Institute of Actuaries of India

CT3: Probability and Mathematical Statistics

May 2009 Examination

Indicative Solutions

1. a) 12 14 19 20 21 28 29 30 32.8(mean) 55 63

___._.______._.__._____.__.__.____.____________________________.________ 63._______

28 (median)

b) Median 28

Mean 32.18

Locations in dot plot

c) Sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

$= 19.07$

[6]

2. $n_1 + n_2 = 150$

$$\frac{(n_1 \times 70) + (n_2 \times 55)}{n_1 + n_2} = 60$$

$\Rightarrow 70n_1 + 55n_2 = 9000$

Solving $n_1 = 50 \quad n_2 = 100$

[3]

3. Given that $P(A) > 0$ and $P(B) > 0$

If A and B are mutually exclusive, $A \cap B = \phi$

$\Rightarrow P(A \cap B) = 0 \neq P(A)P(B)$. Hence not independent

If A and B are independent $P(A \cap B) = P(A)P(B)$

$\Rightarrow P(A \cap B) > 0$. Hence, $A \cap B \neq \phi$. Therefore not mutually exclusive **[2]**

4. $P(X = x) = \binom{3}{x}(0.4)^x (0.6)^{3-x}; x = 0,1,2,3.$

$0.196 < F(0) = 0.216$, $0.351 < F(1) = 0.648$ and $0.975 < F(3) = 1$

Hence, the simulated values are 0,1,3 **[3]**

5.a) If $B_1, B_2, ..., B_n$ are such that $B_i^{'s}$ are disjoint and $\cup B_i = \Omega$, then for any $A \subset \Omega$

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

If $B_1, ..., B_n$ are mutually exclusive of which one must occur, then

$$P(B_i / A) = \frac{P(B_i)P(A / B_i)}{\sum_{i=1}^{n} P(B_i)P((A / B_i)} \quad ; \quad \text{for } i=1,2,\ldots,n$$

b)

$$\left.\begin{array}{l} P(A) = 0.3, P(B) = 0.45, P(C) = 0.25 \\ P(F / A) = 0.01, P(F / B) = 0.01, P(F / C) = 0.02 \end{array}\right\}$$

$$P(C / F) = \frac{P(F / C)P(C)}{P(F / A)P(A) + P(F / B)P(B) + P(F / C)P(C)}$$

$$= \frac{0.02 \times 0.25}{(0.01 \times 0.3) + (0.01 \times 0.45) + (0.02 \times 0.25)}$$

$= 0.4$ **[5]**

6. a) $\int_0^{Q_1} \frac{x}{a^2} e^{\frac{-x^2}{2a^2}} dx = \frac{1}{4}$

$\Rightarrow Q_1 = a\sqrt{2}\sqrt{\log\left(\frac{4}{3}\right)}$

Similarly, $\int_0^{Q_{31}} \frac{x}{a^2} e^{\frac{-x^2}{2a^2}} dx = \frac{3}{4}$

$\Rightarrow Q_3 = a\sqrt{2}\sqrt{\log 4}$

Hence, $IQR = Q_3 - Q_1 = a\sqrt{2}\left[\sqrt{\log 4} - \sqrt{\log(4/3)}\right]$

b) $E(X) = \int_0^{\infty_1} x \frac{x}{a^2} e^{\frac{-x^2}{2a^2}} dx = a\sqrt{\pi/2}$

$E(X^2) = \int_0^{\infty_1} x^2 \frac{x}{a^2} e^{\frac{-x^2}{2a^2}} dx = 2a^2$

SD= $\sqrt{E(X^2) - (E(X))^2} = \sqrt{(2 - \pi/2)}$

Hence, $\frac{Q_3 - Q_1}{\sqrt{V(X)}} = \frac{\sqrt{2}\left[\sqrt{\log 4} - \sqrt{\log 4/3}\right]}{\sqrt{(2-\pi/2)}}$ is free of a **[6]**

7. $H_0 : \mu = 80.0 \quad H_1 : \mu \neq 80.0$

$n = 100 \quad \sigma = 8.40$

Critical region : $\overline{X} < 78.0$

Prob (Type I error) : $P_{H_0}(\overline{X} < 78.0)$

$= P(Z < \frac{78.0 - 80.0}{8.4/10})$

$= P(Z < -2.381) = 0.0087$

Power at $\mu = 82$: $P_{\mu=82}(\overline{X} < 78.0)$

$= P(Z < \frac{78. - 82}{8.4/10})$

$= P(Z < -4.762) \cong 0$ **[3]**

8. A compound distribution is a distribution that consists of a random number of random variables.

a) N has negative binomial distribution with $k = 3$ and $p = 0.9$

Hence, $M_N(t) = \left(\frac{0.9}{1 - 0.1\, e^t}\right)^3$

The claim amounts X follow Gamma (6,2). Hence

$M_X(t) = (1 - t/2)^{-6}$

Let $Y = X_1 + ... + X_N$

$M_Y(t) = E\left[e^{tY}\right] = E\left\{E\left[e^{t\Sigma X_i} \mid N\right]\right\} = E\left[(1 - t/2)^{-6N}\right]$

$$= E\left[e^{-6N\log(1-t/2)}\right]$$

$$= \left[\frac{0.9}{1-0.1\left(1-\frac{t}{2}\right)^{-6}}\right]^3$$

b) $E(X)=3$, $V(X)=1.5$, $E(N)=1/3$, $V(N)=10/27$

$V(Y)=E(N).V(X)+E^2X.V(N)$

The standard deviation of Y is $\sqrt{3\frac{5}{6}}$

[6]

9.a) The *pmf* of X :

$$p(x)=\sum_{y=0}^{x}p(x,y)=\frac{\lambda^x e^{-\lambda}}{x!}\sum_{y=0}^{x}\frac{x!\,p^y(1-p)^{x-y}}{y!(x-y)!}$$

$$= \frac{\lambda^x e^{-\lambda}}{x!}\sum_{y=0}^{x}\binom{x}{y}p^y(1-p)^{x-y};\ x=0,1,2,...$$

$$= \frac{e^{-\lambda}\lambda^x}{x!};\ x=0,1,2,...$$

The *pmf* of Y *:*

$$p(y)=\sum_{x=y}^{\infty}p(x,y)=\sum_{x=y}^{\infty}\frac{\lambda^x e^{-\lambda}p^y(1-p)^{x-y}}{y!(x-y)!};(x\ge y)$$

$$= \frac{e^{-\lambda}(\lambda p)^y}{y!}\sum_{x=y}^{\infty}(x-y)!$$

$$= \frac{e^{-\lambda}(\lambda p)^y}{y!}e^{\lambda(1-p)} = \frac{e^{-\lambda p}(\lambda p)^y}{y!};\ y=0,1,2,...$$

b) Conditional distribution of Y given X :

$$p(y/x)=\frac{p(x,y)}{p(x)}=\binom{x}{y}p^y(1-p)^{x-y};\ x\ge y,\ y=0,1,...,x$$

The conditional distribution is binomial *(x,p)*

The conditional distribution of X given Y :

$$p(x/y)=\frac{p(x,y)}{p(y)}=\frac{e^{-\lambda(1-p)}(\lambda(1-p))^{x-y}}{(x-y)!};\ x\ge y$$

The conditional distribution is Poisson rv $X-Y$ with parameter $\lambda(1-p)$

c) $E(Y/X=x)=xp$

$E(X/Y=y)=y+\lambda(1-p)$

[10]

10.a) Likelihood $L(\theta / \underline{x}) = \begin{cases} \dfrac{1}{(2\theta)^n} & ; \ -\theta \le x_{(1)} \le x_{(n)} \le \theta \\ 0 & otherwise \end{cases}$

The MLE cannot be found by the usual calculus method

The likelihood is maximum if $-\theta \le X_{(1)} \ or \ X_{(n)} \le \theta$

The MLE is $Max(-X_{(1)}, X_{(n)})$

b) $EX = 0 = \overline{X}$

$$V(X) = \frac{4\theta^2}{12} = \frac{1}{n}\Sigma(X_i - \overline{X})^2$$

$$\Rightarrow \hat{\theta} = \sqrt{\frac{3(n-1)}{n}S^2}\ ; S^2 \text{ being sampling variance}$$

(Alternatively $\hat{\theta}$ can be estimated using $\sqrt{\frac{3}{n}\Sigma X_i^2}$

c) CRLB : $\dfrac{1}{E_\theta\left(\dfrac{\partial}{\partial\theta}\log f_\theta(\underline{x})\right)^2}$ or $\dfrac{1}{-E_\theta\left(\dfrac{\partial^2}{\partial\theta^2}\log f_\theta(\underline{x})\right)}$

CRLB cannot be applied in this case, since the <u>regularity conditions</u> are not satisfied.

d) ML estimate : *Max* (4.5, 3.8) = 4.5

Moment estimate : $\sqrt{\frac{3(n-1)S^2}{n}} = \sqrt{3 \times \frac{9}{10} \times 6.42} = 4.16$

where $S^2 = \frac{1}{n-1}\Sigma(x_i - \overline{x})^2 = 6.42$

(Alternate moment estimate is $\sqrt{\frac{3}{10} \times 102.16} = 5.54$

[10]

11.a) $H_0 : \mu_X = \mu_Y = \mu_Z$ H_1 : At least one of the equalities does not hold

Assumptions :
1. The samples are drawn from normal populations
2. The populations are independent
3. The population variances are equal

ANOVA				
Source	S.S.	Df	MSS	F
Between Fertilizers	2.465	2	1.233	13.5
Error	1.092	12	0.091	
Total	3.557	14		

Table value of $F_{(2,12)}$ at 5% level is 3.89

H_0 is rejected. There is evidence that the fertilizers produce different mean yields.

b) Combining the observations under *X* and *Z* together and calling it as *U*; and keeping *Y* separately we have $H_0 : \mu_u = \mu_y$ $H_1 : \mu_u \ne \mu_y$ (or $\mu_u < \mu_y$)

$\bar{u} = 3.96 \quad \bar{y} = 4.82 \quad s_1^2 = 0.07 \quad s_2^2 = 0.17$

The pooled *s.d* is $s = 0.2892$

Hence, $t = \dfrac{\bar{u} - \bar{y}}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

$= -5.417$

The table value of *t* for 13 *df* at 5% level (two sided test) is 2.160
(The table value of t for 13 *df* at 5% level (one sided test) is 1.771)

Reject H_0

[7]

12. The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process having rate $\lambda, \lambda > 0$, if

i) $N(0) = 0$ ii) has stationary and independent increments

iii) $P(N(h) = 1) = \lambda h + 0(h)$ iv) $P(N(h) \geq 2) = 0(h)$

The *pgf* $p(s)$ of X_i is given by $p(s) = \sum_{k=1}^{\infty} \frac{s^k}{2^k} = \frac{s}{2-s}$

Let T be the total number of persons involved in accidents during a week given that the number of accidents follows a Poisson process with rate 2 per day.

Therefore, the *pgf* of *T* is given by

$$G(s) = \exp\left\{2 \times 7\left(\frac{2s-2}{2-s}\right)\right\}$$

On differentiation we get

$$G'(s) = \exp\left\{14\left(\frac{2s-2}{2-s}\right)\right\} 14 \frac{2}{(2-s)^2}$$

$$G''(s) = \exp\left\{28\left(\frac{s-1}{2-s}\right)\right\}\left(\frac{28^2}{(2-s)^4} + \frac{56}{(2-s)^3}\right)$$

Hence, Mean: $G'(1) = 28$

Variance: $G''(1) + G'(1) - (G'(1))^2 = 840 + 28 - (28)^2$

$= 84$

Alternatively :

$E(N(t)) = 2t \quad E(X_1) = 2 \quad V(N(t)) = 2t \quad V(X_1) = 4$

$E(T) = E(X_1)E(N(7))$

$= 2 \times 2 \times 7 = 28$

$V(T) = E(N(7))V(X_1) + \{E(X_1)\}^2 V(N(7))$

$= 84$

[5]

13. $n = 20$, $s_X^2 = 1.2011,$ $s_Y^2 = 2.2958$

$s_{(X-Y)}^2 = 1.8673$

$s_{(X-Y)}^2 = s_x^2 + s_y^2 - 2s_{xy}$

a) Hence $S_{xy} = \frac{1}{2}\left[s_x^2 + s_y^2 - s_{x-y}^2\right]$

$= \frac{1}{2}\left[1.2011 + 2.2958 - 1.8673\right]$

$= 0.8148$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{0.8148}{\sqrt{1.2011}\sqrt{2.2958}} = 0.491$$

b) 95% confidence interval for the population correlation coefficient ρ is

$$r \pm t_{(n-2),\alpha/2}\frac{\sqrt{(1-r^2}}{\sqrt{n-2}}$$

$$0.491 \pm 2.101 \times 0.206 = (0.058, 0.924)$$

c) Since the confidence interval does not include 0, we reject $H_0 : \rho = 0$

[5]

14.a) $H_0 : \mu_1 = 2\mu_2$, $H_1 : \mu_1 \neq 2\mu_2$

$$Z = \frac{\bar{x} - 2\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{4\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$= \frac{52 - 72}{\sqrt{\frac{70}{20} + \frac{4 \times 90}{17}}} = \frac{-20}{4.97} = -4.024$$

Reject H_0

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

Government school : $s_1^2 = \frac{1312}{19} = 69.05$

Private school : $s_2^2 = \frac{1401}{16} = 87.56$

$$F = \frac{s_2^2}{s_1^2} = 1.27$$

Table value of $F_{(16,19)}$ at 5% level is 2.21

Do not reject H_0

c) $H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{1/n_1 + 1/n_2}} = \frac{52 - 36}{8.89\sqrt{\frac{1}{20} + \frac{1}{17}}} = \frac{16}{2.936}$$

$= 5.45$

Table value of t for 35 *df* at 5% level is 2.031

Reject H_0

Comment: If H_0 in b) is rejected, H_0 in c) cannot be tested using t statistic.

d) Confidential interval for $(\mu_1 - \mu_2)$ at 95% level is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(52 - 36) \pm 2.031 \times 8.9 \sqrt{\frac{1}{20} + \frac{1}{17}}$$

$$16 \pm 5.963 = (10.037, 21.963)$$

[10]

15.a) $\Sigma x = 5.5$, $\Sigma y = 134.3$, $\Sigma xy = 78.88$
$\Sigma x^2 = 3.85$ $\Sigma y^2 = 1834.25$

$$S_{xy} = \Sigma(x - \bar{x})(y - \bar{y}) = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 5.015$$

$$S_{xx} = \Sigma(x - \bar{x})^2 = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 0.825$$

$$S_{yy} = \Sigma(y - \bar{y})^2 = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 30.60$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{5.01}{0.82} = 6.078$$

$$\hat{\alpha} = \bar{y} - 6.078 = 13.43 - (6.078 \times .55) = 10.087$$

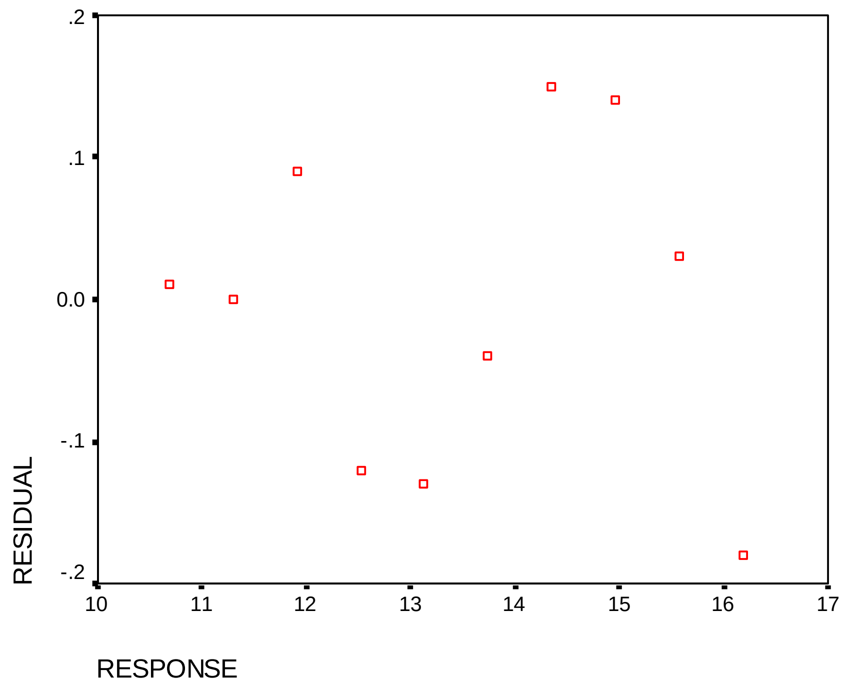
b) The least squares estimates in a) continue to be ML estimates, since errors are *iid* and normally distributed

c)
$$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{YY} - \frac{S_{xy}^2}{S_{xx}}\right)$$

$$= \frac{1}{8}\left(30.6 - \frac{(5.015)^2}{0.825}\right) = 0.1198 \approx 0$$

d)

Y	$\hat{Y}$	R
10.7	10.69	0.01
11.3	11.30	0.00
12.0	11.91	0.09
12.4	12.52	- 0.12
13.0	13.13	- 0.13
13.7	13.74	- 0.04
14.5	14.35	0.15
15.1	14.96	0.14
15.6	15.57	0.03
16.0	16.18	-0.18



The fit is good

[12]

16.a) H_0 : The number of accidents is equally distributed among various times.

Observed frequency ; *O :*	14	16	24	22	24	20
Expected frequency ; *E :*	20	20	20	20	20	20

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

$$= \frac{36}{20} + \frac{16}{20} + \frac{16}{20} + \frac{4}{20} + \frac{16}{20} + 0 = 4.4$$

The table value of χ^2 at 5% level for 5 *df* is 11.07

Do not reject H_0

b) H_0 : There is no association between accidents proneness and colour blindness

Observed frequency :	22	5	38	15
Expected frequency :	20.25	6.75	39.75	13.25

$$\chi^2 = \frac{1.75^2}{20.25} + \frac{1.75^2}{6.75} + \frac{1.75^2}{39.75} + \frac{1.75^2}{13.25}$$

$= 0.913$

The table value of χ^2 for 1 *df* at 5% level is 3.841

Do not reject H_0

[7]

[Total 100 Marks]

****