 2024

Q.39)

Days	Sun	Mon	Tue	Wed	Thu	Fri	Sat
No. of Customer	6	4	9	7	8	10	12

Test at 5% level of significance whether customer visits are uniformly distributed over the days.

→

H_0 : Customer visits are uniformly distribut
H_1 : Customer visits are not uniformly distributed.

Significance level $(\alpha) = 0.05$

observed values

Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
6	4	9	7	8	10	12	56

~~Expected values~~ observed values

Days	Customers
Sun	6
Mon	4
Tue	9
Wed	7
Thu	8
Fri	10
Sat	12
Total	56

Expected frequencies (E):
Since uniform dist. is assumed

$E = \frac{56}{7} = 8$ for each day

Chi-square test statistic formula

Day	O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
Sun	6	8	-2	4	0.5
mon	4	8	-4	16	2
Tue	9	8	1	1	0.125
wed	7	8	-1	1	0.125
Thu	8	8	0	0	0
fri	10	8	2	4	0.5
sat	12	8	4	16	2
					5.25

Decision

Calculated $X^2 = 5.25$

Critical value at 5% los = 6 → $X^2_{0.05,6} = 15.592$

Since

$5.25 \nless 15.592$

Conclusion:

We fail to reject Ho,
There is no significant evidence to suggest that customer visits are not uniformly distributed over the days.

Q.4c) In a batch of 500 articles, produced by a machine, 16 articles are found defective. After overhauling the machine it is found that 3 articles are defective in a batch of 100. Has the machine improve.

$\rightarrow$ $n_1 = 500, \ n_2 = 100$

$p_1 = \frac{16}{500} = 0.032$

$p_2 = \frac{3}{100} = 0.03$

$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{19}{600} = 0.032$

$Q = 1 - P = 1 - 0.032 = 0.968$

Since we have to test whether the machine is improved or not

$\therefore$ H_0 : machine has not improved after overhauling i.e. $P_1 = P_2$

H_1 : $P_1 > P_2$ (right tailed)

under H_0,

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.032 - 0.03}{\sqrt{0.032 \times 0.968 \times \left(\frac{1}{500} + \frac{1}{100}\right)}}$$

$= 0.1037$

Since, $0.1037 < 1.96$

we failed to reject H_0

At 5% los, there is not enough evidence to conclude that machine was improved.

Q41) In two independant samples of size 8 and 10 the sum of squares deviation of sample values from the respective sample means were 84.4 and 102.6. Test whether the diff. of variances of the population is significance or not (Given $F_{t, 0.05} = 3.29$)

Q42) In

→ $n_1 = 8$, $n_2 = 10$

Sum of squares deviations: $SS1 = 84.4$
$SS2 = 102.6$

for sample variance

$$S^2 = \frac{\text{sum of squares}}{n-1}$$

So $S_1^2 = \frac{84.4}{7} = 12.06$

$S_2^2 = \frac{102.6}{9} = 11.4$

$H_0 : \sigma_1^2 = \sigma_2^2$ (variances are equal)

$H_1 : \sigma_1^2 \neq \sigma_2^2$ (variances are different)

F-statistic $F = \frac{\text{Larger sample variance}}{\text{Smaller sample variance}}$

$= \cancel{10.2} \; \frac{12.06}{11.4}$

$= 1.508$

Compare with critical value

Since

$F = 1.508 < 3.29$

we fail to reject null hypothesis.

There is no significant difference betn population variances.

Q4) In	Round & green	Wrinkled & green	Round & yellow	Wrinkled yellow	Total
	222	120	32	150	524

proportion : 8:2:2:1 , [Given : $\chi^2_{3,0.05} = 7.815$]

→ Total = 524 Seeds
Theoretical ratio : 8:2:2:1 (total parts = 13)
Critical value $\chi^2_{3,0.05} = 7.815$

Expected frequencies
Round & green : $8/13 \times 524 = 322.46$
Wrinkled & green: $2/13 \times 524 = 80.61$
Round & yellow: $2/13 \times 524 = 80.61$
Wrinkled & yellow : $1/13 \times 524 = 40.30$

Chi-Square statistic formula

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

Category	O	E	$(O-E)^2/E$
R & G	222	322.46	31.29
W & G	120	80.61	19.24
R & Y	32	80.61	29.31
W & Y	150	40.30	298.61
			378.45

$\chi^2_{calculated} = 378.45 > \chi^2_{3,0.05} = 7.815$

We Reject H_0 (null hypothesis)
There is significant difference between observed and expected frequencies.
The theory does not fit the experiment

Q43) for Sample I : $n_1 = 1000$, $\Sigma x = 49000$, $\Sigma(x-\bar{x})^2 = 7,84,000$
for Sample II : $n_2 = 1500$, $\Sigma x = 70500$, $\Sigma(x-\bar{x})^2 = 24,00,000$
Discuss the significance. difference between mean score. (Given $z_\alpha = 1.96$)

$\longrightarrow$ Given: Sample 1: $n_1 = 1000$, $\Sigma x_1 = 49000$, $\Sigma(x-\bar{x}_1)^2 = 784000$

$\Rightarrow \bar{x}_1 = \frac{49000}{1000} = 49$, $\Rightarrow \sigma_1^2 = \frac{784000}{1000} = 784$

Sample 2:
$n_2 = 1500$, $\Sigma x_2 = 70500 \Rightarrow \bar{x}_2 = \frac{70500}{} = 47$
$\Sigma(x-\bar{x})^2 = 2400000 = \sigma_2^2 = \frac{2400000}{1500} = 1600$

H_0: There is no significant difference betn the means $\mu_1 = \mu_2$
H_1: There is a significant difference $\mu_1 \neq$

z-test for difference of means

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{49-47}{\sqrt{\frac{784}{1000} \frac{1600}{1500}}}$$

$$= \frac{2}{\sqrt{0.784 + 1.067}} \approx 1.47$$

$z_\alpha = 1.96$ at 5% los (two tailed)
Calculated $z = 1.47$

Since $z_{calculated} = 1.47 < z_{critical} = 1.96$
we fail to reject null hypothesis.

There is no significant difference betw the two mean scores.

Q.49) Samples of size 10 and 14 were taken from two normal populations with S.D 3.5 & 5.2, Sample means is 203.0 & 186 find whether the means of two population are at same level, [Given $t_{22, 0.05} = 2.07$]

→ Given:

$n_1 = 10$, $s_1 = 3.5$, $\bar{x}_1 = 20.3$

$n_2 = 14$, $s_2 = 5.2$, $\bar{x}_2 = 18.6$

Degrees of ~~22~~ freedom = 22

Critical value $t_{22, 0.05} = 2.07$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

test statistic: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \dfrac{20.3 - 18.6}{\sqrt{\frac{3.5^2}{10} + \frac{5.2^2}{14}}}$

$= 0.9586$

Calculated $t \approx 0.956$

Critical $t = 2.07$ at $df = 22$, $\alpha = 0.05$

Since $0.956 < 2.07$, we fail to reject H_0

∴ There is no significant difference between the two population means.

Q45> Find the F-statistics

Sample	Size(n)	Total observation Σx	Sum of Squares obs.
1	8	9.6	61.52
2	11	16.5	73.26

Q.46

→ i) Calculate sample means

$\bar{x}_1 = \frac{9.6}{8} = 1.2$, $\bar{x}_2 = \frac{16.5}{11} = 1.5$

ii) Calculate variance for each sample

$$S^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}$$

~~format~~ for sample $S_1^2 = \frac{61.52 - \frac{(9.6)^2}{8}}{8-1}$

$= 7.14$

Sample 2 : $S_2^2 = \frac{73.26 - \frac{(16.5)^2}{11}}{11-1}$

$= 4.\cancel{73}851$

iii) Calculate F-statistic

Since Sample 1 has the larger varian

$$F = \frac{S_1^2}{S_2^2} = \frac{7.14}{\cancel{4.73}\ 4.851}$$

$$\boxed{= 1.47}$$

Q.46) Random Sample of 400 men & 600 women were asked whether they would have a school near their residence, 200 men & 325 women were in favor of proposal. Test the hypo. that proportion of men & women in front of proposal is same at 5% los. ($z\alpha = 1.96$ at 5% los.)

→ i) $H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$

ii) For men: $n_1 = 400,\ x_1 = 200 \Rightarrow P_1 = \frac{200}{400} = 0.5$
for women: $n_2 = 600,\ x_2 = 325 \Rightarrow P_2 = \frac{325}{600} = 0.547$

iii) pooled proportion: $p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{525}{1000} = 0.525$
$q = 1 - p = 0.475$

iv) Standard error (S.E.):
$$S.E. \sqrt{p \cdot q\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$= \sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}$$
$$= 0.0322$$

v) Z-statistics : $Z = \frac{P_1 - P_2}{SE}$
$$= \frac{0.5 - 0.5417}{0.0322} = -1.295$$

vi) Conclusion : $z\alpha = 1.96$
Since $|z| = 1.295 < 1.96$, we fail to reject H_0
∴ There is no significant difference between the proportions of men and women in favor of the proposal at 5% L.O.S.

Q.47) The values given below are:
i) Observed frequencies dist
ii) Expected value
apply x^2 test of goodness-of-test

a)	1	5	20	28	42	22	15	5	2
b)	1	6	18	25	40	25	18	6	1

(Given $x^2 = 12.592$ at 5% l.o.s)

→ $$x^2 = \Sigma \frac{(O_i - E_i)^2}{E}$$

Category	O(a)	E(b)	$(O-E)^2$	$(O-E)^2/E$
1	1	1	0	0
2	5	6	1	0.16676
3	20	18	4	0.222
4	28	25	9	0.36
5	42	40	4	0.1
6	22	25	9	0.36
7	15	18	9	0.5
8	5	6	1	0.1666
9	2	1	1	1

Total $x^2 = 2.8754$

Calculated $x^2 = 2.88$ (approx.)
Critical value $x^2_{0.05,8} = 12.592$
Since $2.88 < 12.592$, we fail to reject the null hypothesis.

∴ There is no significant difference between the observed and expected freq.
The observed data fits the distribution well.

Q.48) Fertilizers | yeilds

A	8.0	7.6	8.2	7.8	8.3	8.4	8.2	7.8	7.1	8.0
B	7.4	8.1	7.6	8.1	7.5	7.6	7.3	7.2	–	–

(Given: $t_{0.05} = 2.201$ at dof 16)
Test using t-test whether in effects of the fertilizer as reflected in the mean

→ $H_0 = \mu_1 = \mu_2$
$H_1 = \mu_1 \neq \mu_2$

fertilizer A ($n_1 = 10$)
fertizer B ($n_2 = 8$)
means & variances
$\bar{x}_1 = \frac{79.4}{10} = 7.94$

$s_1^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{1.344}{9} = 0.493$

$\bar{x}_2 = \frac{60.8}{8} = 7.60$

$s_2^2 = \frac{0.8}{7} = 0.1142$

pooled variance
$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(9)(0.493)+(7)(0.1142)}{16}$
$= 0.1339$

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{7.94-7.60}{\sqrt{0.1339\left(\frac{1}{10}+\frac{1}{8}\right)}} = 1.958$

calculated $t = 1.958 < 2.120$, we fail to reject H_0
∴ There is no significant diff. in the mean yeilds due to the two fertlizers at 5% level of s.

Q49) The average marks in mathematics of a Sample of 100 Students was 51. with S.D = 6 marks. Could this have a random sample from the population with average marks 50%

→ Given:

Sample mean $\bar{x} = 51$

population mean $\mu = 50$

Sample standard deviation $s = 6$

Sample size $n = 100$

Significances level = 5%

$z_\alpha = 1.96$

H_0 : Sample is from a population with $\mu = 50$

H_1 : Sample is not from a population with $\mu \neq 50$

Test Statistic

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{51 - 50}{6/\sqrt{100}} = 1.6667$$

Decision Rule

Critical z-value at 5% level = 1.96

Calculated z = 1.67

since 1.67 < 1.96, we fail to reject H_0

∴ The observed difference is not significant at 5% l.o.s.

So this Could be a random sample from a population with average markss 50

Q.50) A Coin is tossed 160 times.

No. of heads	0	1	2	3	4
Expected freq.	17	52	54	31	6
Observed freq.	10	40	60	40	10

Find χ^2

→

No of heads	O	E	$(O-E)^2$	$(O-E)^2/E$
0	10	17	49	2.8823
1	40	52	144	2.7692
2	60	54	36	0.666
3	40	31	81	2.6129
4	10	6	16	2.666

$\Sigma(O-E)^2/E = 11.597$

$\therefore \chi^2 = 11.597$

Q.51) A Random sample of 16 new comers gave a mean of 1.67m & S.D 0.10m. Is the mean height of newcomers significantly different from the mean height of student population of previous year? ($t_{0.05,15} = 2.13$) (we assume $\mu_0 = 1.75m$) p.y. mean height

→ $n = 16$, $S = 0.10$, $df = n-1 = 15$

Sample mean $\bar{x} = 1.67$, Critical value $t_{0.05,15} = 2.13$

$H_0 : \mu = \mu_0$ (no diff)

$H_1 : \mu \neq \mu_0$ (Significant diff)

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{1.67-1.75}{0.16/\sqrt{16}} = \frac{-0.08}{0.04} = -2.0$$

Since $|t| = 2.0 < 2.13$, we fail to reject H_0

There is no Significant diff. betn height of newcomers

Chi-square$(0.05, 5) = 11.071$

Q.51>

Day	No. of books	E	$(O-E)^2$	$(O-E)^2/E$
			0	0
Mon	120	120	100	0.8333
Tue	130	120	100	0.8333
Wed	110	120	25	0.2083
THU	115	120	225	1.875
Fri	135	120	100	0.8333
Sat	110	120		

H_0: The no. of books issued is same on each

H_1: The no. of books issued is not same on each

Total books issued $= 720$

No. of days 6

Expected freq. for each day $= \frac{720}{6} = 120$

Total $\chi^2 = \boxed{4.5832}$

Since $4.5832 < 11.071$, we fail to reject H_0

∴ there is no significant diff. in no. of books issued across different days at 5% lo

Q.52> A random sample of 900 members has mean 3.4 →
Can it be reasonable regarded as a sample from a large population of mean 3.2cms & S.D 2.3

→ $n = 900$, $\bar{x} = 3.4^{cm}$, $\mu = 3.2^{cm}$, $\sigma = 2.3$cm

H_0: Sample comes from the population w/ mean

H_1: Sample does not come from that pop

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.4 - 3.2}{2.3/\sqrt{900}} = 2.6086$$

At 5% level of significance, critical z value ±

Since $|z| = 2.61 > 1.96$, we reject the null hypo

∴ The sample cannot be reasonably regarded as coming from the population with mean 3.2 at 5% ...

Q.53)

Days	Mon	Tue	Wed	Thu	Fri	Sat	(Chi-square)$_{0.05,5}$
accidents	14	18	12	11	15	14	= 11.09

→ Ho: Accidents are uniformly distributed
H1: Accidents are not uniformly distributed
$\chi^2 = \sum (O-E)^2/E$

Expected freq = $14+18+12+11+15+14/6 = 14$

$= \frac{(14-14)^2}{14} + \frac{(18-14)^2}{14} + \frac{(12-14)^2}{14} + \frac{(11-14)^2}{14} + \frac{(15-14)^2}{14} + \frac{(14-14)^2}{14}$

$= 2.1428$

Since $\chi^2_{calculated} = 2.1428 < 11.09$, fail to reject Ho
∴ There is no significant evidence to suggest that the no. of accidents is not uniformly distributed over the week. Thus the dist. of accidents can be considered uniform.

Q.54) A normal population has mean 6.8, S.D: 1.5. A sample of 400 members gave a mean of 6.75 Is the difference significant? $Z\alpha = 1.96$ at 5% level of significance.

→ $\mu = 6.8$, $\sigma = 1.5$, $n = 400$, $\bar{x} = 6.75$, sig. level = 5%

$H_0 = \mu = 6.8$

$H1 = \mu \neq 6.8$ $\quad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{6.75-6.8}{1.5/\sqrt{400}}$

$= -0.6667$

$|z| = 0.667 < 1.96$

∴ we fail to reject the null hypothesis
∴ The difference between sample mean & population mean is not significant. at 5% l.os.
∴ The sample can be regarded as coming from the population with mean 6.8

Q55) Suppose that sweet are sold in packages fixed weight contents. The procedure of the packages is interested in testing the average weight of content in package in 1 kg. Sum of squares of deviations from mean of 12 samples is 0.011967. Using above data should we conclude the average. Given $\bar{x} = 0.9883$, $t_{0.05, 11} = 2.201$.

→ $H_0: \mu = 1$, $H_1: \mu \neq 1$

Given:

$\bar{x} = 0.9883$

$\mu = 1$

$\Sigma(x-\bar{x})^2 = 0.011967$

$n = 12$

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{0.011967}{11}} = 0.03298$$

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{0.9883-1}{0.03298/\sqrt{12}} = \cancel{0.03187}$$

$$= -1.2289$$

Since

$|t| = 1.2289 < t_{0.05} = 2.201$ d.f 11

Ho is accepted at 5% L.o.s.

∴ The average weight of contents of package taken as 1 kg.

Q.56) A set of five similar coins is tossed 210 times and the result is given in the following table. Use Chi-square test to test the hypo that data follows a bionomial dist. (chi-square = 11.07 at 5% (0.05))

No. of heads	0	1	2	3	4	5
frequency	2	5	20	60	100	31

H_0 : Coins follow bionomial dist. H_1 : Does not follow ......

→ No. of trials $N = 210$, tosses of 5 coins $n = 5$

Assume fair coins $\Rightarrow p = 0.5,\ q - 1 = 0.5$

Bionomial prob. formula : $P(x) = \binom{n}{x} p^x \cdot q^{n-x}$

x	O	E	$(O-E)^2/E$
0	2	6.56	3.1697
1	5	32.81	23.57
2	20	65.62	31.71
3	60	65.62	0.48
4	100	32.81	137.59
5	31	6.56	91.053
			287.5736

$\chi^2 = 287.5736$

Since

$\chi^2 = 287.57 > 11.07$, we reject the H_0

$\therefore$ The given data does not follow a bionomial dist. at 5% L.o.s.

Q.57) for the given data below, intelligence tests of two groups of boys and girls gave the following results. Examine the difference in significance.

Given $ZA = 1.96$ at 5% Los.

	mean	SD	size
Girls	70	10	70
Boys	75	11	110

→ Given: mean for girls $(\bar{x}_1) = 70$

S.D for girls $(\sigma_1) = 10$

Sample size for girls $(n_1) = 70$

mean for boys $(\bar{x}_2) = 75$

S.D. for boys $(\sigma_2) = 11$

Sample size for boys $(n_2) = 110$

Significance level $\alpha = 0.05$, $z_{0.05} = 1.96$

H_0: There is no significant diff. in intelligence scores

$\mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$S.E = \sqrt{\frac{\sigma_1^2}{n_1} \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10^2}{70} + \frac{11^2}{110}} = 1.59$$

Compute z-value: $z = \frac{\bar{x}_1 - \bar{x}_2}{SE} = \frac{70-75}{1.59} \approx -3.1446$

Since $|z| = 3.14 > 1.96$, we reject the null hypothesis.

∴ There is significant difference in intelligence scores between boys & girls at the 5% l.o.s.

All The Best !!

- Karan Salunkhe ☺