Faculty of Actuaries **Institute of Actuaries**

EXAMINATION

September 2005

Subject CT3 — Probability and Mathematical Statistics Core Technical

EXAMINERS' REPORT

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 Flaherty
Chairman of the Board of Examiners

15 November 2005

1 (a) $\bar{x} = \dfrac{150}{100} = 1.5$

(b) Median is $(101/2)^{th}$ observation i.e. the mean of the 50^{th} and 51^{st} observations so median = 1

(c) Mean is higher than the median as the data are positively skewed (skewed to the right).

2 Let X denote the number in the sample who support party A.

X ~ Binomial(275, 0.45)
$\therefore$ $E[X]$ = 275 × 0.45 = 123.75
$V[X]$ = 275 × 0.45 × 0.55 = 68.0625

The normal approximation to the binomial gives, using a continuity correction,

$$P(X \geq 116) = P(X > 115.5) \approx 1 - \Phi\left(\frac{115.5 - 123.75}{\sqrt{68.0625}}\right) = 1 - \Phi(-1) = \Phi(1) = 0.841$$

3 Gamma(120, 1.2) has mean $\dfrac{120}{1.2} = 100$ and variance $\dfrac{120}{1.2^2} = 83.333$

$\therefore X \approx N(100, 9.129^2)$ by the Central Limit theorem (since the gamma variable is the sum of 120 independent gamma(1,1.2) variables)

$$P\left(X > 120\right) \cong P\left(Z > \frac{120 - 100}{9.129} = 2.191\right) = 1 - 0.98578 = 0.0142$$

4 Sample proportion = 0.83

99% CI for the population proportion is 0.83 ± [2.5758 × $(0.83 \times 0.17/100)^{1/2}$]
i.e. 0.83 ± 0.0968 i.e. (0.733, 0.927)

99% CI for percentage is thus (73.3% , 92.7%)

5 n = 500 is very large, so the Central Limit Theorem justifies normality.

95% CI is $\bar{x} \pm 1.96 \dfrac{s}{\sqrt{n}}$

$$\Rightarrow \quad 237 \pm 1.96 \frac{137}{\sqrt{500}} \quad \Rightarrow \quad 237 \pm 12.0 \quad \text{or} \quad (£225.0, £249.0)$$

6

Source of variation	*d.f.*	*SS*	*MSS*
Between groups	3	840	280
Residual	60	660	11
Total	63	1500	

F = 280/11 which equals 25.45.

Therefore there is overwhelming evidence to suggest that the four population means are not equal ($F_{3,60}$(5%) = 2.758, $F_{3,60}$(1%) = 4.126).

7 (a) $\bar{x} = \dfrac{1}{20}(3256) = 162.8$

$$s^2 = \frac{1}{19}\left\{866600 - \frac{3256^2}{20}\right\} = 17711.7 \quad \therefore s = 133.1$$

(b) Distribution must have strong positive skewness since the s.d. is large relative to the mean and the amounts must be positive.

8 (i) Using C for a claim, N for no claim, then

P(premium = 400) = P(C in year 3, regardless of the first 2 years) = p

P(premium = 400k) = P(CN in years 2/3, regardless of the first year) = $p(1-p)$

P(premium = 400k^2) = P(NN in years 2/3, regardless of the first year) = $(1-p)^2$

[These probabilities may be derived in other ways, such as via a tree diagram]

(ii) $E(\text{premium}) = 400.p + 400k.p(1-p) + 400k^2.(1-p)^2$
$= 400\{p + kp(1-p) + k^2(1-p)^2\}$

(iii) For $E(\text{premium}) = 300$ when $p = 0.1$
then $0.1 + 0.09k + 0.81k^2 = 0.75$ $\quad\therefore\ 0.81k^2 + 0.09k - 0.65 = 0$

$$\therefore k = \frac{-0.09 \pm \sqrt{0.09^2 + 4(0.81)(0.65)}}{1.62} = \frac{-0.09 + 1.454}{1.62} \quad \text{for} \quad 0 < k < 1$$

$$\therefore k = 0.84$$

9 Pooled estimate of common population variance = $\dfrac{10\times 59 \ + \ 14\times 42}{10+14} = 49.0833$

$t_{24}(0.025) = 2.064$

95% CI for $\mu_1 - \mu_2$ is given by

$$(124-105) \ \pm \ 2.064\times\left(49.0833\left(\frac{1}{11}+\frac{1}{15}\right)\right)^{1/2} \ \text{ i.e. } 19\pm 5.74 \ \text{ i.e. } (13.3\ ,\ 24.7)$$

10 $X \sim U(0,1000)$, $Y = \min(X,800)$

(i) $$P(X < x \mid X < 800) = \frac{P(X < x \ \text{ and } \ X < 800)}{P(X < 800)}$$

$$= \frac{P(X < x)}{800/1000} \quad \text{for } 0 < x < 800$$

$$= \frac{x/1000}{800/1000} = \frac{x}{800} \quad \text{for} \quad 0 < x < 800$$

so the conditional distribution is $U(0,800)$

[other reasonable arguments were given credit, e.g. "the conditional distribution is simply a scaled version of the original uniform distribution on a restricted range".]

(ii) $E[Y] = E[X|X < 800]\ P(X < 800) + 800\ P(X \geq 800)$

$$= 400\left(\frac{800}{1000}\right) + 800\left(\frac{200}{1000}\right)$$

$= 480$

(iii) $\overline{Y}$ is approximately normal with expectation 480 by Central Limit Theorem

(iv) $E[X \mid X < 800] = 400$ whereas $E[Y] = 480$.

The higher value for $E[Y]$ results from 20% of the Y values being 800 (and 80% being between 0 and 800).

11 (a) For $0 \le r < 0.55 \quad \Rightarrow n = 0$
$0.55 \le r < 0.8 \quad \Rightarrow n = 1$
$0.8 \le r < 0.95 \quad \Rightarrow n = 2$
$0.95 \le r \le 1 \quad \Rightarrow n = 3$

[OR any equivalent allocation which reflects the probabilities of the 4 values of N.]

(b) $0.6221 \Rightarrow n = 1$
$0.1472 \Rightarrow n = 0$
$0.9862 \Rightarrow n = 3$

12 (i) (a) The x_i's are known to be such that $x_i \ge 5$, therefore have density which is a scaled form of $\lambda e^{-\lambda x}$ for $5 < x < \infty$.

The scaling constant k is such that $\int_5^\infty k.\lambda e^{-\lambda x} dx = 1$

$\therefore k[-e^{-\lambda x}]_5^\infty = 1 \quad \therefore k.e^{-5\lambda} = 1 \quad \therefore k = e^{5\lambda}$

[*Note*: this can be argued in other ways; e.g. by referring to a conditional density and dividing by $P(X > 5)$]

(b) $L(\lambda) = \prod_{i=1}^{n} \lambda e^{5\lambda} e^{-\lambda x_i} = \lambda^n e^{5n\lambda} e^{-\lambda \Sigma x_i}$

$\log L(\lambda) = n \log \lambda + 5n\lambda - \lambda \Sigma x_i$

$\frac{d}{d\lambda} \log L(\lambda) = \frac{n}{\lambda} + 5n - \Sigma x_i$

equate to zero for MLE $\Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i - 5n}$

[OR It could be noted that $X - 5 \sim \exp(\lambda)$ and that the MLE is therefore the reciprocal of the mean of the data $(x_i - 5)$ giving the required answer]

(c) $n = 10$, $\Sigma x_i = 71 \Rightarrow \hat{\lambda} = \frac{10}{71 - 50} = 0.476$

(ii) (a) $L(\lambda) = (1 - e^{-5\lambda})^m (e^{-5\lambda})^n$

$\log L(\lambda) = m \log(1 - e^{-5\lambda}) + n \log(e^{-5\lambda})$

$\frac{d}{d\lambda} \log L(\lambda) = \frac{5me^{-5\lambda}}{1 - e^{-5\lambda}} - 5n$

equate to zero for MLE $\Rightarrow \frac{e^{-5\lambda}}{1 - e^{-5\lambda}} = \frac{n}{m}$

$\therefore e^{-5\lambda} = \frac{n}{m+n} \quad \therefore \hat{\lambda} = \frac{1}{5} \log(\frac{m+n}{n})$

[OR Reason via the MLE for a binomial $p = P(X > 5)$ such that $\hat{p} = \frac{n}{m+n}$ and $p = e^{-5\lambda}$]

(b) $m = 120$, $n = 10 \Rightarrow \hat{\lambda} = 0.513$

(iii) (a) $(1-e^{-5\lambda})^m$ is the likelihood of observing m policies with duration < 5

$\prod_{i=1}^{n} \lambda e^{-\lambda x_i}$ is the likelihood of observing the actual durations $x_1, \ldots, x_n$

and independence leads to the product of these

$$\log L(\lambda) = m \log(1-e^{-5\lambda}) + n \log \lambda - \lambda \Sigma x_i$$

$$\frac{d}{d\lambda} \log L(\lambda) = \frac{5me^{-5\lambda}}{1-e^{-5\lambda}} + \frac{n}{\lambda} - \Sigma x_i$$

equate to zero and the solution gives the MLE.

(b) All three are re-assuringly close.

The pooled estimate is between the first two (as expected, but it is closer to 0.513).

13

	purchase	*no purchase*	
return	69	68	137
no return	32	51	83
	101	119	220

(i) (a) 69/101 (= 0.6832)
(b) 68/119 (= 0.5714)
(c) 69/137 (= 0.5036)

(ii) H_0: population proportions of those who intend to return are equal
v H_1: not H_0

Proportion of purchasers $\hat{\theta}_1 = 69/101$; proportion of non-purchasers $\hat{\theta}_2 = 68/119$

Under H_0, estimate of common proportion who intend to return = 137/220

Observed value of $D = \hat{\theta}_1 - \hat{\theta}_2 = 0.1117$

Estimated standard error of $D = \left[\frac{137}{220} \times \frac{83}{220}\left(\frac{1}{101} + \frac{1}{119}\right)\right]^{1/2} = 0.06558$

P-value = $2 \times P(D > 0.1117) = 2 \times P(Z > 0.1117/0.06558) = 2 \times P(Z > 1.70)$
$= 2 \times (1 - 0.95543) = 0.08914$ (i.e. 8.9%)

There is not sufficient evidence (using a two-sided test) to justify rejecting H_0 i.e. there is not sufficient evidence to justify concluding that the intention to return depends on whether or not a purchase was made.

(iii) H_0: no association between attributes v H_1: not H_0

Expected frequencies under H_0 in brackets:

	purchase	*no purchase*	
return	69 (62.9)	68 (74.1)	137
no return	32 (38.1)	51 (44.9)	83
	101	119	220

Test statistic = $6.1^2\left(\frac{1}{62.9}+\frac{1}{74.1}+\frac{1}{38.1}+\frac{1}{44.9}\right) = 2.90$

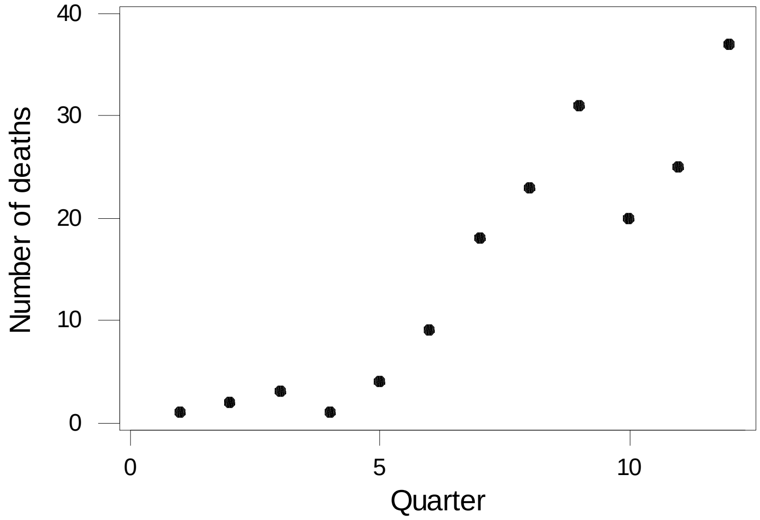
P-value = $P\left(\chi_1^2 > 2.90\right) = 1 - 0.9114 = 0.0886$ (i.e. 8.9%)

There is not sufficient evidence to justify rejecting H_0 i.e. there is not sufficient evidence to justify concluding that the intention to return is associated with purchasing status.

(iv) The two approaches complement each other:
the P-values are the same
the conclusions are the same.

[*Note:* there is a formal connection: the χ_1^2 value in (iii) (2.90) is the square of the z value in (ii) (1.70).]

14 (i) (a) Points are shown on scatterplot.



(b) The mean number of deaths increases with an increasing rate with quarter.

The variance also appears to increase with quarter.

(ii) (a) $q = \Sigma(n_i - \gamma i^2)^2$

$$\frac{dq}{d\gamma} = -2\Sigma i^2(n_i - \gamma i^2)$$

$$\frac{dq}{d\gamma} = 0 \Rightarrow -2\Sigma i^2(n_i - \gamma i^2) = 0$$

$$\therefore\ \Sigma i^2 n_i - \gamma \Sigma i^4 = 0$$

$$\therefore\ \hat{\gamma} = \frac{\Sigma i^2 n_i}{\Sigma i^4}.$$

$$\left(\frac{d^2q}{d\gamma^2} = 2\Sigma i^4 > 0 \quad \therefore \text{ minimum.}\right)$$

(b) $$q^* = \sum \frac{(n_i - \gamma i^2)^2}{i^2} = \sum \left(\frac{n_i}{i} - \gamma i \right)^2$$

$$\frac{dq^*}{d\gamma} = -2\sum i \left(\frac{n_i}{i} - \gamma i \right)$$

$$\frac{dq^*}{d\gamma} = 0 \Rightarrow \sum n_i - \gamma \sum i^2 = 0$$

$$\tilde{\gamma} = \frac{\sum n_i}{\sum i^2}$$

$$\left(\frac{d^2 q^*}{d\gamma^2} = 2\sum i^2 > 0 \quad \therefore \text{ minimum.} \right)$$

(c) $$\hat{\gamma} = \frac{\sum i^2 n_i}{\sum i^4} = \frac{15694}{60710} = 0.259$$

$$\tilde{\gamma} = \frac{\sum n_i}{\sum i^2} = \frac{174}{650} = 0.268$$

(iii) (a) $$E[N_i] = \gamma i^\theta$$

Taking logs gives

$$\log E[N_i] = \log(\gamma i^\theta) = \log\gamma + \theta \log(i) = \log\gamma + \theta x_i$$

Thus $\alpha = \log \gamma$ and $\beta = \theta$.

[OR $\gamma = e^\alpha$ and $\theta = \beta$.]

(b) $\hat{\beta} = 1.6008$ s.e.$(\hat{\beta}) = 0.2525$

H_0: $\beta = 2$ v H_1: $\beta \neq 2$

$$t = \frac{\hat{\beta} - 2}{\text{s.e.}(\hat{\beta})} = \frac{1.6008 - 2}{0.2525} = -1.58$$

Compare with a *t*-distribution with 10 d.f.

As the 5% critical value of a two-tailed test is 2.228, do not reject the null hypothesis.

Therefore, the model used in (ii) with $\theta = \beta = 2$ seems appropriate.

END OF EXAMINERS' REPORT