Institute of Actuaries of India

Subject CT3 – Probability & Mathematical Statistics

September 2018 Examination

INDICATIVE SOLUTION

Introduction

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 points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1:

Combined mean $\bar{x}$ = $(n_1\,\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2)$

= (130.2×28+140.7×15)/(28+15)

= (5756.1)/(43)

= 133.9 (lakhs of Rupees)

Using the fact that $\sum {x_i}^2 = (n_i - 1)s_i^2 + n_i\,{\overline{x_i}}^2; i = 1,2$ we have, for the combined set

$\sum x^2$ = (27× 55.1^2 +28× 130.2^2) + (14× 66.2^2+15× 140.7^2) = 914930.9

Therefore the variance for the combined set

$$S^2 = \frac{914930.9 - \frac{5756.1^2}{43}}{42}$$

= 144403/42

= 3438.7.

Thus, S = 58.6 (lakhs of Rupees)

[4 Marks]

Solution 2:

i) Let Success: = Getting one passenger to go to *B* from *A.*

p = Probability [Getting one passenger to go to *B* from *A*.] = 0.3;

q = $1 - p$ =0.7

k = the number of successes (number of passengers going to Town *B)* = 4

X+*k* = the number of trials=15

X = the number of failures = 11

X follows *NB* (*k*=4, *p* = 0.3)

$[X \sim NB(k,p) \rightarrow P(X = x) = \binom{x+k-1}{x} p^k q^x; = x = 0,1, \ldots$]

Hence, $P(X = 11) = \binom{11+4-1}{11} 0.3^4 0.7^{11}$

= $\frac{14!}{11!3!}\, 0.3^4 0.7^{11}$=0.0583. [3]

ii) The average number of persons to be asked in order to get 4 passengers

$= E$ (the number of trials)

$= E(X+k) = E(X) + k = \frac{kq}{p} + k = \frac{k}{p}$

$= \frac{4\times 0.7}{0.3} + 4 = \frac{4}{0.3}$

$= 13.33 \sim 14$ persons. [3]

[6 Marks]

Solution 3:

i) Mode: For fixed $\vartheta > 0$, the density function $f(x)$ is an increasing function of x.

Thus, $f(x)$ has maximum at the right end point of the interval $[0, \vartheta]$.

Hence the mode of this distribution is ϑ.

Median:

½ $= \int_0^q f(x)\,dx$

$= \int_0^q \frac{3x^2}{\theta^3}\,dx = [x^3/\theta^3]$ from 0 to q

Thus, ½ $= q^3/\theta^3$ implies $q = \frac{\theta}{2^{\frac{1}{3}}}$ [5]

ii) Let A be the ratio of the mode of this distribution to the median

A = mode/median = $\theta \times \frac{2^{\frac{1}{3}}}{\theta} = 2^{\frac{1}{3}} = 1.2599$

$P(X < A) = \int_0^A f(x)\,dx$

$= \int_0^A 3x^2/\theta^3\,dx$

$= [x^3/\theta^3]$ from 0 to A

$= A^3/\theta^3$

$$= \begin{cases} \frac{2}{\theta^3} & if \quad \theta > 2^{1/3} = 1{\cdot}2599 \\ 1 & otherwise \end{cases}$$ [3]

[8 Marks]

Solution 4:

The moment generating function of $X_1 = M_{X_1}(t_1) = E(e^{t_1 X_1})$

$$= M_X(t_1, 0) \ = \frac{1}{3}\left(1 + e^{(t_1+2\times 0)} + e^{(2t_1+0)}\right)$$

$$= \frac{1}{3}\left(1 + e^{t_1} + e^{2t_1}\right)$$

The expected value of $X_1$is obtained bytaking first derivative of its MGF and evaluating at t_1=0.

Thus, $$E(X_1) = M_{X_1}(0)=\frac{1}{3}\left(1 + e^0 + e^0\right) = 1$$

Similarly using the mgf of X_2 , $E(X_2)$ is shown to be 1

$E(X_1 X_2)$ is computed by taking the second cross-partial derivative of joint moment generating function evaluated at $(t_1, t_2) = (0,0)$:

$$\frac{\partial^2 M_{X_1,X_2}(t_1,t_2)}{\partial t_1 \partial t_2} = \frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}\left(\frac{1}{3}[1 + \exp(t_1 + 2t_2) + \exp(2t_1 + t_2)]\right)\right)$$

$$= \frac{\partial}{\partial t_1}\left(\frac{1}{3}[2\exp(t_1 + 2t_2) + \exp(2t_1 + t_2)]\right)$$

$$= \frac{1}{3}[2\exp(t_1 + 2t_2) + 2\exp(2t_1 + t_2)]$$

Thus, $E(X_1 X_2) = \frac{4}{3}$

$$Cov(X_1 X_2) = E(X_1 X_2) - E(X_1)E(X_2) = \frac{4}{3} - 1 \times 1 = \frac{1}{3} = 0.33$$

[7 Marks]

Solution 5: By transformation, the two normal random variables X_1 and X_2, can be written as

$X_1 = 4Z_1$, $X_2 = 4Z_2$ where Z_1 and Z_2 are standard normal random variables.

Thus, we can write $P(X_1^2 + X_2^2 > 8) = P(16\, Z_1^2 + 16\, Z_2^2 > 8)$

$$= P(Z_1^2 + Z_2^2 > \tfrac{1}{2})$$

The sum $Z_1^2 + Z_2^2$ has chi square distribution with 2 *df*.

Therefore, $P(X_1^2 + X_2^2 > 8) = P(Z_1^2 + Z_2^2 > \frac{1}{2})$

$$= 1 - P(Z_1^2 + Z_2^2 < \tfrac{1}{2})$$

$$= 1 - F_Y(1/2)$$

where $F_Y(1/2)$ is the distribution function of a chi square random variable *Y* with 2 *df*.

Evaluated at $y = 1/2$, the value of $F_Y(1/2) = 0.2212$. **[4 Marks]**

Solution 6: $X \sim b(2, \theta)$; $H_0 : \theta = \frac{1}{2}$ and $H_1 : \theta = \frac{3}{4}$

A single observation *x* is taken; Critical region: *x* = 1 *or* 2

i) Probability of Type I error = *P* (Rejecting H_0 | H_0 is true)

$$= P(x = 1 \; or \; x = 2 | \; \theta = \tfrac{1}{2})$$

$$= \binom{2}{1}(\tfrac{1}{2})^1(\tfrac{1}{2})^1 + \binom{2}{2}(\tfrac{1}{2})^2(\tfrac{1}{2})^0 = \tfrac{3}{4}$$ [2]

ii) Probability of Type II error = *P*(Not Rejecting H_0 | H_1 is true)

$$= P(x = 0 \mid \theta = \tfrac{3}{4}) = \binom{2}{0}(\tfrac{3}{4})^0(\tfrac{1}{4})^2 = \tfrac{1}{16}$$

Power of the test = $1 - P(\text{Type II error}) = 1 - \frac{1}{16} = \frac{15}{16}$. [3]

[5 Marks]

Solution 7:

i) $F_C(c) = P(C \le c) = \sum P(C \le c / \; N = n).P(N = n)$

$$= \sum P[W_1 + W_2 + \cdots + W_N \le c/N = n].P(N = n)$$

$$= \sum P[W_1 + W_2 + \cdots + W_N \le c] \; P(N = n)$$

$$= \sum P_{W^*}(c) P_N(n) \; [W^* \text{ is the } n \text{ fold convolution of } W].$$

[3]

ii) Let $M_C(t)$ be the mgf of C.

$$M_C(t) = E[e^{tc}] = E_N[E(e^{ct} \;\; / N]$$

$$= E_N[E\{e^{(W_1+W_2+\cdots+W_N)t}/N\}]$$

$$= E_N[E(e^{w_1 t})E(e^{w_2 t}) \ldots E(e^{w_N t})]$$

$$= E_N[\; E(e^{w_1 t})^N]$$

$$= E_N[\left(M_W(t)\right)^N]$$

$$= E_N[e^{N \log M_W(t)}]$$

$$= M_N[\log M_N(t)].$$

Hence, log $M_C(t)$ = $\psi_C(t) = \psi_N(\psi_W(t))$ which is the cumulant generating function of *C.*

[3]

[6 Marks]

Solution 8:

i) The likelihood function is $L(\lambda) = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_{i!}}$

$\Rightarrow$ log $L(\lambda)$ = $-n\lambda$ + $(\sum x_i)$ log λ + constant.

$\frac{\partial \, logL(\lambda)}{\partial\lambda}$ =0 imples$-n + \frac{\sum x_i}{\lambda} = 0$ giving $\widehat{\lambda}$=$\bar{X}$.

$\frac{\partial^2 logL(\lambda)}{\partial\lambda^2} = -\frac{\sum x_i}{\lambda^2} < o$

Hence, MLE $\widehat{\lambda}$=$\bar{X}$ [4]

ii) $\frac{\partial^2 logL(\lambda)}{\partial\lambda^2} = -\frac{\sum x_i}{\lambda^2}$ and $-E[\frac{\partial^2 logL(\lambda)}{\partial\lambda^2}]$=$-E\left[-\frac{\sum x_i}{\lambda^2}\right] = \frac{n}{\lambda}$

CRLB:$\frac{1}{-E[\frac{\partial^2 logL(\lambda)}{\partial\lambda^2}]} = \frac{\lambda}{n}$ [4]

iii)

a) $E[\,\hat{\lambda}\,] = E[\bar{X}] = E\left[\frac{\sum X_i}{n}\right] = \frac{n\lambda}{n} = \lambda$

$V[\,\hat{\lambda}\,] = V[\bar{X}] = V\left[\frac{\sum X_i}{n}\right] = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$ which is CRLB. [2]

b) The theory of asymptotic distribution of MLEs (and in this case CLT) gives $\hat{\lambda} \sim N(\lambda, \frac{\lambda}{n})$ approximately.

[1]

iv)

a) Large sample approximate 95% confidence interval for λ is given by

$\hat{\lambda} \pm (1.96 \times se\ (\widehat{\lambda})$

With n = 100, we get the 95% CI as

$$\bar{x} \pm (1.96 \times \sqrt{\frac{\bar{x}}{100}})$$

That is, $\bar{x} \pm 0.196\sqrt{\bar{x}}$ [2]

b) $\bar{x}$ = 215/100 = 2.15,

Hence, a confidence interval is 2.15 ± 0.196 $(2.15)^{1/2}$
i.e. 2.15 ± 0.287
i.e. (1.86, 2.44). [3]

[16 Marks]

Solution 9:

i) The marginal *pdf* of Y:

i. $f_1(y) = \int_0^y 2\,dx = \begin{cases} 2y & 0 < y < 1 \\ 0, & otherwise. \end{cases}$ [1]

ii) The conditional *pdf* of X given Y: $f(x/y)$ is

i. $f(x/y) = \frac{f(x,y)}{f_1(y)} = \begin{cases} \frac{2}{2y}\ ;\ 0 < x < y;\ 0 < y < 1 \\ \quad 0\ \ otherwise. \end{cases}$ [2]

iii) The conditional mean: $E(X/Y = 2) = \int_0^y x\,\frac{1}{y}dx = \frac{y}{2},\ 0 < y < 1$

1. $= \frac{1}{4} =$ 0.25 when $y = \frac{1}{2}$ [2]

iv) The conditional variance: $V(X/\ Y = y) = E(X^2/Y = y) - (E(X/Y = 2))^2$

a. $E(X^2/Y = y) = \int_0^y x^2 \frac{1}{y}\,dx = \frac{y^2}{3}\ ; 0 < y < 1$

1. $= \frac{1}{12} = 0.083$ when $y = \frac{1}{2}$.

Hence, $V(X/\ Y = \frac{1}{2}) = \frac{1}{12} - (\frac{1}{4})^2 = \frac{1}{48} = 0.0208$. [3]

[8 Marks]

Solution 10:

i) We use *t*- test for testing $H_0: \mu_A = \mu_B\ \ vs\ \ H_1: \mu_A \neq \mu_B$

Data: $\bar{y}_A = \frac{56.1}{12} = 4.675\ ;\ \bar{y}_B = \frac{59.1}{12} = 4.925$

$$s_A^2 = \frac{1}{11}\left(266.33 - \frac{56.1^2}{12}\right) = 0.369\ ; s_B^2 = \frac{1}{11}\left(297.03 - \frac{59.1^2}{12}\right) = 0.54205$$

Assuming that the two samples are coming from normal populations with same Variance, the pooled variance is computed as:

$$s_p^2 = \frac{11s_A^2 + 11s_B^2}{22} = 0.4555.$$

Hence, the *t* statistic is : $\frac{\bar{y}_A - \bar{y}_B}{\sqrt{s_p^2\left(\frac{1}{12}+\frac{1}{12}\right)}} = -0.907$

The critical value of $t_{22}(0.025) = -2.074$

So we do not have enough evidence to reject H_0 and conclude that the mean delay times are the same for claims associated with the two causes for illness. [6]

ii) The distribution of times can be skewed to the right and we need log transformation for the data to be normally distributed to have a valid test.

[2]

[8 Marks]

Solution 11:

i) $H_0 : \mu_A = \mu_B = \mu_A$; H_1 : at least one pair is not equal.

Data: $\sum y_A = 27$; $\sum y_B = 14$; $\sum y_C = 52$; $\sum y = 93$ *and* $\sum y^2 = 865$

$$SS_T = 865 - \frac{93^2}{15} = 288.4$$

$$SS_B = (27^2 + 14^2 + 52^2)/5 - \frac{93^2}{15} = 149.2$$

ANOVA

Source of variation	SS	df	MSS	*F* ratio
Between	149.2	2	74.6	$F = \frac{74.6}{11.6}$ = 6.431
Residuals	139.2	12	11.6	
Total	288.4	14		

Thus, the calculated value of F for $(2,12)$ df is 6.431.

The critical value of F for $(2,12)$ df at 5% $level$ is 3.885 .
The critical value of F for $(2,12)$ df at 1% $level$ is 6.927.

We reject the null hypothesis at 5% level of significance. However, we do not reject the null hypothesis at 1% level of significance. [6]

ii) The $(1-\alpha)\%$ confidence for $\mu_A - \mu_C$ is:

$$(\overline{y}_A - \bar{y}_C) \pm t_{n-k,\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

The 95% confidence interval for $\mu_A - \mu_C$:

The value $t_{12}(0.025) = 2.179$.

The estimate of σ^2 from ANOVA table is: 11.6.

The 95% confidence interval for $\mu_A - \mu_C$ is :

$$(5.4 - 10.4) \pm 2.179\{11.6\left(\frac{1}{5}+\frac{1}{5}\right)\}^{0.5}$$

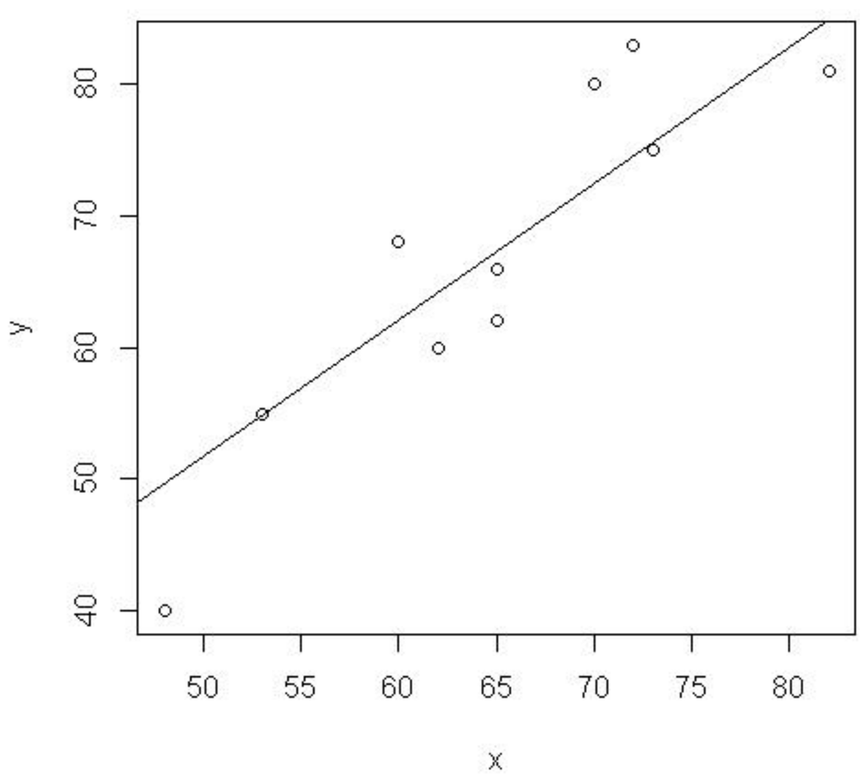
$$= -5 \pm 4.694 = (-9.694, -0.306)$$

[4]

[10 Mark]

<u>Solution 12</u>:

i) Scatter diagram



The scatter diagram indicates that linear regression model can be fitted to the given data. [2]

ii) Linear regression model :

$\sum x_i = 650$; $\sum y_i = 670$; $\sum x_i{}^2 = 43144$; $\sum y_i{}^2 = 46524$; $\sum x_i\, y_i = 44648$

$n = 10$; $\bar{x} = 65$, $\bar{y} = 67$,

$$S_{xy} = \sum x_i\, y_i - \frac{\sum x_i \sum y_i}{n} = 44648 - \frac{650\times 670}{10} = 1098$$

$$S_{xx} = \sum x_i{}^2 - \frac{(\sum x_i)^2}{n} = 894.$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = \frac{1098}{894} = 1.2282$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 67 - 1.2282 \times 65 = -12.833$$

Hence, the fitted regression line is: $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x = -12.833 + 1.2282x$.

[4]

iii) Estimate of σ^2 and confidence interval for σ^2:

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 46524 - \frac{670^2}{10} = 1634.$$

$$SS_{Res} = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = 1634 - \frac{1098^2}{894} = 285.4497$$

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n-2} = \frac{285.4497}{8} = 35.6812125 = MS_{Res}$$

The 95% confidence interval for σ^2 is given by:

$$\frac{(n-2)MS_{Res}}{\chi^2_{0.025,n-2}} \le \sigma^2 \le \frac{(n-2)MS_{Res}}{\chi^2_{0.0975,n-2}}$$

$$\frac{8\times 35.6812125}{17.53} \le \sigma^2 \le \frac{8\times 35.6812125}{2.18}$$

i.e $$\frac{285.4497}{17.53} \le \sigma^2 \le \frac{285.4497}{2.18}$$

$$16.2835 \le \sigma^2 \le 130.9402.$$ [4]

iv) Testing for $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

The test statistic is : $t = \frac{\widehat{\beta_1}}{\sqrt{MS_{Res}/S_{xx}}} = 6.148$

The critical value of *t* at 5% level of significance for 8 *df* is2.306.

Hence, we reject the hypothesis that $\beta_1 = 0$. [4]

v) Sample correlation coefficient $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{1098}{\sqrt{894 \times 1634}} = 0.9084$

Testing for $H_0: \rho = 0 \ vs \ H_1: \rho \neq 0$, we have the *t* statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2} \ \ df$$

$$= \frac{0.9084\sqrt{8}}{\sqrt{1-0.9084^2}}$$ = 6.1477 with 8 *df*

The critical value of *t* for 8 *df* is 2.306

Hence, we reject the null hypothesis that $\rho = 0$. [4]

[18 Marks]

****