Subject CT3 — Probability and Mathematical Statistics Core Technical

EXAMINERS' REPORT

April 2008

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

M A Stocker
Chairman of the Board of Examiners

June 2008

1 $n = 80$, $\Sigma fx = 22$, $\Sigma fx^2 = 40$

$\bar{x} = 22/80 = 0.275$

$$s^2 = \frac{1}{79}\left(40 - \frac{22^2}{80}\right) = \frac{33.95}{79} = 0.42975 \Rightarrow s = 0.656$$

2 $\Sigma x = n\bar{x} = 29(461.5) = 13383.5$

$\Sigma x^2 = (n-1)s^2 + n\bar{x}^2 = 28(618.8)^2 + 29(461.5)^2 = 16898062$

removing the outlier of 3657.5 gives

$\Sigma x = 13383.5 - 3657.5 = 9726$

$\Sigma x^2 = 16898062 - 3657.5^2 = 3520756$

$$\therefore \bar{x} = \frac{9726}{28} = £347.4$$

$$s^2 = \frac{1}{27}[3520756 - \frac{9726^2}{28}] = 5272.6 \quad \therefore s = £72.6$$

3 (i) $\bar{x} = \dfrac{23778}{13} = 1829.08.$

Median = 7^{th} ordered observation = 1614.

(ii) The median should be preferred, as it is not sensitive to the extreme observed claim of £4320.

4 $$P(A\,|\,B) = \frac{P(A \cap B)}{P(B)}$$

Maximum value of $P(B)$ is 0.8 in which case $P(A\,|\,B) = \dfrac{0.1}{0.8} = 0.125$

Minimum value of $P(B)$ is 0.1, in the case $B \subset A$. Then $P(A\,|\,B) = 1$

5 Define the following events:

C: Policy results in a claim;

B_i: Policy comes from portfolio i, $i = 1, 2, 3, 4$.

Then the required probability is $P(B_3|C)$, and using Bayes' theorem:

$$P(B_3 \mid C) = \frac{P(C \mid B_3)P(B_3)}{P(C)} = \frac{P(C \mid B_3)P(B_3)}{\sum_i P(C \mid B_i)P(B_i)},$$

which gives

$$P(B_3 \mid C) = \frac{0.02 \times \frac{13}{30}}{0.08 \times \frac{4}{30} + 0.05 \times \frac{7}{30} + 0.02 \times \frac{13}{30} + 0.04 \times \frac{6}{30}} = \frac{0.26/30}{1.17/30} = \frac{0.26}{1.17} = 0.222 .$$

[OR It is possible to argue straight to
260/(320 + 350 + 260 + 240) = 260/1170 = 0.222
which is correct and gets full marks.]

6 (i) $$f(y \mid x) = \frac{f(x,y)}{f(x)}$$

$$= \frac{\frac{4}{3}(1-xy)}{\frac{2}{3}(2-x)} = 2\frac{(1-xy)}{(2-x)}, \quad 0 < y < 1$$

(ii) (a) $$E(Y \mid X = x) = \frac{2}{(2-x)}\int_0^1 y(1-xy)dy$$

$$= \frac{2}{(2-x)}\left[\frac{y^2}{2} - x\frac{y^3}{3}\right]_0^1 = \frac{2}{(2-x)}\left(\frac{1}{2} - \frac{x}{3}\right) = \frac{(3-2x)}{3(2-x)}$$

$$E(Y) = \int_0^1 \frac{(3-2x)}{3(2-x)} \frac{2}{3}(2-x)dx$$

$$= \frac{2}{9}\int_0^1 (3-2x)dx = \frac{2}{9}[3x - x^2]_0^1 = \frac{4}{9}$$

(b) $E(Y) = \frac{2}{3}\int_0^1 y(2-y)dy = \frac{2}{3}[y^2 - \frac{y^3}{3}]_0^1 = \frac{2}{3}\cdot\frac{2}{3} = \frac{4}{9}$

7 (i) $M_X(t) = E(e^{tX}) = (1-4t)^{-3}$ from yellow book

Let $Y = \frac{1}{2}X$.

$\therefore M_Y(t) = E(e^{tY}) = E(e^{tX/2}) = M_X(t/2) = (1-2t)^{-6/2}$

which is the m.g.f. of a gamma(3,1/2) or χ_6^2 variable

(ii) $P(X > 20) = P(Y > 10)$

= 1 – 0.8753 = 0.1247

8 Let L be the length of the metal bar and Z_i be the error that arises at the i^{th} cut.

Length of 1st post cut = $L + Z_1$

Length of 2nd post cut = $L + Z_1 + Z_2$

Length of 100th post cut = $L + Z_1 + Z_2 + \ldots + Z_{100}$

Error in length of last post cut is $E = Z_1 + Z_2 + \ldots + Z_{100}$

$E \sim N(0,900)$ approximately, by CLT

$P(|E| < 15) \approx P(|Z| < 15/30) = P(|Z| < 0.5) = 2 \times 0.1915 = 0.383$

So P(error exceeds 15mm) ≈ 1 – 0.383 = 0.617

9 (i) Probability of type I error is

$\alpha = P(\text{reject } H_0 | H_0 \text{ is true}) = P(X = 0 \text{ or } X = 4 | H_0 \text{ is true})$,

which gives

$$\alpha = P\{X = 0 | P(Heads) = 0.5)\} + P\{X = 4 | P(Heads) = 0.5)\}$$
$$= \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^4 = 0.125.$$

(ii) Probability of type II error of the test at $P(Heads) = 0.7$ is

$$\beta = P(\text{accept } H_0 \,|\, H_1 \text{ is true}) = 1 - P\{X = 0 \text{ or } X = 4 \,|\, P(Heads) = 0.7\}$$

$$\Rightarrow \beta = 1 - \left(0.3^4 + 0.7^4\right) = 0.7518.$$

[OR using $P\{X = 1 \text{ or } X = 2 \text{ or } X = 3 \,|\, P(heads) = 0.7\}$]

10

range	0–1	1–2	2–3
observed frequency	45	35	20
expected frequency	100/3	100/3	100/3

$\chi^2 = [(45 - 100/3)^2 + (35 - 100/3)^2 + (20 - 100/3)^2]/(100/3) = 9.50$ on 2df

P-value = $P(\chi_2^2 > 9.50) < 0.01$

Reject model (at the 1% level of testing) as not providing a good fit to the data.

OR 5% point of χ^2_2 is 5.991, so we reject model at 5%
OR 1% point of χ^2_2 is 9.210, so we reject model at 1%

11 (i) (a) The required probability is

$$P(T > t_k) = 1 - P(T \le t_k) = 1 - F_T(t_k)$$

$$= 1 - (1 - e^{-\lambda t_k}) = e^{-\lambda t_k} \quad \text{(using formulae or by integration).}$$

(b) The likelihood function is given by:

$$L(\lambda) = \prod_{i=1}^{k} f(t_i) \prod_{j=k+1}^{n} P(T > t_k)$$

$$= \prod_{i=1}^{k}\left(\lambda e^{-\lambda t_i}\right) \prod_{j=k+1}^{n}\left(e^{-\lambda t_k}\right) = \lambda^k e^{-\lambda \sum_{i=1}^{k} t_i} e^{-(n-k)\lambda t_k}$$

For the MLE:

$$l(\lambda) = \log L(\lambda) = k\log(\lambda) - \lambda\sum_{i=1}^{k} t_i - (n-k)\lambda t_k$$

$$l'(\lambda) = \frac{k}{\lambda} - \sum_{i=1}^{k} t_i - (n-k)t_k$$

$$l'(\lambda) = 0 \Rightarrow \hat{\lambda} = \frac{k}{\sum_{i=1}^{k} t_i + (n-k)t_k}.$$

[And $l''(\lambda) = -\frac{k}{\lambda^2} < 0$]

(c) For the observed data,

$$n = 20,\ k = 5,\ t_k = 21.54,\ \sum_{i=1}^{k} t_i = 54.82.$$

$$\hat{\lambda} = \frac{k}{\sum_{i=1}^{k} t_i + (n-k)t_k} = \frac{5}{54.82 + 15\times 21.54} = 0.0132\ .$$

(ii) (a) We have n policies with independent durations, and each will have expired by the time of termination with probability

$$p = P(T \le t_0) = 1 - e^{-\lambda t_0},$$

or will have not expired with probability 1 - p.

Therefore, $K \sim \text{bin}\left(n,\ 1 - e^{-\lambda t_0}\right)$

(b) $$L(\lambda) \propto \left(1 - e^{-\lambda t_0}\right)^k \left(e^{-\lambda t_0}\right)^{n-k}$$

$$l(\lambda) = \log L(\lambda) = k\log\left(1 - e^{-\lambda t_0}\right) - (n-k)\lambda t_0$$

$$l'(\lambda) = \frac{kt_0 e^{-\lambda t_0}}{1 - e^{-\lambda t_0}} - (n-k)t_0$$

$$l'(\lambda) = 0 \Rightarrow e^{-\lambda t_0} = \frac{n-k}{n} \Rightarrow \hat{\lambda} = -\frac{1}{t_0}\log\left(1 - \frac{k}{n}\right)$$

[OR, observed proportion (k/n) is the MLE of corresponding proportion/probability $\{1 - \exp(-\lambda t_0)\}$; solving for λ leads to same estimate as above.]

(c) Now $t_0 = 24$ and all other involved quantities are as before.

$$\hat{\lambda} = -\frac{1}{t_0}\log\left(1-\frac{k}{n}\right) = -\frac{1}{24}\log\left(1-\frac{5}{20}\right) = 0.0120 .$$

12 (i) (a) $F = 1149/289 = 3.98$ on (2, 12) degrees of freedom

From Yellow Tables pages 172/3, *P-value* of the data is between 0.05 and 0.025.

We can reject H_0 (the "no schools effects" hypothesis) at the 5% level of testing but not at the 1% level. We have some evidence against the "no schools effects" hypothesis – and conclude that there are school effects (i.e. differences among the underlying means).

(b) School 1 mean = 598/5 = 119.6

$t_{12}(0.025) = 2.179$

95% CI for school 1 mean is $119.6 \pm 2.179\times(289/5)^{1/2}$

i.e. 119.6 ± 16.6 or (103.0, 136.2)

(ii) (a) $y_{1\bullet} = 598$, $y_{2\bullet} = 485$, $y_{3\bullet} = 629$, $y_{4\bullet} = 566$

$y_{\bullet\bullet} = 2278$, $\Sigma y^2 = 266{,}788$

$SS_T = 266788 - 2278^2/20 = 7323.8$

$SS_B = (598^2 + 485^2 + 629^2 + 566^2)/5 - 2278^2/20 = 2301$

$\Rightarrow SS_R = 7324 - 2301 = 5023$

Source of variation	*d.f.*	*SS*	*MSS*
Between schools	3	2301	767
Residual	16	5023	314
Total	19	7324	

$F = 767/314 = 2.44$ on (3, 16) degrees of freedom

From Yellow Tables pages 172/3, *P-value* of the data is just more than 0.1 (>10%)

We do not have sufficiently strong evidence against the "no schools effects" hypothesis, which can stand.

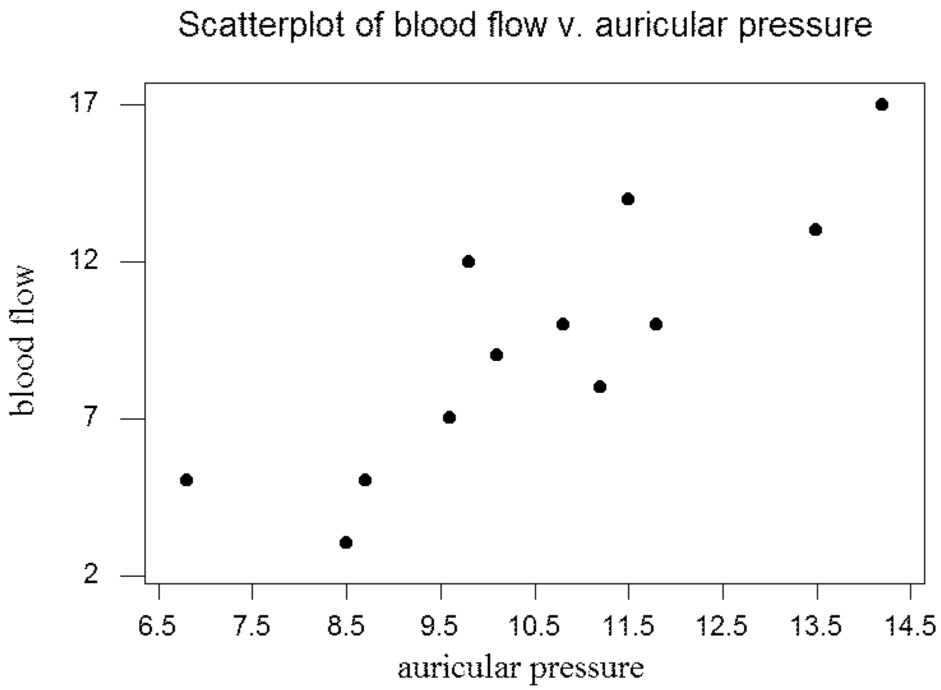
(b) With only three schools involved, the results from one of them (School 2) are sufficiently different from those of the other two to allow us to detect a difference among underlying means. However, the results for the fourth school range across the results for the original three schools – with all four schools in the comparison, the "between schools" sum of squares is no longer so high relative to the residual and we fail to detect differences.

(c) $t_{16}(0.025) = 2.120$

95% CI is $\left(119.6-97\right)\pm 2.120\left\{314\left(\frac{1}{5}+\frac{1}{5}\right)\right\}^{0.5}$

i.e. 22.6 ± 23.76 or $(-1.2, 46.4)$

13 (i) A clearly labelled scatterplot:



There seems to be a positive linear relationship between blood flow and auricular pressure.

(ii) $n = 12$

$$S_{xx} = 1381.85 - \frac{126.5^2}{12} = 48.3292$$

$$S_{yy} = 1251 - \frac{113^2}{12} = 186.9167$$

$$S_{xy} = 1272.2 - \frac{(126.5)(113)}{12} = 80.9917$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{80.9917}{48.3292} = 1.676$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{1}{12}(113 - 1.6758 * 126.5) = -8.249$$

Fitted line is $y = -8.249 + 1.676x$

(iii) (a) $$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2} \quad \text{where} \quad \hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$

$$P\left[-t_{n-2}(2.5\%) < \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} < t_{n-2}(2.5\%)\right] = 0.95$$

Rearrangement results in the 95% confidence interval for β

$$\hat{\beta} \pm t_{n-2}(2.5\%)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

Here: $\hat{\sigma}^2 = \frac{1}{10}\left(186.9167 - \frac{(80.9917)^2}{48.3292}\right) = 5.1188$

95% CI is $1.676 \pm 2.228(0.3254) \Rightarrow 1.676 \pm 0.725 \Rightarrow (0.95, 2.40)$

(b) As 1.5 lies comfortably inside this confidence interval, then there is no evidence at all against the hypothesis that $\beta = 1.5$.

(iv) (a) $$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

$$P[\chi^2_{n-2}(97.5\%) < \frac{(n-2)\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-2}(2.5\%)] = 0.95$$

Rearrangement results in the 95% confidence interval for σ^2

$$\frac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2}(2.5\%)} < \sigma^2 < \frac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2}(97.5\%)}$$

Here 95% CI is $\frac{10(5.1188)}{20.48} < \sigma^2 < \frac{10(5.1188)}{3.247}$

$\Rightarrow (2.50, 15.76)$

(b) 95% CI for σ is $(\sqrt{2.50}, \sqrt{15.76}) \Rightarrow (1.58, 3.97)$

END OF EXAMINERS' REPORT