Practical-I

- **Aim:** To construct the SPRT for the given parameters of normal distribution & draw decision line, OC & ASN curves.

- **PROBLEM:** Construct SPRT to test $H_0: \theta_0 = 40$ against $H_1: \theta_1 = 45$ is a random sample drawn from the $N(\theta, 100)$. take $\alpha = 0.01$ & $\beta = 0.3$. Draw decision lines OC & ASN curve.

- **Theory & Formula:** SPRT for testing $H_0: \theta = \theta_0$ V/s $\theta = \theta_1$ $\forall$ $\theta_1 > \theta_0$

For m observations, we have-

(i) Accept H_0 & stop taking additional observation if

$$\sum x_i \leq \frac{\sigma^2 \log B}{\theta_1 - \theta_0} + \frac{m(\theta_1 + \theta_0)}{2} \quad \text{where } B = \frac{\beta}{1-\alpha}$$

(ii) Reject H_0 & stop taking additional observation if

$$\sum x_i \geq \frac{\sigma^2 \log A}{(\theta_1 - \theta_0)} + \frac{m(\theta_1 + \theta_0)}{2} \quad \forall \; A = \frac{1-\beta}{\alpha}$$

(iii) Continue sampling if-

$$\frac{\sigma^2 \log B}{\theta_1 - \theta_0} + \frac{m(\theta_1 + \theta_0)}{2} \leq \sum x_i \leq \frac{\sigma^2 \log A}{\theta_1 - \theta_0} + \frac{m(\theta_1 + \theta_0)}{2}$$

→ For OC function.

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \; ; \quad \text{where } h(\theta) = \frac{\theta_1 + \theta_0 - 2\theta}{\theta_1 - \theta_0}$$

→ For ASN function.

$$E(n) = \frac{L(\theta) \log B + (1 - L(\theta)) \log A}{E(Z)}$$

$$\forall \; E(Z) = \frac{2\theta(\theta_1 - \theta_0) + (\theta_0^2 - \theta_1^2)}{2\sigma^2}$$

- Calculation: As done in excel
- Result: Hence, we have plotted OC curve, decision line and ASN curve in the graph.

15/11/24

Practical-2

- **Aim:** To construct SPRT for the Bernoulli distribution with parameter θ. & draw decision line, OC & ASN curve.

- **Problem:** Construct SPRT to test $H_0: \theta = 0.05$ V/s $H_1: \theta = 0.1$ is random sample drawn from the bernoulli popn. with parameter θ. Take $\alpha = 0.08$ & $\beta = 0.01$. Draw decision lines, OC, ASN curves.

- **Theory & formula:** Here null hypothesis $H_0: \theta_1 = 0.05$ V/S alt. hypothesis $H_1: \theta_1 = 0.1$ &
$\alpha = 0.08$
$\beta = 0.01$

SPRT for testing observation $H_0: \theta = 0.05$ V/S $H_1: \theta = 0.01$ after m observations.

(i) terminates the process & accept H_0 if

$$\sum_{i=1}^{m} z_i \le \log B.$$

$$\Sigma x_i \le \frac{\log B - m\log\left(\frac{1-\theta_1}{1-\theta_0}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right) - \log\left(\frac{1-\theta_1}{1-\theta_0}\right)} = A_m.$$

(ii) Continue taking obs. if-

$$\log B < \sum_{i=1}^{m} z_i < \log A$$

where $A = \frac{1-\beta}{\alpha}$, $B = \frac{\beta}{1-\alpha}$

$$A_m < \Sigma x_i < R_m$$

~~for OC function~~

$$\cancel{L(\theta)} = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

We regard h as a parameter & solve for θ

$$\theta = \frac{1 - \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)}}{\left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} - \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)}} = \theta(h)$$

$$L(\theta) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} = L(\theta, h)$$

→ ASN function :-

$$E(n) = \frac{L(\theta) \log B + (1 - L(\theta)) \log A}{E(z)}$$

$$E(z) = \theta \log \frac{\theta_1}{\theta_0} + (1-\theta) \log \left(\frac{1-\theta_1}{1-\theta_0}\right)$$

$$E(n) = \frac{L(\theta) \log B + (1 - L(\theta)) \log A}{\theta \log \frac{\theta_1}{\theta_0} + (1-\theta) \log \left(\frac{1-\theta_1}{1-\theta_0}\right)}$$

- Calculation: as done in excel sheet.

- Result: Hence, we have plotted decision line, OC & ASN curves graphically.

15/4/24

Practical-3

- **Aim:** To construct SPRT for exponential sample & draw decision line, OC & ASN Curve.

- **Problem:** Construct SPRT to test $H_0: \theta = 6$ against $H_1: \theta = 3$ is a random sample drawn from the exponential distribution with parameter θ. Take $\alpha = 0.1$ & $\beta = 0.3$. Draw decision line, OC & ASN curves

- **Theory & Formula:** For exponential distribution, the SPRT for testing H_0 v/s H_1 is defined as follow-

After taking m observation-

(i) terminate the process & accept H_0 if -

$$\sum x_i \geq \frac{m \log\left(\frac{\theta_1}{\theta_0}\right) - \log B}{\theta_1 - \theta_0} = A_m \quad ; \; B = \frac{\beta}{1-\alpha}$$

(ii) Terminate the process & reject H_0 if-

$$\sum x_i \leq \frac{m \log\left(\frac{\theta_1}{\theta_0}\right) - \log A}{(\theta_1 - \theta_0)} = R_m.$$

(iii) Continue sampling if-

$$R_m \leq \sum x_i \leq A_m$$

→ For OC function-

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

$$\theta = \frac{(\theta_1 - \theta_0)\, h(\theta)}{\left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} - 1}$$

ASN function-

$$E(n) = \frac{L(\theta) \log B + (1 - L(\theta)) \log A}{E(Z)}$$

where $E(Z) = \log\left(\frac{\theta_1}{\theta_0}\right) + \frac{(\theta_0 - \theta_1)}{\theta}$

- **Calculation:** As done in excel sheet.
- **Result:** Hence, we have plotted the decision line, OC & ASN curves graphically.

15/4/24

Practical : 4

- **Aim:** To construct SPRT for Poisson sample & draw decision lines, OC, & ASN curves.

- **Problem:** Construct SPRT to test $H_0: \theta = 3$ & $H_1: \theta = 4$ in a random sample drawn from poisson distn. with parameter θ. Take $\alpha = 0.08$ & draw decision lines, OC, ASN curves.

- **Theory & formula:-** For poisson dist, the SPRT for deciding 2 alternative H_1 & H_0 is defined as—

After m observations—

(i) Terminate the process & accept H_0 if

$$\sum_{i=1}^{m} x_i \leq \frac{\log B + m(\theta_1 - \theta_0)}{\log(\theta_1/\theta_0)} = A_m \quad ; \quad B = \frac{\beta}{1-\alpha}$$

(ii) Terminate the process & accept H_1 if

$$\sum_{i=1}^{m} x_i \geq \frac{\log A + m(\theta_1 - \theta_0)}{\log\left(\frac{\theta_1}{\theta_0}\right)} = R_m \quad ; \quad A = \frac{1-\beta}{\alpha}$$

(iii) Continue sampling if—

$$R_m \leq \sum_{i=1}^{m} x_i < A_m.$$

→ Construction of OC function—

$$\theta = \frac{h(\theta_1 - \theta_0)}{\left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} - 1} \quad ; \quad L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

→ for ASN curve →

$$E(n) = \frac{L(\theta) \log B - \log A(1 - L(\theta))}{E(Z)}$$

where, $E(Z) = \theta \log\left(\frac{\theta_1}{\theta_0}\right) - (\theta_1 - \theta_0)$

- **Calculation:** As done in excel sheet.

- **Result:** Hence, we have plotted the decision line, OC & ASN curves.

15/4/24

Practical-5

- **Aim:** To draw decision lines, OC & ASN curves for given exponential distribution data.

- Problem: Let 0.11236, 0.26316, 0.11837, 0.21327, 0.0119, 0.16747. be a random sample from exponential dist. with parameter θ use SPRT procedure to test $H_0: \theta = 5$ v/s $H_1: \theta = 3$ when $\alpha = 0.1$ & $\beta = 0.3$. Draw decision lines, OC & ASN curves.

- Theory & formula: The SPRT for testing m observations-

(i) terminates the process & accept H_1 if-

$$\Sigma x_i \geq \frac{\log\left(\frac{1-\beta}{\alpha}\right) - m\log\left(\frac{\theta_1}{\theta_0}\right)}{\theta_0 - \theta_1} = A_m$$

(ii) Terminate the process & reject H_1 if-

$$\Sigma x_i \leq \log\left(\frac{1-\beta}{\alpha}\right) - m\log\left(\frac{\theta_1}{\theta_0}\right) = R_m.$$

(iii) Continue taking observation.

$$R_m \leq \Sigma x_i \leq A_m.$$

For OC function:

$$E\left[\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}\right]^{h(\theta)} = 1$$

$$\theta = \frac{(\theta_1 - \theta_0)\ h(\theta)}{\left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} - 1}$$

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

for ASN function:-

$$E(N) = \frac{L(\theta)\log B + (1 - L(\theta))\log(A)}{E(Z)}$$

$$E(Z) = \log\left(\frac{\theta_1}{\theta_0}\right) + \left(\frac{\theta_0 - \theta_1}{\theta}\right)$$

- Calculation:

as done in excel.

- Result:

Hence, we have plotted decision line, OC & ASN curve.

Practical-6

- **Aim:** To apply Run test.
- **Experiment:** Test the randomness of following sample size 30 using Run test.

15	69	76	28	86	85
77	58	20	26	06	40
01	40	00	46	17	51
63	81	84	66	43	40
19	16	22	31	49	10

For testing.

H_0: sample values come from a random sequences.

H_1: Sample values come from a non-random sequence.

- **Theory & formula:**

For a Run test,

Let one kind of letters denote sample size 'n_1' & second kind of letters denote sample size 'n_2'

Total sample size $= n_1 + n_2$

Compare the observed no. of runs (R) in the sample with two critical values of R for given values of n_1 & n_2 at a pre-determined level of significance.

- **Decision Rule:**

Reject H_0, if $R \leq C_1$ or $R \geq C_2$

Otherwise, Accept H_0,

Here C_1 & C_2 are two critical values.

We obtain these values from table of critical values for diff values of n_1 & n_2

- **Calculations:** Performed in SPSS software.

- **Result:-** Since P value is greater than 0.05, so Accept H_0.

Practical-7

- **Aim:** To perform Run test.
- **Problem:** the win-lose, record of a certain basketball team for their last 50 consecutive game was follows.

WWWWWW L WWWWWW L WLWWWLL W WWW LWWWLL WWW WWW LLLWWLL L WLWL WWW.

Apply Run test to test that sequence of wins & losses is random.

- **Theory & formula:-** For the sample size, i.e n_1 or n_2 is greater than 20. The sampling distn of R can be approximate by normal distn. with mean & variance.

Here, $n_1 = 3l$, $n_2 = 17$ } use Large ~~small~~ sample run test.

- **Formulas:-**

$$E[R] = \frac{2n_1 n_2}{n_1+n_2} + 1$$

$$V[R] = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}$$

$$Z = \frac{R - E[R]}{\sqrt{V[R]}}$$

Decision Rule:- Accept H_0, if $Z_{cal} < Z_{tab}$
Reject H_0, if $Z_{cal} > Z_{tab}$.

Practical-8

- **Aim:** To perform one-sample sign test.
- **Problem:** The following table represents observations on heights & weights of 15 females using sign test we have to test the following 2 hypothesis:

(i) the Height of the females can be taken to be equal to 64 inches.

(ii) The weight of the females can be taken to be 140 lbs.

Heights	weights.
58	115
59	117
60	120
61	123
62	126
63	132
64	139
65	142
66	146
67	150
68	154
69	159
70	154
71	159
72	164

- **Theory & formula:**

Let $x_1, x_2, x_2 - x_n$ be the given set of observation

(i) Then compute $D_i = x_i - U_{e0}$, $i = 1, 2 -- n$.

$U_{e0} \to$ given popn. median

(ii) Compute r = no. of D_i's greater than 0.

(iii) Compute s = no. of D_i's less than 0, ignore D_i's $= 0$

(iv) set up hypothesis:

H_0; $U_e = U_{e0}$ (median of given sample is equal to given popn. median].

H_1: $U_e > U_{e0}$ [median of given sample is greater to given popn median]

(V) Compute p-value.

if p-value $> 0.05 \rightarrow$ Accept H_0

p value $< 0.05 \rightarrow$ Reject H_0

· Calculations:-

As performed in SPSS file,

Go to Analyze → Non parametric test → → Binomial

· Results:

Practical-9

- **Aim:** To perform two sample Sign tests.
- **Problem:** The following data represents the marks given to the sample set of 8 students by two diff. professors in same exam.
  Using sign. test, we have to test if the grading of both the profs. can be taken to be same.

Student	Prof A	Prof B.
1	74	75
2	78	80
3	68	87
4	72	81
5	76	72
6	69	73
7	71	80
8	74	76

- **Theory & formula:-**

Let $x_1, x_2 - - x_n$ (Prof A) & $y_1, y_2 - - y_n$ (Prof B) be the given set of observations.

(i) compute $D_i = x_i - y_i$, $i = 1, 2, - - n$, where ν_0 is given popl^n median.

(ii) Compute, S = no. of D_i's less than 0. Ignore all D_i's $= 0$

(iv) Set up hypothesis,

H_0: Grading of both the prof. is same.
H_1: Grading of both profs not same

(v) Compute p-value.

if p-value > 0.05 → Accept H_0 at 5% LoS.

if pvalue < 0.05 → Reject H_0 at 5% LoS.

- **Calculations:** As performed in SPSS file.

Go to analyze → Non-parametric test → Legacy → 2 paired samples → sign test.

Practical-10

- **Aim:** To apply wald wolfowitz test.
- **Problem:** Following is the data for prices in rupees of a certain commodity in a sample of 15 randomly selected shops from city A & those of 15 randomly selected shops from city B.

City A (price in Rs.)	City B
7.41	7.08
7.77	7.49
7.44	7.92
7.40	7.04
7.38	6.92
7.93	7.22
7.58	7.68
8.28	7.24
7.23	7.74
7.52	7.81
7.82	7.28
7.71	7.83
7.84	7.47.
7.63	
7.68	

- **Theory & formula:**

H_0: Prices in city A & city B have same prob. distn.
H_1: Prices in city A & city B dont have same distn.

Test statistic:

- Let r denote no. of runs to obtain r, list the n_1+n_2 obs. from two samples in order of magnitude. Denote observations from one sample by x's and other by y's count no. of runs.
- $r_c \Rightarrow$ critical value of r.
- if $r \leq r_c$, reject H_0

if Tie, in case x and y observations have same value, place observation x/y first if run of observation x/y is continuity.

for large samplesize ($n_1, n_2 > 20$)

$$r_c = \mu - 1.96\sigma \text{ at } 5\% \text{ LOS.}$$

$$\mu = 1 + \frac{2 n_1 n_2}{n_1 + n_2}$$

$$\sigma = \sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}}$$

if $r < r_c$, reject H_0.

if $(n_1 \geq 10,\ n_2 \geq 10)$

then we use normal test

1. set up H_0, H_1
2. complete U_x, U_y
3. $z = \frac{U - \mu_U}{\sigma_U} \sim N(0,1)$

where $U = \min(U_x, U_y)$

$$\mu_U = \frac{n_1 n_2}{2}$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

if $z > z_{\alpha/2}$ then reject H_0

- Calculations:

Done in SPSS

p value = 0.539 > 0.05
So we accept H_0

- Result:

Both teaching methods are equally effective.

Problem-11

- **Aim:** To apply mann-whitney test.
- **Problem:** Marks in Reading ability test

First Group	Second Group.
227	202
176	14
252	165
194	171
16	292
55	271
237	151
194	235
247	147
92	99
184	63
147	284
88	53
161	228
171	271

Apply U-test to test if two diff - teaching methods can be taken as equally effective or not.

- **Theory & formula:**

H_0: 2 diff teaching methods are equally effective
H_1: 2 diff teaching methods are not equally effective

Assumption: Median variables are continous

Procedure: Let T_x: Sum of ranks in x sample.
T_y: Sum of ranks in y sample.

Mann whitney
$$\begin{cases} U_x = T_x - \dfrac{n_1(n_1+1)}{2} \\ U_y = T_y - \dfrac{n_2(n_2+1)}{2} \end{cases}$$

$n_1 \to$ no. of obs. in 1st sample
$n_2 \to$ no. of observation in 2nd sample

$U = \min(U_x, U_y)$

Remarks: i) $U_x + U_y = n_1 n_2$

ii) $T_x + T_y = \dfrac{(n_1+n_2)(n_1+n_2-1)}{2}$

if $U \le T_c$ then reject H_0 ($T_c \to$ critical value of statistic)

• Calculation:

Done in spss.

Here p value

• Result:

Prices in city A & city B have same probability distn.

Problem - 12

- **Aim:** To apply Kruskal Wallis test.
- **Problem:** To identify whether samples arise from common parent population.

Sample 1	Sample 2	Sample 3
1.7	13.6	13.4
1.9	19.8	20.9
6.1	25.3	25.1
12.5	46.2	29.1
16.5	46.2	29.7
25.1	61.1	25.7
30.5		
42.1		
82.5		

Apply Kruskal Wallis test.

- **Theory & formula:**

H_0: All medians are equal $(\mu_1 = \mu_2 = \cdots = \mu_k)$

H_1: Atleast one median is not equal.

- Kruskal Wallis test is non parametric alternative to one way ANOVA test when normality and equality of variance assumptions are violated.

Assumptions: i) There are atleast three independently drawn random samples.
ii) Each sample has atleast 5 observations, $[n_i \geq 5]$ ~~observations~~

Sample	Observation
1	
2	
i	
k	

k: no. of samples
n_i: Size of ith sample $(i = 1, 2 \cdots k)$

$n = n_1 + n_2 + \cdots n_k$ Total no. of obs. in k samples.

R_i: Sum of ranks of ith sample

- Test statistic (under H_0)

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{r_i^2}{n_i} - 3(n+1) \sim \chi^2_{(k-1)}$$

- if $H > \chi^2_{(k-1)}(\alpha)$ we reject H_0

- Calculations.
  - Done in SPSS
    - P value =
      $\Rightarrow$
- Results:

All medians are equal.

Practical-13

- **Aim:** To apply KS one sample test.

- **Problem:** For following samples test if they are drawn from Normal, Exp, Poisson & Uniform distn. respectively

Obs.	Sample 1	Sample 2	Sample 3	Sample 4.
1	1.089	0.046	4	10.25
2	1.9627	0.02969	2	10.125
3	1.72	0.0138	1	17.417
4	1.63	0.01497	1	18.87
5	0.14	0.2168	2	15.32
6	0.23	0.549	3	11.97
7	1.68	0.0075	4	16.79
8	3.63	0.147	3	14.9
9	1.81	0.29	3	14.105
10	1.68	0.027	4	14.99
11	1.65	0.163	3	12.68
12	0.83	0.81	1	10.62
13	1.42	0.078	4	15.598
14	1.12	0.153	4	17.59
15	1.55	0.016	5	10.602
16	0.24	0.086	2	17.17
17	0.47	0.501	4	11.59
18	3.013	0.104	5	14.11
19	2.73	0.02	0	19.6
20	2.58	0.153	5	17.203

- **Theory & formula:**

H_0: $F_0(x) = F_n(x)$

(there is no difference b/w emperical distn. $F_0(x)$ and theoritical distn. $F_n(x)$)

H_1: $F_0(x) \neq F_n(x)$

Test statistic

$D = \max |F_0(x) - F_n(x)|$

Step (3)

if $D >$ critical value of D (From K-S table)
⇒ Reject H_0

if $D <$ critical value
⇒ Accept H_0

- Calculations:

Done in SPSS.

- For normal Distn ⇒ p-value = 0.819 > 0.05 ⇒ Accept H_0.
- For exponential Distn ⇒ p value ⇒ 0.855 > 0.05 ⇒ Accept H_0
- For Uniform Distn ⇒ p value ⇒ 0.777 > 0.05 ⇒ Accept H_0.
- For poisson Distn ⇒ p value = 1 > 0.05 ⇒ Accept H_0.

- Results:

- Sample 1. follows Normal Distn.
- sample 2 follows Exp. Distn.
- sample 3 follows Uniform Distn
- sample 3 follows Poisson Distn.

Practical-14

- **Aim:** To apply median test.

- **Problem:** The scores of two groups of students are given below:

  G1 : 61, 64, 59, 53, 51, 52, 57, 54, 51, 54, 60, 62
  G2 : 60, 58, 57, 64, 67, 70, 73, 62, 52, 53, 52

  Apply median test to see if groups different significantly

- **Theory & formula:**
  - Median test is NPT used to compare medians of two independent samples.

Assumptions:

i) The observations are independent both within & between samples.
ii) The obs. come from popn with cond. distn fn.

H_0: $F_1(x) = F_2(y)$ (two samples have drawn from same popn)
H_1: $F_1(x) \neq F_2(y)$

Arrange data in ascending order and complete median of series.

$$\text{Median} \begin{cases} \left(\frac{N+1}{2}\right)^{th} & \text{if } N \text{ is odd} \\ \dfrac{\left(\frac{N}{2}\right) + \left(\frac{N}{2}+1\right)}{2} & \text{if } N \text{ is even.} \end{cases}$$

Construct 2×2 const table as.

	Sample 1	Sample 2
No. of obs ≥ Median	m_1	m_2
No. of obs < median	$n_1 - m_1$	$n_2 - m_2$
Total	n_1	n_2

- When either $n_1, n_2 < 10$ then under H_0:
  Joint distn. of m_1 & m_2 is hypergeometric distn. with prob. fn.

$$P = \frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n_1+n_2}{m_1+m_2}}$$

However if both $m_1, n_2 \geq 10$

then use χ^2 test for testing H_0

for 2×2 const. ratio.

$$\chi^2 = \frac{N\,(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

O.W.

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

in case any of cell freq. is less than 5,

Apply Yates correlation.

$$\chi^2 = \frac{N\left[\,|ad-bc| - N/2\,\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$

• Conclusion:

case 1:

• if $p < \alpha$ we reject H_0

case 2:

• if $\chi^2 > \chi^2_{(r-1)(c-1)}(\alpha)$, we reject H_0.

• Calculations:

Done in SPSS

p value =>

• Results:

Groups do not differ significantly and have been drawn from same popn.

Practical: 1(A)

- **Aim:** To find parameters of distribution of form bivariate & trivariate normal.

- **Problem:** Let $f(x) = ae^{-Q/2}$ where $Q = 3x^2 + 2y^2 - 2xy - 32x + 4y + 92$
and $Q = 2x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 - 2x_1x_3 - 4x_2x_3 - 6x_1 - 2x_2 + 10x_3 + 8$
Find μ & Σ

- **Theory:**

Multivariate distribution:

$$\text{Pdf} = f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{p/2}} \exp\left\{ -\frac{1}{2} (\underset{\sim}{x} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}) \right\}$$

$\underset{\sim}{x}$ is $N_p(\underset{\sim}{\mu}, \Sigma)$ where $\underset{\sim}{\mu} \rightarrow$ mean vector
$\Sigma \rightarrow$ Cov. matrix.

Bivariate Pdf:

$p = 2$

$$f(x) = \frac{1}{2\pi |\Sigma|^{n/2}} \exp\left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\}$$

Where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} ; \quad \underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} ; \quad \underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

Which reduces to form;

$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)} \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{x_2-\mu_2}{\sigma_2} \right) \right] \right.$$

~~Triv~~ • Trivariate Pdf:

$p = 3$

$$f(x) = \frac{1}{2\pi^{3/2} |\Sigma|^{3/2}} \exp\left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right]$$

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} ; \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}$$

- Calculations:

1) $Q = 3x^2 + 2y^2 - 2xy - 32x + 4y + 92$

$\frac{\partial Q}{\partial x} = 6x - 2y - 32 = 0$ — (1) $\Rightarrow 3x - y = 16$

$\frac{\partial Q}{\partial y} = 4y - 2x + 4 = 0$ — (2) $\Rightarrow -x + 2y = -2$

Solving (1) & (2) $\rightarrow$

$x = 6$, $y = 2$

$$\Sigma^{-1} = \begin{bmatrix} \text{coeff of } x_1^2 & \frac{1}{2}\text{ coeff of } x_1 x_2 \\ \frac{1}{2}\text{ coeff of } x_1 x_2 & \text{coeff of } x_2^2 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 2y \\ x_2 2y \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

So, $\sigma_1^2 = 1$ $\quad \sigma_2^2 = 3$ $\quad \rho\sigma_1\sigma_2 = 1$

$\sigma_1 = 1$ $\quad \sigma_2 = \sqrt{3}$ $\quad \rho = \frac{1}{\sqrt{3}}$

$\Rightarrow$ For Q_1 model is $N\left(6, 2, 1, 3, \frac{1}{\sqrt{3}}\right)$

for $Q_2 = 2x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 - 2x_1x_3 - 4x_2x_3 - 6x_1 - 2x_2 + 10x_3$

$\frac{\partial Q_2}{\partial x_1} = 4x_1 + 2x_2 - 2x_3 - 6 = 0$

$\frac{\partial Q_2}{\partial x_2} = 6x_2 + 2x_1 - 4x_3 - 2 = 0$

$\frac{\partial Q}{\partial x_3} = 8x_3 - 2x_1 - 4x_2 + 10 = 0$

Practical: 2(b)

- **Aim:** To find values of probability of certain problems.

- **Problem:** $x_1, x_2 \sim N(60, 75, 6^2, 12^2, 0.55]$

Find $P[x_1 \geq 70)$, $P[X_2 \leq 70)$ $P[65 \leq x_1 \leq 75]$

$P[71 \leq X_2 < 80 \mid X_1 = 55]$ & $P[\,|X_1 - X_2| > 15]$

- **Theory:** $X \sim BVN(60, 75, 36, 14, 0.55]$ $[\underset{\sim}{x} = [x_1, x_2]]$

Marginal of $x_1 \sim N(60, 75)$

Marginal of $x_2 \sim N(75, 144]$

$P[X_2 \mid X_1 = x] = N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), (1-\rho^2)\sigma_2^2]$

$x_1, x_2 \sim N(\mu - \mu_2, \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$

- **Calculation:**

① $P[X_1 > 70] = P\left[\frac{x_1 - 60}{6} > \frac{70 - 60}{6}\right] = 1 - \Phi[5/3]$

② $P[X_2 \leq 70] = P\left[\frac{x_2 - 75}{12} \leq \frac{70 - 75}{12}\right] = 1 - \Phi(5/12)$

③ $P[65 \leq x_1 \leq 75] = P\left[5/6 \leq Z < 10/6\right] = \Phi[5/3] - \Phi[5/6]$

④ $P[x_2 / x_1 = 55] \sim N\left[75 + 0.55 \cdot \frac{12}{6}[55 - 60], (1 - 0.55)^2 \cdot 12^2\right]$

$X_1 / X_2 \sim N[69.5, 100.44]$

$P\left[\frac{71 - 69.5}{\sqrt{100.44}} \leq Z \leq \frac{80 - 69.5}{\sqrt{100.44}}\right]$

$x_1 - x_2 \sim N[-15, 36 + 144 - 2(0.55) - 6.12)$

$\sim N(-15, 79.2)$

⑤ $P[\,|X_1 - X_2| > 15$

$\Rightarrow 1 - P[\,|X_1 - X_2| \leq 15] \Rightarrow 1 - P[-15 \leq Y \leq 15] \Rightarrow 1 - P\left[\frac{-30}{\sqrt{79.25}} \leq Y < 0\right]$

$\Rightarrow 1 - \left[\frac{1}{2} - \Phi\left[\frac{-30}{\sqrt{79.25}}\right]\right]$

- **Result:** All required values are calculated in the calculation.

Practical: 36

- **Aim:** To find the distn of given form of bivariates.

- **Problem:** $f(x,y) = K \exp\left[\frac{-1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right]$

Obtain K, distribution of $x+y$, distribution of x/y & y/x curve of regression of y on x.

- **Theory:**

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left[\frac{x_1-\mu_1}{\sigma_1}\right]\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right]$$

$X, Y \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

distn of X & $Y \sim BVN(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$

distn of $X/Y \sim BVN\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y-\mu_y), (1-\rho^2)\sigma_x^2\right)$

distn of $Y/X \sim BVN\left(\mu_2 + \rho\frac{\sigma_1}{\sigma_2}(x-\mu_1)(1-\rho^2)\sigma_x^2\right)$

Curve of Regression of Y on X →

$$E[Y/X] = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

Curve of regression of X on Y →

$$E[X|Y] = \mu_x + \frac{\rho\sigma_x}{\sigma_y}(y-\mu_y)$$

$X+Y$ & $X-Y$ are independent of each other iff X & Y follow distn with same variance & distn is $N(0,1)$ with correlation ρ.

- **Calculation:** $\iint f(x,y)\,dx\,dy = 1$

$$K\iint \exp\left[\frac{-1}{2(1-\rho^2)}(x^2+y^2-2xy)\right]dxdy = 1$$

$$\Rightarrow K\iint \exp\left[\frac{-1}{2(1-\rho^2)}(x^2+y^2+x^2\rho^2-y^2\rho^2-2\rho xy)\right]dxdy = 1$$

$$\Rightarrow K\int_{-\infty}^{\infty}\exp\left[\frac{-1}{2(1-\rho^2)}(x^2-x^2\rho^2)\right]\left(\int_{-\infty}^{\infty}\exp\left(\frac{-1}{2(1-\rho^2)}(y-x\rho)^2dy\right)\right)dx = 1$$

$$\Rightarrow K\sqrt{2\pi}(\sqrt{1-\rho^2})\int_{-\infty}^{\infty}\exp\left(\frac{-1}{2(1-\rho^2)}x^2(1-\rho^2)\right) = 1$$

$$\Rightarrow K\sqrt{2\pi}\sqrt{1-\rho^2}\int_{-\infty}^{\infty}\exp\left(\frac{-x^2}{2}\right) = 1$$

$$K = \frac{1}{2\pi\sqrt{1-\rho^2}}$$

Practical: 4(B)

- **Aim:** To solve the BVN question & find solutions for given problem.
- **Problem:** If $(X, Y) \sim BVN(0, 0, 16, 9, 0.8)$. Then find regression lines of Y on X & X on Y & Joint distn of X+Y & X−Y.
- **Theory:** $X, Y \sim BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$Y|X \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X), (1-\rho^2)\sigma_Y^2\right)$$

$$X|Y \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y), (1-\rho^2)\sigma_X^2\right)$$

→ Regression of line Y on x = $E[Y|X]$

→ Regression of lines x on y = $E[X|Y]$

If $[X, Y] \sim BVN[0, 0, 1, 1, \rho]$ that, $X \sim N(0,1)$ $Y \sim N(0,1)$ with correlated ρ, then $X+Y$ & $X-Y$ are independent distributed

$$f(X+Y, X-Y) = C_1 \exp\left[\frac{-(X+Y)^2}{4(1+\rho)}\right] C_2 \exp\left[\frac{-(X-Y)^2}{4(1-\rho)}\right]$$

- **Calculation:**

$X, Y \sim BVN(0, 0, 16, 9, 0.8)$

$E[Y/X] = x(0.8)\,3/4 \Rightarrow 0.6x$

$E[X/Y] = 0.8(4/3)\,Y \Rightarrow \frac{3.2}{3} \Rightarrow 1.066x$

Jt. distribution ⇒ $C_1 \exp\left[\frac{-(x+y)^2}{4(1.8)}\right] C_2 \exp\left[\frac{-(x-y)^2}{4(0.2)}\right]$

⇒ ~~Cexp~~

$C_1 \exp\left[\frac{-(x+y)^2}{7.2}\right] C_2 \exp\left[\frac{-(x-y)^2}{0.8}\right]$

- **Results:** Hence, we have obtained the required values

PRACTICAL – 5 (b)

Aim – To find multiple & partial correlation of randomly generated no.

PROBLEM – Generate 20 numbers randomly of following:-

1] N(50, 9) 2] U(40,60)

3] N[40,16] 4] U(10, 20)

Calculate the partial correlation coeff. between-

i] variable 1 & 2 partially due to correlation of 3 & 4 with both 1 & 2

ii] variable 1 & 3 partially due to correlation of 2 & 4 with both 1 & 3

iii] variable 1 & 4 partially due to correlation of 2 & 3 with both 1 & 4

Also calculate coefficient of multiple correlation between 1 & joint effect of 2, 3 & 4.

THEORY – Coefficient of multiple correlation-

$$R^2_{1.234} = 1 - \frac{\omega}{\omega_{11}} = \frac{r^2_{12} + r^2_{13} + r^2_{14} - 2r_{12}r_{13}r_{14}r_{23}r_{24}r_{34}}{1 - r^2_{23} - r^2_{24} - r^2_{34} + 2r_{23}r_{24}r_{34}}$$

$$\text{where } \omega = \begin{vmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{vmatrix}$$

$$\omega_{1,2} = \begin{vmatrix} 1 & r_{23} & r_{24} \\ r_{32} & 1 & r_{34} \\ r_{42} & r_{43} & 1 \end{vmatrix}$$

coefficient of partial correlation

$$r_{12.34} = \frac{+\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}} \qquad r_{13.24} = \frac{-\omega_{13}}{\sqrt{\omega_{11}\omega_{33}}} \qquad r_{14.32} = \frac{-\omega_{14}}{\sqrt{\omega_{11}\omega_{44}}}$$

CALCULATIONS –

1. Random no. generated of required distn using data analysis tool in excel.
2. Calculate correlation matrix b/w 4 variables.
3. We find ω, ω_{11}, ω_{12}, ω_{13}, ω_{14}, ω_{22}, ω_{33}, ω_{44}
4. find partial correlation coeffi.

$R_{12.34} = -0.27325$, $R_{13.24} = 0.2017$, $R_{14.32} = 0.35126$

Find multiple correlation coefficient as $R^2_{1.234} = 0.096125$

RESULT All required values are calculated above.

PRACTICAL 6(b)

AIM - Calculate the population principal components.

PROBLEM - To find the population principal components b/w three r.v. namely Y_1, Y_2, Y_3 having covar. matrix as-

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

THEORY - A PCA is concerned with explaining the variance-covariance structure of a set of variables through a few ~~~~ linear combination of these variables.

Result 1:

Let Σ be a cov. matrix associated with random vector $x' = [x_1\ x_2 - - x_p]$. let Σ have the eigen value - eigen vector pairs $(\lambda_1, e_1), (\lambda_2, e_2) \cdots (\lambda_p, e_p)$ where, $\lambda_1 > \lambda_2 - - > \lambda_p > 0$. Then, i^{th} principle component. is given by,

$$Y_i = e_i' X = e_{i1} X_1 + e_{i2} X_2 - - - + e_{ip} X_p \quad ; \quad i = 1, 2 - - p.$$

$$Var(Y_i) = e_i' \Sigma e_i = \lambda_i \quad ; \quad i = 1(1)p.$$

$$cov(Y_i, Y_k) = e_i' \Sigma e_k = 0 \quad ; \quad i \neq k.$$

Result 2:

if $Y_1 = e_1' X,\ Y_2 = e_2' X, - - - Y_p = e_p' X$

A the principal components obtained from the covariance matrix Σ, then

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad ; \quad i, k = 1(1)p$$

& the correlation coefficient b/w the components Y_i & the variable X_k.

$(\lambda_1, e_1), (\lambda_2, e_2) - - (\lambda_p, e_p)$ are the eigen value - vector pairs for Σ

CALCULATION - Eigen value → Eigen vector pairs are

$\lambda_1 = 5.83 \qquad e_1' = [0.383, -0.924, 0]$

$\lambda_2 = 2.00 \qquad e_2' = [0, 0, 1]$

$\lambda_3 = 0.17 \qquad e_3' = [0.924, 0.383, 0]$

Principle components are-

$$Y_1 = e_1' X = 0.383 X_1 - 0.924 X_2$$

$$Y_2 = e_2' X = X_3$$

$$Y_3 = e_3' X = 0.924 X_1 + 0.383 X_2$$

X_3 is one of p.c. because it is uncorrelated with other 2 variable.

$Var(Y_1) = Var[0.383 x_1 - 0.924 x_2)$
$= 0.583 = \lambda_1$

$Cov(Y_1, X_2) = cov(0.383 x_1 - 0.924 x_2, x_3)$
$= 0.$

$\sigma_{11} + \sigma_{22} + \sigma_{33} = 1 + 2 + 5 = 8$

$\lambda_1 + \lambda_2 + \lambda_3 = 5.83 + 2 + 0.17 = 8$

→ Proportion of total var accounted for by first.

principal component $= \dfrac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \dfrac{5.83}{8} = 0.73.$

→ First two component ~~component~~ account for proportion $= \dfrac{5.83+2}{8} = 0.98$ of popu var.

→ Y_1 & Y_2 could replace the original three variables

Now, $\rho_{Y_1, X_1} = \dfrac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \dfrac{0.388\sqrt{5.83}}{\sqrt{1}} = 0.925$

$\rho_{Y_1, X_2} = \dfrac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \dfrac{-0.924\sqrt{5.83}}{\sqrt{5}} = -0.998$

→ X_2 contributes more to the determination of Y_1 than does X_1

& $\rho_{Y_2 X_1} = \rho_{Y_2 X_2} = 0$ & $\rho_{Y_2 X_3} = 1 = \dfrac{\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}}$

→ Third component is unimportant.

RESULT - Hence, Principal components are X_1 & X_2.