Institute of Actuaries of India

Subject CT3-Probability and Mathematical Statistics

May 2008 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.

1. a. Mean 140.4167; Median 128

b. The least two points are 113 and 116. The largest two points are 179 and 239.

So, the trimmed mean is [140.42*12-(113+116+179+239)]/8 = 129.75

c. Comment:

i. Mean gives equal weight to all the observations so it is more likely to get impacted by the extreme observations.

ii. Median doesn't depend on the 'size' of the observations but only on the 'order'. So it is least likely to get impacted by the extreme observations.

iii. Impact of outliers on the trimmed mean depends on the number of observations being trimmed. In any case it is less impacted by extreme observations than the simple mean.

d.

$$
\begin{aligned}
&\frac{11S^2}{1300} = X \sim \chi^2_{11}, Now \\
&P(S^2 > 1300) \\
&= P(\frac{11S^2}{1300} > 11) \\
&= P(X > 11) \\
&= 0.4433
\end{aligned}
$$

Total [8]

2.
$$
\begin{aligned}
P(A_1 \cap A_3 / A_2) &= P(A_1 \cap A_2 \cap A_3) / P(A_2) \\
&= P(A_1 / A_2 \cap A_3) P(A_2 \cap A_3) / P(A_2) \\
&= P(A_1 / A_2 \cap A_3) P(A_3 / A_2) \\
&= (0.4)\,(0.8) \quad = 0.32
\end{aligned}
$$

Total [2]

3.

$$
\left.
\begin{aligned}
P(A) &= \frac{18}{36} = \frac{1}{2} \\
P(B) &= \frac{18}{36} = \frac{1}{2} \\
P(C) &= \frac{4}{36} = \frac{1}{9}
\end{aligned}
\right\}
$$

$$
P(A \cap B \cap C) = \frac{1}{36} = P(A)P(B)P(C)
$$

However,

$$\left.\begin{aligned} P(A\cap B) &= \frac{1}{6} \neq P(A)P(B) \\ P(B\cap C) &= \frac{1}{12} \neq P(B)P(C) \\ P(A\cap C) &= \frac{1}{36} \neq P(A)P(C) \end{aligned}\right\}$$

A, *B* and *C* are not mutually independent.

Total [4]

4. $G_X(t) = E(t^X) = \sum_x \frac{1}{2} pq^{|x|-1}\ t^x \qquad ; x = \pm 1, \pm 2 \ldots$

$$= \frac{1}{2}\ \frac{p}{q} \sum_x t^x q^{|x|}$$

$$= \frac{1}{2}\left(\frac{p}{q}\right)\left[\left(tq + t^{-1}q\right) + \left(t^2q^2 + t^{-2}q^2\right) + \ldots\right]$$

$$= \frac{p}{2q}\left[q\left(t + t^{-1}\right) + q^2\left(t^2 + t^{-2}\right) + \ldots\right]$$

$$= \frac{p}{2q}\left[\sum_{r=1}^{\infty}(qt)^r + \sum_{r=1}^{\infty}\left(\frac{q}{t}\right)^r\right]$$

$$= \frac{p}{2q}\left[(qt)(1-qt)^{-1} + \frac{q}{t}\left(1 - \frac{q}{t}\right)^{-1}\right]$$

$$= \frac{pq}{2q}\left[\frac{t}{(1-qt)} + \frac{1}{(t-q)}\right];\ q < |t| < \frac{1}{q}$$

$$= \frac{p}{2}\left[\frac{t}{(1-qt)} + \frac{1}{(t-q)}\right];\ q < |t| < \frac{1}{q}$$

Total [3]

5. a. $f(x) = \dfrac{1}{x\sqrt{2\pi}} \quad e^{-[\log x - 3.2]^2/2} \ ; x \geq 0$

$= 0 \quad for\ x < 0$

b. $EX = e^{\mu + \sigma^2/2} = e^{3.2 + 1/2} = e^{3.7}$

$Var\ X = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) = e^{7.4}(e-1)$

c. $P(X > 8) = 1 - P(X \leq 8)$. Since $\log X \sim N(\mu = 3.2, \sigma = 1)$,

$P(X \leq 8) = \Phi\left[\dfrac{\log 8 - 3.2}{1}\right]$

$= \Phi\ (-1.12)$, Φ denotes the *cdf* of standard normal.

$= 0.1314$

$P(X > 8) = 1\text{-}0.1314 = 0.8686$

Total [4]

6. a. Most suitable distribution for N is Binomial (90, p) where p is the probability of head.
Estimate of p is 50/90. Mean of N is 90*(50/90) = 50.
Variance is 90*(50/90) * (1-50/90) = 50*40/90 = 22.22.

b. N approximately follows Normal with mean 45 and variance 22.5. Using continuity correction,

$P_{Bin}(N > 50)$

$= P_{Nor}(N \geq 50.5)$

$= P_{Nor}\left(Z \geq \frac{(50.5-45)}{\sqrt{22.5}}\right)$ where $Z \sim N(0, 1)$

$= P_{Nor}(Z \geq 1.16)$

$= 0.123$

c. Using the above probability, the P-value of the test is 0.123. We do not have sufficient evidence to reject H_o at 5% significance level.

d. H_o should be rejected for P-value < 0.05. Let n be the desired number of heads. Then

$$P(N \geq n) = 0.05$$

$$\Rightarrow P\left(\frac{(N-45)}{\sqrt{225}} \geq \frac{(n-45)}{\sqrt{225}}\right) = 0.05$$

$$\Rightarrow \frac{(n-45)}{\sqrt{225}} = 1.64$$

$$\Rightarrow n = 1.64 * \sqrt{225} + 45$$

$$\Rightarrow n = 5277$$

So, H_o should be rejected for the number of heads $\geq$ 53.

Total [7]

7. $X_i \sim U(0,1)$

$EX_i = \frac{1}{2}$, $Var\ X_i = \frac{1}{12}$; $i = 1,2,\ldots,20$

Given that $Y = \sum_{i=1}^{20} X_i$ and hence $Y \sim N\left(10, \frac{5}{3}\right)$

a. $P[Y \leq 9.1] = P\left[\frac{Y-10}{\sqrt{5/3}} \leq \frac{9.1-10}{\sqrt{5/3}}\right]$

$= P[Z \leq -0.697]$

$= \Phi(-0.697) = 0.2423$

b. $P[8.5 \le Y \le 11.7] = P\left[\frac{8.5-10}{\sqrt{5/3}} \le \frac{Y-10}{\sqrt{5/3}} \le \frac{11.7-10}{\sqrt{5/3}}\right]$

$= P[\,-1.162 \le Z \le 1.317]$

$= \Phi(1.317) - \Phi(-1.162)$
$= 0.9061 - 0.1226 = 0.7835$

Total [3]

8. a. P (Wait for ten minutes)

$=P$(0,1,2 passengers will arrive within the next ten minutes)
$$= e^{-\frac{10}{3}} + e^{-\frac{10}{3}}(\frac{10}{3}) + e^{-\frac{10}{3}}(\frac{10}{3})^2 \frac{1}{2!} = 0.3528$$

b. Solving this part depends on the memoryless property of Poisson process. The waiting period before the arrival of next passenger is exponentially distributed with an expected value of three minutes.
P(having to wait another five minutes)= P (nobody arrives in the next five minutes)
$$= e^{-\frac{5}{3}} = 0.1883$$

Total [5]

9. a. $\int_0^{\theta} \frac{kx}{\theta^2} dx$

$$= \frac{k}{\theta^2}\int_0^{\theta} x dx$$
$$= \frac{k}{\theta^2}[\frac{x^2}{2}]_0^{\theta}$$
$$= \frac{k}{\theta^2} * \frac{\theta^2}{2}$$
$$= \frac{k}{2}$$

For the above integral to be 1, k should be 2.

b. The likelihood function is

$$L(\underline{x}, \theta) = 2^n \theta^{-2n} \prod_{i=1}^{n} x_i \quad ; \quad 0 < x_i < \theta$$

(Differentiating this wrt θ and equating to 0 will not give a solution for θ in terms of x).
L is a decreasing function of θ
L is positive only when $\theta \ge Max(x_i)$
So L is the maximum when $\theta = Max(X_i)$; which is the MLE of θ.

Total [5]

10. a. For Exponential λ, $f(x,\lambda) = \lambda e^{-\lambda x}$ and $F(x) = 1 - e^{-\lambda x}$, x > 0

The likelihood function is

$$L(\lambda, x) = \lambda^7 * e^{-\lambda(23+27+39+52+68+89+95)} * e^{-300\lambda}$$

b. The log-likelihood is

$$\ell(\underline{x}, \lambda) = 7\ln\lambda - 693\lambda$$

$$\frac{\partial \ell(\underline{x}, \lambda)}{\partial x} = \frac{7}{\lambda} - 693$$

Equating above to 0 gives MLE of λ = 7/693 =.0101

c.

$$\frac{\partial^2 \ell(\underline{x}, \lambda)}{\partial^2 x} = -\frac{7}{\lambda^2}$$

$$So, E\left[\frac{\partial^2 \ell(\underline{x}, \lambda)}{\partial^2 x}\right] = -\frac{7}{\lambda^2}$$

CRLB for the MLE is $\dfrac{1}{-E\left[\dfrac{\partial^2 \ell(\underline{x}, \lambda)}{\partial^2 x}\right]} = \dfrac{\lambda^2}{7} =$ .000015 which is same as the asymptotic variance of the MLE

d. The 95% confidence interval for λ is $\hat{\lambda} \pm 1.96.S.D.(\hat{\lambda})$.

Using the MLE of λ and the variance the interval is:

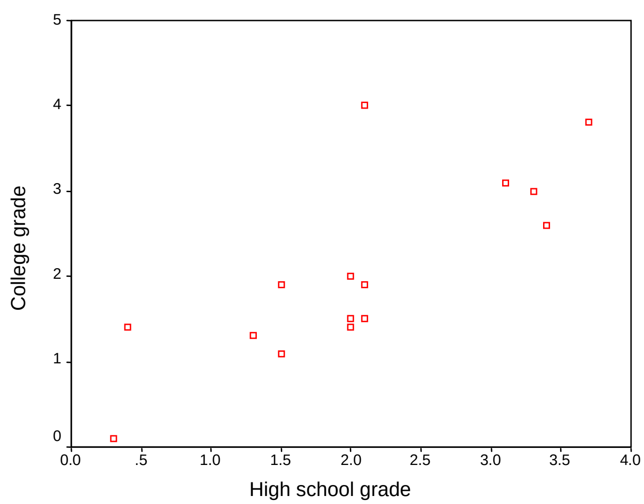
(0.0101 1.96*0.0039) i.*e*.(0.0025,0.0177)

e. The likelihood is

$$L(\lambda; \underline{x},) = [1 - e^{-25\lambda}]^3 * \lambda^7 * e^{-\lambda[(23+27+39+52+68+89+95)+7*25]} * e^{-[300+3*25]\lambda}$$

Total [12]

11.a. The Scatter Plot



b. $\hat{\alpha} = 0.346$; $\hat{\beta} = 0.825$; $\hat{r} = 0.778$;

c. As the estimate of the slope is lower than 1, the expected performance in college is lower than the performance in high school.

d. $\hat{y} = \hat{\alpha} + \hat{\beta}x = 0.346 + 0.825 * 2.1 = 2.0785$.
Actual college grade for a high school grade of 2.1 for observation 10 is much higher than the expected. This observation can be treated as an outlier while fitting the regression.

Total [11]

12. $$\bar{Y}_{..} = \frac{(n_A * \bar{Y}_{A.} + n_B * \bar{Y}_{B.} + n_C * \bar{Y}_{C.})}{(n_A + n_B + n_C)}$$

$$\bar{Y}_{..} = \frac{8*100 + 10*110 + 9*95)}{8 + 10 + 9} = \frac{2755}{27} = 102.037$$

a. $\hat{\mu} = \bar{Y}_{..} = 102.037$

$$\hat{\tau}_A = \bar{Y}_{A.} - \bar{Y}_{..} = 100 - 102.037 = -2.037$$
$$\hat{\tau}_B = \bar{Y}_{B.} - \bar{Y}_{..} = 110 - 102.037 = 7.963$$
$$\hat{\tau}_C = \bar{Y}_{C.} - \bar{Y}_{..} = 95 - 102.037 = -7.037$$
$$\hat{\sigma}^2 = \frac{SSR}{n_{..} - k} = \frac{1075}{24} = 44.79$$

b. $SS_B = \sum_i n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$ =8*4.15+10*63.41+9*49.52= 1112.98

$SS_T = SS_B + SS_R = 1112.98 + 1075 = 2187.98$

c. The ANOVA table is:

Source	DF	*SS*	*MS*	*F*
Between	2	1112.98	556.49	12.42
Residual	24	1075.00	44.79	
Total	26	2187.98		

$F_{0.05, (2, 24)}$ is 3.403. So there is enough evidence to reject the null hypothesis of equality of all fertilizer effects.

d. The 95% confidence interval for $\mu + \tau_A$ is given by

$$(Y_{A.} - t_{0.025,24}\frac{\hat{\sigma}}{n_A}, Y_{A.} + t_{0.025,24}\frac{\hat{\sigma}}{n_A})$$
$$= (100 - 2.064 * \frac{44.79}{8}, 100 + 2.064 * \frac{44.79}{8})$$
$$= (88.44, 111.56)$$

Total [10]

13. a. $\Sigma X = 378$, $\Sigma Y = 333$, $\Sigma X^2 = 16112$, $\Sigma Y^2 = 1347$, $\Sigma XY = 14340$

$$r = \frac{\frac{1}{n}\Sigma XY - \overline{X}\,\overline{Y}}{\sqrt{\frac{1}{n}\Sigma X^2 - \left(\overline{X}\right)^2}\sqrt{\frac{1}{n}\Sigma Y^2 - \left(\overline{Y}\right)^2}}$$

$= 0.679514 = 0.680$ (App)

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

Test statistic is $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

$$= \frac{0.68\sqrt{7}}{\sqrt{1-0.68^2}} = 2.453$$

The table value of t distribution for 7 df at 5% level is 1.895.

Conclusion: Reject H_0.

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

(Usually the larger variance is taken in the numerator)

$\overline{X} = 42$; $s_X^2 = 29.5$ and $\overline{Y} = 37$; $s_Y^2 = 143.75$

$F_{cal} = S_Y^2 / S_X^2 = 4.87$ and $F_{tab}(8,8) = 3.44$

Conclusion: Reject H_0

c. 90% confidence interval for $\frac{\sigma_Y^2}{\sigma_X^2}$ is $\left(\frac{1}{F_{\alpha/2}(8,8)}\frac{s_Y^2}{s_X^2}, F_{\alpha/2}(8,8)\frac{s_Y^2}{s_X^2}\right)$

$= [1.416, 16.757]$, $F_{0.5}(8,8) = 3.44$

(The CI for $\sigma_X^2 / \sigma_Y^2 = [.0596, 0.7059]$

d. Changes in Scores (d): 5 -7 -11 8 - 11 – 21 0 -9 1

(d = Y-X)

$\overline{d} = -5; s_{\overline{d}} = 9.206$ and $n = 9$.

$H_0 : \mu_d = 0$ Vs $H_1 : \mu_d > 0$

The Test statistic is : $t = \frac{\overline{d}}{\left(\frac{s_{\overline{d}}}{\sqrt{n}}\right)} = \frac{-5}{\left(\frac{9.206}{3}\right)} = -1.629$

Table value $t_{0.05}(8) = 1.86$; Conclusion: Do not reject H_0

e. 95% confidence interval for the difference between the two population means $(\mu_y - \mu_x)$ is

$$\{(\mu_y - \mu_x) \pm (t_{\alpha/2, n-1})\left(\frac{s_{\bar{d}}}{\sqrt{n}}\right)\}$$

which simplifies to (-12.077,2.077).

(The confidence interval for $(\mu_x - \mu_y)$ will therefore be (-2.077, 12.077).)

Total [14]

14 a. Note that value of k in the density does not affect the conditional expectation.

E(X/Y=y):

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

$$= k\int_{10}^{Y}(x-10)(y-10)dx$$

$$= k\,(y-10)^3; \quad 10 < y < 30$$

Hence, $f_{X/Y}(x/y) = \dfrac{2(x-10)}{(y-10)^2}; \quad 10 < x < y$

$= 0$; otherwise

$$E(X/Y=y) = \int_{10}^{y} x\frac{2(x-10)}{(y-10)^2}\,dx$$

$$= 10 + \frac{2}{3}(y-10); \; 0 < y < 30$$

b. $P(X > 20/Y = 25) = \displaystyle\int_{0}^{20} \frac{2(x-10)}{(25-10)^2}dx \quad = 5/9$

Total [7]

15. The *pmf* of *X*

$$p_X(x) = \binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x}$$

$$= \binom{n}{x}\left(\frac{1}{2}\right)^n; x = 0,1,2,...n$$

and $P(N = n) = \dfrac{n}{3}; n = 1,2$

$$P(X = x, N = n) = P(X = x/N = n)P(N = n)$$

$$= \binom{n}{x}\left(\frac{1}{2}\right)^n\frac{n}{3}; x = 0,1,2,...n; n = 1,2$$

Hence, $P(X = x) = \sum_{n} P(X = x, N = n)$

Then, $P(X = 0) = P(X = 0, N = 1) + P(X = 0, N = 2) = \frac{1}{6} + \frac{1}{6} = 1/3$

$P(X = 1) = P(X = 1, N = 1) + P(X = 1, N = 2) = \frac{1}{6} + \frac{1}{3} = 1/2$

$P(X = 2) = P(X = 2, N = 2) = 1/6$

Mean of X: 5/6

Variance of X: 17/36

Total [5]

****