Institute of Actuaries of India

Subject CT3 – Probability & Mathematical Statistics

May 2014 Examinations
Indicative Solutions

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other approaches leading to a valid answer and examiners have given credit for any alternative approach or interpretation which they consider to be reasonable.

Sol. 1 : The sizes of corpus on retirement (in ₹'000) in a DCS Pension Scheme for the workers of an auto component unit are:

208 232 226 219 235 220 234 226 213 220 231 210 218 222 227

i)

$$\text{Mean} = \frac{3{,}341}{15} = 222.73$$

As n = 15, median = $\frac{15+1}{2}^{th} = 8^{th}$ observation = 222 (2)

ii)

$$Q_1 = \left(\frac{15+1}{4}\right)^{th} \text{observation} = 218$$

$$Q_3 = \left(\frac{3*15+3}{4}\right)^{th} \text{observation} = 231$$

Thus, the interquartile range (IQR) = $Q_3 - Q_1$ = 231 – 218 =13. (3)

[5]

Sol. 2:

i) Define B_i as the event that a battery is of type i, for i = 1, 2, 3.
Define A as the event that a battery lasted over 100 hours.

$$P(A) = \sum_{i=1}^{3} P(A|B_i) * P(B_i) = 0.7 * 0.2 + 0.4 * 0.3 + 0.3 * (1 - 0.2 - 0.3) = 0.41$$

(2)

ii)

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(A|B_1) * P(B_1)}{P(A)} = \frac{0.7 * 0.2}{0.41} = 0.34$$

(2)

[4]

Sol. 3 : X is a random variable with cumulative distribution function F(x).

Distribution	Mean	Variance	x	F(x)
Binomial	10	5	*	0.0207
Normal	0	1	*	0.1190
Chi-Square	*	72	51	*

Binomial: X ~ Bin(n, p)

Mean = n p = 10 & Variance = n p (1 - p) = 5
This means: 1 - p = 0.5 i.e. p = 0.5; n = 10/0.5 = 20

$$x = F^{-1}(0.0207) = 5 \quad [from\ tables\ in\ page\ 188]$$

Normal: X ~ N(0, 1)

$$
\begin{aligned}
x &= F^{-1}(0.119)\\
&= F^{-1}(1-0.881)\\
&= F^{-1}\big(1-F(1.18)\big) \qquad [from\ tables\ in\ page\ 160]\\
&= F^{-1}\big(F(-1.18)\big)\\
&= -1.18
\end{aligned}
$$

Chi-square: X ~ $\chi^2(\nu)$

Mean = ν & Variance = 2ν = 72. This means: ν = 36.

$F_{\chi2(36)}[51] = 1 - 5\% = 0.95 \quad [from\ tables\ in\ page\ 169]$

[3]

Sol. 4 :

i) $F_V(t) = P(V \le t) = P\,(max\,\{X,Y\} \le t)$
$= P(X \le t\ and\ Y \le t) = P(X \le t)P(Y \le t)\ \ as\ X\ and\ Y\ are\ independent$
$= F_X(t) * F_Y(t)$ (2)

ii) $F_W(t) = P(W \le t) = 1 - P(min\{X,Y\} > t)$
$= 1 - P(X > t\ and\ Y > t)$
$= 1 - P(X > t)P(Y > t) \quad as\ X\ and\ Y\ are\ independent$
$= 1 - [1 - P(X \le t)]\,[1 - P(Y \le t)]$
$= 1 - (1 - F_X(t))\,(1 - F_Y(t))$
$= F_X(t) + F_Y(t) - F_X(t)F_Y(t)$ (2)

iii) The random variable X has an exponential distribution with mean 0.25 and, independently, Y has an exponential distribution with mean 0.25. This means the parameter λ for both these random variables is 4.

$F_X(t) = F_Y(t) = 1 - e^{-4t}$

$$
\begin{aligned}
F_W(t) &= 1 - e^{-4t} + 1 - e^{-4t} - (1 - e^{-4t})^2\\
&= 2 - 2e^{-4t} - 1 - e^{-8t} + 2e^{-4t}\\
&= 1 - e^{-8t}
\end{aligned}
$$

This is the CDF of an exponential random variable with parameter λ = 8.
Deriving Median (x_m) from 1st principles: $0.5 = 1 - e^{-8x_m} \Rightarrow x_m = -\frac{\log_e 0.5}{8} = 0.0866.$ (3)

[7]

Sol. 5 : i) Let X be a random variable which follows a Geometric distribution with parameter p (0 < p < 1).

The random variable X has the probability mass function:

$$P(X = x) = p\,(1-p)^x \qquad x = 0, 1, 2, \dots$$

The probability generating function is therefore given by:

$$G(s) = E[s^X] = \sum_{x=0}^{\infty} s^x p\,(1-p)^x = \sum_{x=0}^{\infty} p\,\{s(1-p)\}^x = \frac{p}{1-s(1-p)}$$

$$\text{where } |s| < \frac{1}{1-p} \text{ for convergence}$$

To compute mean and variance:

$$G'(s) = \frac{p(1-p)}{[1-s(1-p)]^2}, \qquad \text{thus } \mathrm{E}[X] = G'(1) = \frac{p(1-p)}{[1-(1-p)]^2} = \frac{1-p}{p}$$

$$G''(s) = \frac{2p(1-p)^2}{[1-s(1-p)]^3}, \quad \text{thus } \mathrm{E}[X(X-1)] = G''(1) = \frac{2p(1-p)^2}{[1-(1-p)]^3} = \frac{2(1-p)^2}{p^2}$$

$$\begin{aligned}\text{So, } \mathrm{Var}[X] &= E[X(X-1)] + E[X] - (E[X])^2 \\ &= \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2 \\ &= \frac{1-p}{p^2}.\end{aligned}$$ (4)

ii) The pair of random variables X and Y have the following properties:

- X is a discrete random variable
- Y is a continuous random variable
- The unconditional distribution of Y is exponential with parameter λ (> 0)
- Conditional on Y = y, X follows a Poisson distribution with expected value y.

a. $$P(X = x|y) = \frac{e^{-y}y^x}{x!} \quad \text{for } x = 0, 1, 2 \dots$$

$$\begin{aligned}\therefore P(X = x) &= \int_0^{\infty} \frac{e^{-y}y^x}{x!}\lambda e^{-\lambda y}\,dy \\ &= \frac{\lambda}{x!}\int_0^{\infty} y^x e^{-(\lambda+1)y}\,dy \\ &= \frac{\lambda}{(\lambda+1)^{x+1}}\int_0^{\infty} \frac{t^{(x+1)-1}}{\Gamma(x+1)}e^{-t}\,dt \quad \text{where } t = (\lambda+1)y\end{aligned}$$

$$= \frac{\lambda}{(\lambda+1)^{x+1}} * 1$$

$$\left[\text{where} \int_0^\infty \frac{t^{(x+1)-1}}{\Gamma(x+1)} e^{-t}\, dt = 1 \ldots \text{the total probability of a } Gamma(x+1,1)\; r.v.\right]$$

$$= \frac{\lambda}{(\lambda+1)^{x+1}} \quad i.e. \quad \frac{\lambda}{\lambda+1}\left(1 - \frac{\lambda}{\lambda+1}\right)^x$$

Thus X is a Geometric random variable with $p = \frac{\lambda}{\lambda+1} \in (0,1)$. (3)

b. Using the fact that X is a Geometric random variable with $p = \frac{\lambda}{\lambda+1}$:

$$E[X] = \frac{1-p}{p}$$

$$= \frac{1 - \frac{\lambda}{\lambda+1}}{\frac{\lambda}{\lambda+1}} = \frac{1}{\lambda}$$

$$Var[X] = \frac{1-p}{p^2}$$

$$= \frac{\left(1 - \frac{\lambda}{\lambda+1}\right)}{\left(\frac{\lambda}{\lambda+1}\right)^2} = \frac{\lambda+1}{\lambda^2}$$

Using the fact that X | Y = y is a Poisson random variable with expected value y:

$$E(X|Y) = Y$$
$$Var(X|Y) = Y$$

Using the fact that Y is an Exponential random variable with parameter λ:

$$E[Y] = \frac{1}{\lambda}$$
$$Var[Y] = \frac{1}{\lambda^2}$$

Thus:

$$E[E\{X|Y\}] = E[Y] = \frac{1}{\lambda} = E[X]$$

$$E[Var\{X|Y\}] + Var[E\{X|Y\}]$$
$$= E[Y] + Var[Y]$$
$$= \frac{1}{\lambda} + \frac{1}{\lambda^2} = \frac{\lambda+1}{\lambda^2} = Var[X]$$

(5)

[12]

Sol. 6 :

We are given: $\sum_{\mathrm{i}=1} r_i = 198.0; \ \sum_{i=1} (r_i - \bar{r})^2 = 71.08$

i) $$\text{Sample Mean} = \frac{\sum_{\mathrm{i}=1}^{10} r_i}{10} = \frac{198.0}{10} = 19.8$$

$$\text{Sample Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{10} (r_i - \bar{r})^2}{10-1}} = \sqrt{\frac{71.08}{9}} = 2.81$$ (2)

ii) Assuming that the Chairman's claim on the true values of μ and σ is correct, it means R ~ N(21, 4).

a. Under the above assumptions, the sample mean $\bar{R} \sim N(21,\ 0.4)$.

$$\begin{aligned} &P(\bar{R} < 19.8) \\ &= P\left(Z < \frac{19.8 - 21}{\sqrt{0.4}}\right) \\ &= \Phi(-1.90) \\ &= 1 - \Phi(1.90) \\ &= 1 - 0.97128 \\ &= 0.02872 \end{aligned}$$ (3)

b. Under the above assumptions, using the sample variance S^2 the statistic:

$$\frac{9 * S^2}{4} \sim \chi^2 \text{ with 9 degrees of freedom} \quad \text{i.e.} \quad 2.25\, S^2 \sim \chi_9^2 \text{ distribution}$$

$$\begin{aligned} &P(S > 2.81) \\ &= P(\chi_9^2 > 2.25 * 2.81^2) \\ &= P(\chi_9^2 > 17.77) \\ &= 1 - P(\chi_9^2 < 17.77) \\ &= 1 - 0.962 \\ &= 0.038 \end{aligned}$$

Using linear interpolation:

$$\frac{p - 0.9586}{0.9648 - 0.9586} = \frac{17.77 - 17.5}{18.0 - 17.5} \Rightarrow p = 0.962$$ (3)

[8]

Sol. 7 : **i)** Suppose X is a continuous random variable with probability density function f(X):

$$f(x) = \begin{cases} \frac{1}{2}\, x^2 e^{-x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$M_X(t) = E[e^{Xt}]$

$$= \int_0^\infty e^{xt}\, \frac{1}{2}\, x^2 e^{-x}\, dx$$

$$= \frac{1}{2} \int_0^\infty x^2 e^{-(1-t)\,x}\, dx$$

$$= \frac{1}{(1-t)^3} \int_0^\infty \frac{(1-t)^3}{\Gamma(3)} x^{3-1} e^{-(1-t)\,x}\, dx$$

$$= \frac{1}{(1-t)^3} \quad \text{since the integral equals 1 as it is the total probability of a Gamma r.v.}$$ (3)

ii) The size of a motor insurance claim X, is to be modelled using a gamma random variable with parameters α and θ (both > 0) such that the moment generating function of X is given by:

$$M_X(t) = (1 - \theta t)^{-\alpha} \quad \text{for } t < \frac{1}{\theta}$$

The cumulant generating function is given by: $C_X(t) = \log_e M_X(t) = -\alpha\ \log_e(1 - \theta t)$.

$$C_X'(t) = \alpha\theta(1 - \theta t)^{-1}$$

$$C_X''(t) = \alpha\theta^2(1 - \theta t)^{-2}$$

$$C_X'''(t) = 2\alpha\theta^3(1 - \theta t)^{-3}$$

Therefore: $\kappa_2 = C''(0) = \alpha\theta^2$ and $\kappa_3 = C'''(0) = 2\alpha\theta^3$

So, Coefficient of Skewness is given by:

$$\frac{\kappa_3}{\kappa_2^{3/2}} = \frac{2\alpha\theta^3}{(\alpha\theta^2)^{3/2}} = \frac{2}{\sqrt{\alpha}}$$ (3)

[*To gain credit the students must use cumulant generating function*]

[6]

Sol. 8 : A botanist is conducting a crossbreeding experiment with pea plants which resulted in four types of pea pods being grown. He believes that the types are dependent on only two genes – colour and texture, each of which can appear either in dominant or recessive form:

	Colour	Texture
Dominant	Y	R
Recessive	g	w

The actual observed frequencies of the four types YR, Yw, gR, gw in this experiment were as follows:- YR: 315 Yw: 101 gR:108 gw:32

i) He expects that the frequencies of the four types YR, Yw, gR, gw should be in the ratio 9:3:3:1.

	YR	**Yw**	**gR**	**gw**	**Total**
Observed	315	101	108	32	**556**
Expected	312.75[$]	104.25	104.25	34.75	**556**
$\frac{(O-E)^2}{E}$	0.016	0.101	0.135	0.218	**0.470**

[$] (9/16) * 556 = 312.75 and so on ...

The test statistic $\sum\frac{(O-E)^2}{E}$ (with an observed value of 0.470) follows a chi-square distribution with 3 (= 4 - 1) degrees of freedom. From the tables, the p-value lies between 0.9189 and 0.9402, which is large. (Alternately, $\chi^2_{3,0.05} = 7.815$)

So, we can conclude that the experimenter postulation is correct at least at 5% level. (4)

ii) To test if colour and texture are independent traits (H_0), we first reorganize the **observed** data in a contingency table:

Observed	**R**	**w**	**Total**
Y	315	101	**416**
g	108	32	**140**
Total	**423**	**133**	**556**

To calculate **expected** values, the marginal totals are divided up *pro rata* under H_0:

Expected	**R**	**w**	**Total**
Y	316.49[$]	99.51	**416**
g	106.51	33.49	**140**
Total	**423**	**133**	**556**

$\frac{416*423}{556} = 316.49$ and so on ....

Hence, the chi-square statistic table is as below:

$\frac{(O-E)^2}{E}$	R	w	Total
Y	0.007	0.022	**0.029**
g	0.021	0.066	**0.087**
Total	**0.028**	**0.088**	**0.116**

The test statistic $\sum \frac{(O-E)^2}{E}$ (with an observed value of 0.116) follows a chi-square distribution with 1 [= (2 - 1) * (2 - 1)] degrees of freedom. From the tables, the p-value lies between 0.6547 and 0.7518, which is large. (Alternately, $\chi^2_{1,0.05} = 3.841$)

Thus, the null hypothesis of no association between the row and column factors is accepted at least at 5% level. In other words, colour and texture could be independent traits. (4)

[8]

Sol. 9 :

Independent random samples, of size n each, are taken from normal populations $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ respectively, where the parameters μ_1 and μ_2 are unknown and σ^2 is known.

i) The sampling distributions of $\bar{X}_1$ and $\bar{X}_2$ are given as below:

- $\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n}\right)$
- $\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n}\right)$

$\bar{X}_1 - \bar{X}_2$, is the difference between two independent normal random variables and so is itself normal, with mean (μ_1 - μ_2) and variance $\frac{2\sigma^2}{n}$. (1)

ii) Taking Step 1, we determine the two 95% confidence intervals as below:

- μ_1: $\bar{X}_1 \mp 1.96 * \frac{\sigma}{\sqrt{n}}$
- μ_2: $\bar{X}_2 \mp 1.96 * \frac{\sigma}{\sqrt{n}}$

In order to reject the null hypothesis that the two population means are equal (Step 2), we must ensure that the two confidence intervals **do not** overlap. This is equivalent to saying one of the following two scenarios hold:

Scenario 1:

$\bar{X}_2 - 1.96 * \frac{\sigma}{\sqrt{n}}$ $\quad$ $\bar{X}_2 + 1.96 * \frac{\sigma}{\sqrt{n}}$

$$---(----------)---(----------)---\rightarrow$$

$\bar{X}_1 - 1.96 * \frac{\sigma}{\sqrt{n}}$ $\quad$ $\bar{X}_1 + 1.96 * \frac{\sigma}{\sqrt{n}}$

Scenario 2:

$$\begin{array}{c}\qquad\qquad\qquad\qquad\qquad\qquad \bar{X}_1 - 1.96 * \frac{\sigma}{\sqrt{n}} \qquad\qquad\qquad \bar{X}_1 + 1.96 * \frac{\sigma}{\sqrt{n}} \\ ----(----------)----(----------)--\rightarrow \\ \bar{X}_2 - 1.96 * \frac{\sigma}{\sqrt{n}} \qquad\qquad\qquad \bar{X}_2 + 1.96 * \frac{\sigma}{\sqrt{n}} \qquad\qquad\qquad\qquad\qquad\qquad\end{array}$$

The probability of Type 1 error

$= \text{P}(\text{Reject } H_0 | H_o \text{ is true})$

$$= \text{P}\left(\bar{X}_1 + 1.96 * \frac{\sigma}{\sqrt{n}} < \bar{X}_2 - 1.96 * \frac{\sigma}{\sqrt{n}}\right) + \text{P}\left(\bar{X}_2 + 1.96 * \frac{\sigma}{\sqrt{n}} < \bar{X}_1 - 1.96 * \frac{\sigma}{\sqrt{n}}\right)$$

$$= \text{P}\left(\bar{X}_1 - \bar{X}_2 < -2 * 1.96 * \frac{\sigma}{\sqrt{n}}\right) + \text{P}\left(\bar{X}_1 - \bar{X}_2 > 2 * 1.96 * \frac{\sigma}{\sqrt{n}}\right)$$

$$= \text{P}\left[\frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{2} * \frac{\sigma}{\sqrt{n}}} < -\sqrt{2} * 1.96\right] + \text{P}\left[\frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{2} * \frac{\sigma}{\sqrt{n}}} > \sqrt{2} * 1.96\right]$$

$$= \Phi(-\sqrt{2} * 1.96) + [1 - \Phi(\sqrt{2} * 1.96)]$$

$$\because \bar{X}_1 - \bar{X}_2 \sim N\left(0, \frac{2\sigma^2}{n}\right) \textit{ using independence and that under } H_0 \text{: } \mu_1 = \mu_2$$

$$= 2 * [1 - \Phi(\sqrt{2} * 1.96)]$$

$$= 2 * [1 - 0.9972]$$

$$= 0.0056$$

Clearly, the assertion made by the statistician is **incorrect** as the probability of type 1 error is much less than 0.05 using his proposed approach.

This means that using this approach one can reject H_0 at 5% significance based on the fact that the two confidence intervals do not overlap. However, if the confidence intervals do overlap, it is not clear if one can possibly claim that H_0 can be accepted at 5% significance level. (6)

[7]

Sol. 10 :

The local government has commissioned a study of length of hospital stay, in days, as a function of age, type of health insurance and whether or not the patient died while in the hospital. Length of hospital stay is recorded as a minimum of at least one day and the number of days is rounded up to the next integer.

An analyst working on the study believes that the length of hospital stay (X) can be modelled as a discrete random variable with a probability mass function:

$$P(X = k) = c * \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for k} = 1, 2, 3 \ ...$$

Here, c is a constant and λ (> 0) is an unknown parameter.

The analyst collects data for 20 random patients for the study.

i) For a valid probability mass function, we need:

$$\sum_{k=1}^{\infty} P(X = k) = 1.$$

In other words:

$$\begin{aligned}
1 &= \sum_{k=1}^{\infty} P(X = k) \\
&= \sum_{k=1}^{\infty} c * \frac{e^{-\lambda}\lambda^k}{k!} \\
&= c * \left(\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} - e^{-\lambda}\right) \\
&= ce^{-\lambda} * \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - 1\right) \\
&= ce^{-\lambda}(e^{\lambda} - 1) \\
&= c(1 - e^{-\lambda})
\end{aligned}$$

Thus:

$$c = \frac{1}{1 - e^{-\lambda}} \tag{2}$$

ii) Let $X_1, X_2 \ldots X_{20}$ be the patient data collected by the analyst.

The likelihood function is given by:

$$\begin{aligned}
L(\lambda) &= \prod_{i=1}^{20} P(X_i = x_i) \\
&= \prod_{i=1}^{20} \frac{e^{-\lambda}\lambda^{x_i}}{(1 - e^{-\lambda})\, x_i!} = \frac{e^{-20\lambda}\lambda^{\sum x_i}}{(1 - e^{-\lambda})^{20} \prod x_i!}
\end{aligned}$$

The log-likelihood function is given by:

$$\begin{aligned}
l(\lambda) &= \log_e L(\lambda) \\
&= -20\lambda + \left(\sum_{i=1}^{20} x_i\right) * \log_e \lambda - 20 * \log_e(1 - e^{-\lambda}) - \log_e\left(\prod_{i=1}^{20} x_i!\right)
\end{aligned} \tag{3}$$

iii) The maximum likelihood estimator of λ, $\hat{\lambda}_{MLE}$, satisfies the equation:

$$\frac{dl(\lambda)}{d\lambda} = 0$$

i.e.

$$0 = \frac{dl(\lambda)}{d\lambda} = -20 + \frac{\sum x_i}{\lambda} - \frac{20\,e^{-\lambda}}{1 - e^{-\lambda}}$$

i.e.

$$\frac{\sum x_i}{\lambda} = \frac{20}{1 - e^{-\lambda}}$$

(3)

The analyst wrote down the following information obtained using the data on the likelihood equations for various value of λ:

λ_0	1.700	1.750	1.800	1.850
$l(\lambda_0)$	-27.421	-27.391	-27.384	-27.398
$l'(\lambda_0)$	0.824	0.365	-0.072	-0.488
$l''(\lambda_0)$	-9.409	-8.950	-8.527	-8.136

Here:

- $l'(\lambda_0)$ = value of the first derivative of $l(\lambda)$ at $\lambda = \lambda_0$
- $l''(\lambda_0)$ = value of the second derivative of $l(\lambda)$ at $\lambda = \lambda_0$

iv) $\hat{\lambda}_{MLE}$ will satisfy:

$$\left.\frac{dl(\lambda)}{d\lambda}\right|_{\lambda=\hat{\lambda}_{MLE}} = l'(\hat{\lambda}_{MLE}) = 0$$

Looking at the values provided it is clear that $\hat{\lambda}_{MLE}$ lies between 1.75 and 1.80.

Using linear interpolation,

$$\frac{\hat{\lambda}_{MLE} - 1.75}{1.80 - 1.75} = \frac{0 - 0.365}{(-0.072) - 0.365}$$

$$\Rightarrow \hat{\lambda}_{MLE} = 1.75 + 0.05 * \frac{0.365}{0.437} = 1.792$$

(4)

v) It can be shown that the asymptotic variance for $\hat{\lambda}_{MLE}$ is given by:

$$Var(\hat{\lambda}_{MLE}) \approx -\frac{1}{l''(\hat{\lambda}_{MLE})}$$

To evaluate $l''(\hat{\lambda}_{MLE})$ we use linear interpolation again:

$$\frac{l''(\hat{\lambda}_{MLE}) - (-8.950)}{(-8.527) - (-8.950)} = \frac{1.792 - 1.750}{1.800 - 1.750}$$

$$\Rightarrow l''(\hat{\lambda}_{MLE}) = (-8.950) + 0.423 * \frac{0.042}{0.050} = -8.595$$

Alternately, this can be computed without knowing the value of $\hat{\lambda}_{MLE}$ using the fact that $l'(\hat{\lambda}_{MLE}) = 0$:

$$\frac{l''(\hat{\lambda}_{MLE}) - (-8.950)}{(-8.527) - (-8.950)} = \frac{0 - 0.365}{(-0.072) - 0.365}$$

$$\Rightarrow l''(\hat{\lambda}_{MLE}) = (-8.950) + 0.423 * \frac{0.365}{0.437} \approx -8.597$$

Thus:

$$Var(\hat{\lambda}_{MLE}) \approx -\frac{1}{l''(\hat{\lambda}_{MLE})}$$

$$\approx -\frac{1}{(-8.595)}$$

$$\approx 0.11635$$

In large sample, the asymptotic distribution $\hat{\lambda}_{MLE} \sim N[\lambda, Var(\hat{\lambda}_{MLE})]$

Thus, an approximate 95% confidence interval for λ will be given by:

$$\hat{\lambda}_{MLE} \mp 1.96 * \sqrt{Var(\hat{\lambda}_{MLE})}$$

$$i.e.\ 1.792 \mp 1.96 * 0.3411$$

$$\Longrightarrow (1.123, 2.460)$$

(4)

[16]

Sol. 11 : The following shows how a panel of nutrition experts and a panel of housewives ranked 15 breakfast foods on their palatability:

Breakfast Food	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Nutrition Experts (X)	2	13	14	9	1	4	11	3	7	15	6	12	10	8	5
Housewives (Y)	3	15	13	14	2	6	8	5	4	10	11	9	12	7	1

We are given:

$$\sum x = \sum y = 120; \quad \sum x^2 = \sum y^2 = 1240; \quad \sum\sum xy = 1171$$

i)

$$S_{xx} = \sum x^2 - \left(\sum x\right)^2/15 = 1240 - 120^2/15 = 280$$

$$S_{yy} = \sum y^2 - \left(\sum y\right)^2/15 = 1240 - 120^2/15 = 280$$

$$S_{xy} = \sum\sum xy - \left(\sum x\right)\left(\sum y\right)/15 = 1171 - 120 * 120/15 = 211$$

Thus, the sample correlation coefficient **r** is given by:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\,S_{yy}}} = \frac{211}{\sqrt{280 * 280}} = 0.7536$$ (3)

ii) To test for any linear relationship between the two rankings, we set up the following hypothesis test (with ρ being the population correlation coefficient parameter):

H_0: $\rho = 0$ against H_1: $\rho \neq 0$

Assuming that X and Y are normal random variables whose joint distribution is bivariate normal, the test statistic under H_0: $\rho = 0$ is given by:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$ which follows a $\boldsymbol{t}$ distribution with $(\mathrm{n}-2)$ degrees of freedom.

Here: n = 15 (sample size).

Observed value of the test statistic:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.7536 * \sqrt{13}}{\sqrt{1-0.7536^2}} = 4.133.$$

The critical value at 5% significance for a t_{13} distribution is |2.160|. Clearly, we are unable to accept H_0 at 5% level of significance. (4)

Hence, we can conclude that there is some evidence of linear relationship between the two rankings.

iii) In order to compute an approximate 95% confidence interval for the corresponding population parameter ρ, we need to apply the Fisher's transformation of **r**.

We know:

$z_r \sim N\left(z_\rho, \frac{1}{n-3}\right)$ approximately

where $z_r = \tanh^{-1} r = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right)$ and $z_\rho = \tanh^{-1}\rho = \frac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right)$

An *approximate* 95% confidence interval for $\mathbf{z_\rho}$ is given by:

$$z_r \mp 1.96 * \frac{1}{\sqrt{n-3}}$$

$$i.e.\ \tanh^{-1}(0.7536) \mp 1.96 * \frac{1}{\sqrt{12}} \quad or, \quad \frac{1}{2}\log_e\left(\frac{1+0.7536}{1-0.7536}\right) \mp 1.96 * \frac{1}{\sqrt{12}}$$

$i.e.\ \ 0.981 \mp 0.566$

$\Rightarrow (0.415, 1.547)$

Reversing the Fisher's transformation:

$$z_r = \tanh^{-1} r = \frac{1}{2}\log_e\left(\frac{1+r}{1-r}\right) \quad \Leftrightarrow \quad r = \tanh(z_r) = \frac{e^{2z_r}-1}{e^{2z_r}+1}$$

Thus, an *approximate* 95% confidence interval for **ρ** is given by:

(r_{lower}, r_{upper})

$$i.e.\ (\tanh(z_{r}\_lower), \tanh(z_{r}\_upper)) \ \ or, \ \ \left(\frac{e^{2z_{r}\_lower}-1}{e^{2z_{r}\_lower}+1}, \frac{e^{2z_{r}\_upper}-1}{e^{2z_{r}\_upper}+1}\right)$$

$$i.e.\ (\tanh(0.415), \tanh(1.547)) \ \ or, \ \ \left(\frac{e^{2*0.415}-1}{e^{2*0.415}+1}, \frac{e^{2*1.547}-1}{e^{2*1.547}+1}\right)$$

$\Rightarrow (0.393, 0.913)$ (7)

[14]

Sol. 12 :

An Indian market research firm has conducted a survey of five leading mobile brands on various parameters like customer satisfaction, quality of the product and post sales service. Based on the outcome of the survey, it gave scores to each of the brand considering the various parameters. The survey was conducted over six months and following scores were awarded in each month:

Month	Brand				
	A	B	C	D	E
1	168	184	216	192	232
2	176	195	184	184	213
3	141	147	155	181	143
4	166	168	166	194	185
5	136	179	198	149	183
6	133	190	185	189	182
Total	**920**	**1063**	**1104**	**1089**	**1138**

$\sum\sum y_{ij}$ = 5,314; $\sum\sum y_{ij}^2$ = 957,718

i) If we assume the observations are from normal populations with the same variance, we can apply an ANOVA test of the hypotheses:

H_0 : Each brand has the same mean score.
H_1 : There are differences between the mean scores of the different brands.

For this data:

- Number of treatments (brands): k = 5 and
- Sample size: $n_1 = 6, n_2 = 6, n_3 = 6, n_4 = 6, n_5 = 6$ and $n = 30$

$\sum\sum y_{ij} = 5{,}314$; $\sum\sum y_{ij}^2 = 957{,}718$

$$SS_T = 957{,}718 - \frac{5{,}314^2}{30} = 16{,}431.47$$
$$SS_B = \frac{1}{6}(920^2 + 1{,}063^2 + 1{,}104^2 + 1{,}089^2 + 1{,}138^2) - \frac{5{,}314^2}{30} = 4{,}738.47$$
$$SS_R = SS_T - SS_B = 11{,}693$$

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*
Treatments	4,738.47	4	1,184.62	2.53
Residuals	11,693.00	25	467.72	
Total	**16,431.47**	**29**		

Under H_0, $F = \frac{1{,}184.62}{467.72} = 2.53$, using the $F_{4,25}$ distribution.

The 5% critical point is 2.76, so we have insufficient evidence to reject H_0 at the 5% level. Therefore we conclude that there is no difference between the brands' mean scores at the 5% level. (6)

ii) Here: $\frac{SS_R}{\sigma^2} \sim \chi_{30-5=25}$

Using values from the chi-square tables:
$$P\left(13.120 < \frac{SS_R}{\sigma^2} < 40.646\right) = 0.95$$

$\therefore$ 95% confidence interval for σ^2 is given as:

$$\left(\frac{SS_R}{40.646}, \frac{SS_R}{13.120}\right)$$
$$i.e. \left(\frac{11693}{40{,}646}, \frac{11693}{13.120}\right)$$
$$i.e. \quad (287.68, 891.25)$$

Hence a 95% confidence interval for σ is given by:
$$\left(\sqrt{287.68}, \sqrt{891.25}\right) \Leftrightarrow (16.96, 29.85)$$ (4)

[10]

****