Actuarial Society of India

Examinations

November 2006 Examinations

CT3 – Probability and Mathematical Statistics

Indicative Solutions

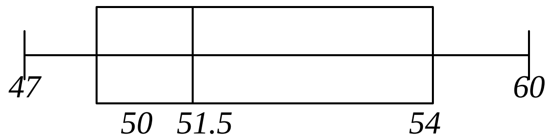
1. i) Minimum : *47* Maximum : *60*

Q_1 : 11^{th} smallest *50*

Q_2 : Average of 21^{st} and 22^{nd} smallest *51.5*

Q_3 : 32^{nd} smallest *54*

BOX PLOT



ii) Mean : $\Sigma fx/n = \dfrac{2174}{42} = 51.76$

iii) Sample variance $s^2 = \dfrac{1}{n-1}\left[\sum fx^2 - n\bar{x}^2\right]$

$$= \frac{1}{41}\left[112890 - 112522\right] = 8.97$$

iv) IQR = $Q_3 - Q_1 = 54 - 50 = 4$

[7]

2. $P(A/B \cap C)P(C) = \dfrac{P(A \cap B \cap C)}{P(B)P(C)}P(C) = \dfrac{P(A \cap B \cap C)}{P(B)}$

$P(A \cap B \cap C^c)P(C) = \dfrac{P(A \cap B \cap C^c)}{P(B)P(C^c)}P(C^c) = \dfrac{P(A \cap B \cap C^c)}{P(B)}$

$$\frac{P(A \cap B \cap C)}{P(B)} + \frac{P(A \cap B \cap C^c)}{P(B)} = \frac{P(A \cap B \cap C) + P(A \cap B \cap C^c)}{P(B)}$$

$$= \frac{P(A \cap B)}{P(B)} = P(A/B)$$

[3]

3. D : Defective bulbs ND : Non Defective

Sample space : { *N, DN, DDN, DDDN*}

with probabilities $\dfrac{5}{8}, \dfrac{3}{8}\times\dfrac{5}{7}, \dfrac{3}{8}\times\dfrac{2}{7}\times\dfrac{5}{6}$ and $\dfrac{3}{8}\times\dfrac{2}{7}\times\dfrac{1}{6}\times\dfrac{5}{5}$

Let X denote number of bulbs chosen with values 1,2,3,4

$EX = 1(5/8) + 2(15/56) + 3(5/56) + 4(1/56)$

$= 12/8 = 1.5$

[4]

4. $f(x,y)=\begin{cases}4y(x-y)e^{-(x+y)} & ; \quad 0<x<\infty, \quad 0\le y\le x \\ 0 & otherwise\end{cases}$

$$f_Y(y)=\int_y^{\infty} 4y(x-y)e^{-(x+y)}dx$$

$$f_{X/Y}{}^{(x/y)}=\frac{f(x,y)}{f_Y(y)}=\frac{4y(x-y)e^{-(x+y)}}{\int_y^{\infty}4y(x-y)e^{-(x+y)}dx}$$

$$=\frac{(x-y)e^{-x}}{\int_y^{\infty}(x-y)e^{-x}dx}$$

The integral $\int_y^{\infty}(x-y)e^{-x}dx$ reduces to $\int_0^{\infty}we^{-(y+w)}dw$ by letting $w = x\text{-}y$. Note that $\int_0^{\infty}we^{-w}dw$ is the expected value of exponential random variable W with mean 1.

Hence, $f_{x/y}(x/y)=(x-y)e^{-(x-y)}; x>y$

$$E\left(X/Y=y\right)=\int_y^{\infty}x(x-y)e^{-(x-y)}dx$$

$$=\int_0^{\infty}w(w+y)e^{-w}dw$$

$$= E(W^2)+ yE(W)$$
$$= 2 + y$$

[7]

5. $X \sim f(x) = e^{-x}\,; x > 0; \qquad Y = 1 - e^{-X}$

$$M_Y(t) = E(e^{tY}) = \int_0^{\infty}e^{t(1-e^{-x})}e^{-x}dx$$

$$= e^t\int_0^{\infty}e^{(-x-te^{-x})}dx \qquad = e^t\int_0^1 y\,e^{-ty}dy \qquad \text{if } y = e^{-x}$$

$$= \frac{e^t-1}{t}; t\in R$$

Since $M_Y(t)$ is the *mgf* of uniform distribution over $(0,1)$, it implies that Y has uniform distribution over $(0,1)$, as the *mgf* determines the distribution uniquely.

[5]

6. Postulates of Poisson Process

$\{N(t)\ ;\ t \geq 0\}$ denote the total number of occurrences of the event E in an interval of duration t.
$\{N(t)\ ;\ t \geq 0\}$ is called a Poisson processes if the following Postulates are satisfied.

1. Independence: $N(t)$ is independent of the number of occurrences in an interval prior to the interval $(0,t)$, *i.e.* future changes in $N(t)$ are independent of the past changes.

2. Homogeneity in time: $P_n(t)$, the probability of the number of occurrences, depends only on the length t of the interval and is independent of where the interval is situated.

3. Regularity: $P_1(t) = \lambda t + o(t)$ and $P_n(t) = o(t); n \geq 2$ where $o(t)/t \rightarrow 0$

Let $N(t)=N_1(t) - N_2(t)$, where $N_1(t)$ and $N_2(t)$ are independent Poisson process.
The *pgf* of $N(t)$ = $E\left(s^{N(t)}\right) = E\left(s^{N_1(t)-N_2(t)}\right) = E\left(s^{N_1(t)}\right) . E\left(s^{-N_2(t)}\right)$

$$= E\left(s^{N_1(t)}\right) E\left(\frac{1}{s}\right)^{N_2(t)}$$

$\Rightarrow N(t)$ is not a Poisson process.

[4]

7. $P\left(\frac{s_1^2}{s_2^2} < 4.03\right)$

$= P(F < 4.03) = 1 - P(F > 4.03)$ with $(9,14)$ *df*

$\simeq 1 - .01 = 0.99$

[3]

8. X denote the number of times 6 will appear when a fair die is tossed

$X \sim B(180, 1/6)$, mean : 30, Variance : 25

$P(29 \leq X \leq 39) = P(28.5 \leq X \leq 39.5)$

(using normal approximation with continuity correction)

$= P(-0.3 \leq Z \leq 1.9)$

$= 0.1179+0.4713 = 0.5892$

$P(X \leq 22) = P(X \leq 21.5)$

$= P(Z < -1.7)$

$= 0.0446$

[4]

9. i) $L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \Sigma x_i}$

$\log L(\lambda) = n \log \lambda - \lambda \Sigma x_i$

$\frac{\partial \log L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0$ gives $\hat{\lambda} = \frac{n}{\Sigma x_i} = \frac{1}{\bar{x}}$

$$\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0 \text{ giving maximum}$$

For the given data $\hat{\lambda} = \frac{20}{6939.5} = 0.0029$

ii) CRLB = $-\frac{1}{E\left[\frac{\partial^2}{\partial \lambda^2}\log L(\lambda)\right]} = \frac{1}{E\left[n/\lambda^2\right]} = \frac{\lambda^2}{n}$

Approximate variance = $\frac{\hat{\lambda}^2}{n} = \frac{(.0029)^2}{20} \simeq 0.4 \times 10^{-6}$

iii) Since $\hat{\lambda} \sim N(\lambda, CRLB)$, the approximate *95%* confidence interval for λ is

$\hat{\lambda} \pm 1.96\sqrt{CRLB}$

$0.0029 \pm 1.96 \times 6.44 \times 10^{-4}$

$0.0029 \pm 0.00126 = (0.00164, 0.00416)$

iv) $L(\lambda) = \prod_{i=1}^{16} \lambda_i e^{-\lambda x_i} \prod_{i=1}^{4} e^{-\lambda y_i}$; $y_i = 600$, $i = 1,2,3,4$

$\log L(\lambda) = 16 \log \lambda - \lambda\left(\sum x_i + \sum y_i\right)$

$\frac{\partial \log L(\lambda)}{\partial \lambda} = 0$ gives $\tilde{\lambda} = \frac{16}{\Sigma x_i + (4 \times 600)}$

$= \frac{16}{6659.7} = 0.0024$

[14]

10.

Note: In the question the distribution of blood group among general public is not given. In fact the proportion of people with different blood groups are known to be different. However, in the absence of this information the problem has been solved as given below under this assumption of equality of different proportions. The law of equal distribution of ignorance is valid, if the proportions of different attributes are not given. The rejection of such a hypothesis would lead to different proportions for the various levels of the attributes in the population.

$n = 200$

Observed frequencies:	$O_A = 92$	$O_B = 20$	$O_{AB} = 4$	$O_0 = 84$
The expected frequencies:	$E_A = 50$	$E_B = 50$	$E_{AB} = 50$	$E_0 = 50$

(under H_0)

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{1}{50}\left[42^2 + 30^2 + 46^2 + 34^2\right] = 118.72$$

Critical value of χ_3^2 at 5% level is 7.815

Reject H_0 that the blood type distribution of people with stomach cancer is same that of the general public.

[5]

11. i) From the data it can be computed as

$s_1 = 40,$ *and* $s_2 = 44$ (approx)

$H_0 : \mu_1 \le \mu_2$; $H_1 : \mu_1 > \mu_2$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(11 \times 1600) + (11 \times 1936)}{22}$$

$$= 1768$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 2.16$$

Critical value for t at $22\ df$ for 5% level is 1.717. Reject H_0

ii) $H_0 : \sigma_1^2 = \sigma_2^2$ $\quad H_1 : \sigma_1^2 \neq \sigma_2^2$

$F = \dfrac{s_2^2}{s_1^2} = 1.21$. Critical value of $F\,(11,11)$ at 5% level is 2.82. Accept H_0

iii) Interval with $(1-\alpha)\,\%$ confidence is

$$\left(\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\, s_{(\bar{x}_1 - \bar{x}_2)}\right)$$

$\bar{x}_1 = 325,\ \bar{x}_2 = 288,\ n_1 = n_2 = 12$

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 17.1659$$

90% Confidence interval is $(325 - 288 \pm (1.717 \times 17.1659)$

$= (7.5261,\ 66.4739)$

[12]

12. Model : $x_{ij} = \mu + \alpha_i + e_{ij}$; $j = 1,2,\ldots,n_I$; $i = 1,2,3$

μ: the overall mean

e_{ij} : random component ~ $N(0.\sigma^2)$

Hypothesis : $H_0 : \alpha_1 = \alpha_2 = \alpha_3$

$H_1 : \alpha_1 \neq \alpha_2 \neq \alpha_3$

$\alpha_i : i = 1,2,3$ are population means at Delhi, Mumbai and Chennai respectively.
Assumptions :

i) Samples are drawn from normal populations
ii) Samples are independent
iii) Population variances are equal

ANOVA

Sources	Sum of squares	d.f.	MSS	F
Treatments	*516*	*2*	*258*	*9.00* (app)
Error	*430*	*15*	*28.67*	
Total	*946*	*17*		

Critical value of $F(2,15)$ at 5% level = 3.682
Reject H_0 . The population means are not equal in these places

[8]

13. $H_0 : \rho = 0 \quad H_1 : \rho > 0$

The test statistic is $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)} df$ and $n = 10$.

If the p-value for one-sided test is 0.025, then observed t value is 2.306

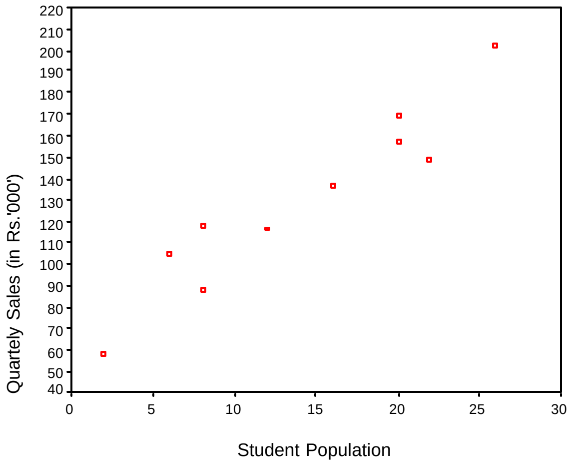
Hence, $2.306 = \dfrac{r\sqrt{8}}{\sqrt{1-r^2}}$ gives

$r^2 = 0.3993$
$r = 0.632$

If Fisher's transformation is used, then $Z = \dfrac{1}{2} log\left(\dfrac{1+r}{1-r}\right)$, which under H_0 has approximately $N(0,1/7)$. This gives $r = 0.630$.

[5]

14. i) SCATTER PLOT



ii) $n = 10,\ \Sigma x_i = 140 \qquad \Sigma y_i = 1300 \qquad \bar{x} = 14 \qquad \bar{y} = 130$

Estimated regression equation is: $y = b_0 + b_1 x$

$$\text{where } b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{2840}{568} = 5$$

$$b_0 = \bar{y} - b_i \bar{x} = 130 - (5 \times 14) = 60$$

$$y = 60 + 5x$$

iii) $H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \quad 0$

The statistic $t = \dfrac{b_1 - \beta_1}{s_{b_1}} \sim t_{(n-2)} df$

$$s_{b_1} = \frac{s}{\sum (x_i - \bar{x})^2} = \frac{13.829}{\sqrt{568}} = 0.5803$$

under H_0 , $t = \dfrac{b_1}{s_{b_1}} = \dfrac{5}{0.5803} = 8.62$

The critical value of t for $n - 2 = 8$ df at $\alpha = 0.01$ is 3.355

Reject H_0 :

iv) 99% confidence interval for β_1 is $b_1 \pm t_{\alpha/2} s_{b_1}$.

Hence the confidence interval is

$$5 \pm (3.355 \times 0.5803)$$

$$= 5 \pm 1.947 \quad = (3.053,\ 6.947)$$

15. *Pgf* of S_N :

i) $p_1(s_N) = E(s^{S_N}) = E(s^{X_1 + X_2 + \ldots + X_N})$

$= E\left[E(s^{X_1+X_2+\ldots+X_N} / N)\right]$

$= \sum_{n=0}^{\infty} E\left[(s^{X_1+X_2+\ldots+X_N} / N)\right] P(N = n)$

$= \sum_{n=0}^{\infty} E(s^{X_1+\ldots+X_N}).P(N = n)$. Since $X_i^{'s}$ and N are independent.

$= \sum_{n=0}^{\infty} (p(s))^n P(N = n)$; $p(s)$ is the *pgf* of X_i

$= P_N(p(s))$; P_N being the *pgf* of N

ii) $p_1^{'}(s_N) = P_N^{'}(p(s))p^{'}(s)$

Hence $E(S_N) = P_N^{'}(p(1))\, p^{'}(1)$

$= P_N^{'}(1)p^{'}(1)$

$= E N. EX$

[12]

****