EXAMINATIONS

September 1999

Subject C — Statistics

Paper One

EXAMINERS' REPORT

Examiners' Comments

The performance in the three parts differed with Part One (multiple choice) clearly producing the best answers and Part Three (long questions) producing the poorest answers.

Of the short answer questions, Q10 and Q12 had good responses but Q8 and Q11 were badly attempted.

Of the long answer questions, Q17 had clearly the best responses and Q16 the worst responses.

PART ONE

1 $Q1 = (40 + 2) / 4\text{th} = 10.5\text{th observation} = (1.27 + 1.63) / 2 = 1.45$

$Q3 = 30.5\text{th observation } (= 10.5\text{th from top end}) = (25.18 + 28.82) / 2 = 27.0$

$IQR = 27.0 - 1.45 = 25.55.$ So **D**.

2 $P(\text{fifth selected is a man})$

$= P(\text{first selected is a man}) = \frac{12}{20} = 0.6.$ So **C**.

3 $E(Y|X = 0) = [0(0.01) + 1(0.10) + 2(0.05)] / 0.16 = 1.25.$ So **B**.

4 Since α is large, Central Limit Theorem allows use of normal approximation.

Mean = 40 / 0.1 = 400, var = $40 / (0.1)^2 = 4000$ and sd = 63.25.

$P(X > 500) = P(Z > (500 - 400) / 63.25) = P(Z > 1.58) = 0.057.$ So **A**.

5 Use $\frac{(n-1)S^2}{\sigma^2} = \chi^2_{n-1}$. $\therefore$ $P\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_{0.95,n-1}\right) = 0.95$ using tables notation.

Here $P\left(\frac{9S^2}{15^2} > 3.325\right) = 0.95$ So required "a" $= \sqrt{\frac{3.325 \times 15^2}{9}} = 9.12.$

So **B**.

6 $130.2 \pm 1.833 \times \frac{25.0}{\sqrt{10}}$, i.e. 115.71 to 144.69. So **C**.

7 All observed frequencies O_i have doubled and so all expected frequencies E_i also double. Hence each term $(O_i - E_i)^2 / E_i$ in the χ^2 statistic also doubles.

So **C**.

PART TWO

8 SSREG / SSTOT = $R^2 = r^2 = 0.36$, so SSRES / SSTOT = 0.64,
so SSREG : SSRES = 0.36 : 0.64, so SSREG = (0.36 / 0.64) × 2160 = 1215.

9 John: $\Sigma x = 15(452) = 6780$
$\Sigma x^2 = 14(205)^2 + 15(452)^2 = 3652910$

Mary: $\Sigma x = 20(385) = 7700$
$\Sigma x^2 = 19(182)^2 + 20(385)^2 = 3593856$

Combined: $\Sigma x = 14480$
$\Sigma x^2 = 7246766$

$\therefore$ mean $= \frac{14480}{35} = 413.7$

s.d. $= \sqrt{\frac{1}{34}\left\{7246766 - \frac{14480^2}{35}\right\}} = 192.2$

10 $E(X)$ = midpoint of [2, 4] = 3

$$\text{Var}(X) = \frac{1}{2}\int_2^4 (x-3)^2\,dx = \left[\frac{(x-3)^3}{6}\right]_2^4 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$E(Y) = 48E(X) = 144$

$\text{Var}(Y) = 48\text{Var}(X) = 16$. So s.d.$(Y) = 4$

So $P(140 \le Y \le 152) = P(-4/4 \le Z \le 8/4)$
$= 0.8413 + 0.9772 - 1 = 0.82$

11 Assuming that the resulting "n" is large enough to use normality, the confidence limits are

$$\pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ .$$

We must allow for the worst case of $\hat{p} = 0.5$.

So we seek n such that

$$1.96\sqrt{\frac{(0.5)(0.5)}{n}} = 0.05 \qquad \text{(for} \pm 5\%\text{)}$$

$$\Rightarrow \sqrt{n} = \frac{1.96(0.5)}{0.05} = 19.6 \qquad \therefore\ n = 385$$

[This is large enough to validate the original assumption that "n" is large enough to use normality.]

12 Observed and expected frequencies

	I		***II***		***III***		
	o	*e*	*o*	*e*	*o*	*e*	*Total*
A	30	35.85	35	35.21	40	33.93	105
B	26	20.15	20	19.79	13	19.07	59
Total	56		55		53		164

H_1 : Representative and Policy Type are independent

H_0 : not independent

Contributions to χ^2

0.95	0.00	1.09
1.70	0.00	1.93

Total = 5.67

5% point on 2 df = 5.99

So just insufficient evidence to reject independence at 5% level.

Perhaps collect more data.

13 (i) Least squares estimate of β minimises

$$q = \sum_{i=1}^{n} [y_i - \beta x_i]^2$$

$$\frac{\partial q}{\partial \beta} = \sum_{i=1}^{n} 2[y_i - \beta x_i]\,(-x_i)$$

equate to zero:

$$\sum_{i=1}^{n} y_i\, x_i = \beta \sum_{i=1}^{n} x_i^2$$

$$\therefore\ \hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

(ii) For the fitted line to pass through $(\bar{x}, \bar{y})$ $\hat{\beta}$ must satisfy the equation

$$\bar{y} = \hat{\beta}\bar{x} \quad \text{or} \quad \Sigma y_i = \hat{\beta}\,.\,\Sigma x_i$$

So, is it true that

$$\Sigma y_i = \frac{\Sigma x_i y_i}{\Sigma x_i^2}\,.\,\Sigma x_i\ ?$$

Obviously **not** in general.

14 $W = (1/2)\log[(1 + r) / (1 - r)] \sim N(\mu, 1/47)$ where $\mu = (1/2)\log[(1 + \rho) / (1 - \rho)]$

$P[-1.96 < (W - \mu)\sqrt{47} < 1.96] = 0.95$

i.e. $P(W - 0.285895 < \mu < W + 0.285895) = 0.95$

Here $w = (1/2)\log(1.58 / 0.42) = 0.662463$

so 95% CI for μ is (0.376568, 0.948358)

Now $\mu = (1/2)\log[(1 + \rho) / (1 - \rho)]$ gives $\rho = [\exp(2\mu) - 1] / [\exp(2\mu) + 1]$

Then the limits for μ give the 95% CI for ρ as (0.360, 0.739).

PART THREE

15 (i)

$$M_X(t) = \int_0^\infty e^{tx} \, . \, \lambda e^{-\lambda x} \, dx$$

$$= \frac{\lambda}{\lambda - t} \int_0^\infty (\lambda - t) \, e^{-(\lambda - t)x} \, dx = \frac{\lambda}{\lambda - t} \quad (t < \lambda)$$

(ii) (a)

$$M_Z(t) = \int_{-\infty}^\infty e^{tz} f(z) \, dz$$

$$= \int_{-\infty}^0 e^{tz} \frac{\alpha\beta}{\alpha + \beta} e^{\beta z} \, dz + \int_0^\infty e^{tz} \frac{\alpha\beta}{\alpha + \beta} e^{-\alpha z} \, dz$$

$$= \frac{\alpha\beta}{\alpha + \beta} \cdot \left[\frac{e^{(\beta + t)z}}{\beta + t} \right]_{-\infty}^0 + \frac{\beta}{\alpha + \beta} \cdot \frac{\alpha}{\alpha - t}$$

$$= \frac{\alpha\beta}{\alpha + \beta} \left\{ \frac{1}{\beta + t} + \frac{1}{\alpha - t} \right\}$$

$$= \frac{\alpha\beta}{(\beta + t)(\alpha - t)} \quad (-\beta < t < \alpha \text{ for convergence})$$

(b) $\log M_Z(t) = \log(\alpha\beta) - \log(\beta + t) - \log(\alpha - t)$

$$\therefore \frac{1}{M_Z(t)} M'_Z(t) = -\frac{1}{\beta + t} + \frac{1}{\alpha - t}$$

$$\therefore E(Z) = M'_Z(0) = \frac{1}{\alpha} - \frac{1}{\beta} \, .$$

$$\frac{M''_Z(t)}{M_Z(t)} - \left(\frac{M'_Z(t)}{M_Z(t)} \right)^2 = \frac{1}{(\beta + t)^2} + \frac{1}{(\alpha - t)^2}$$

$$\therefore \text{Var}(Z) = M''_Z(0) - \left[M'_Z(0) \right]^2 = \frac{1}{\alpha^2} + \frac{1}{\beta^2} \, .$$

(c) Mode is when $z = 0$.

If $\alpha > \beta$, mean is negative.

Mean < mode $\Rightarrow$ negative skewness

If $\alpha < \beta$, mean is positive.

Mean > mode $\Rightarrow$ positive skewness.

(iii) $M_R(t) = E(e^{tR}) = E[e^{t(X-Y)}]$

$= E(e^{tX}) \,.\, E(e^{-tY}) = M_X(t) \,.\, M_Y(-t)$

$= \dfrac{\alpha}{\alpha - t} \cdot \dfrac{\beta}{\beta + t}$

$\therefore$ R has same density as Z in part (ii).

16 (i) $L(\mu, \sigma^2) = \prod_{i=1}^{n} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{1}{2}\dfrac{(x_i - \mu)^2}{\sigma^2}\right\}$

$\therefore \log L(\mu, \sigma^2) = \text{const.} - \dfrac{n}{2} \log(\sigma^2) - \dfrac{1}{2\sigma^2} \Sigma(x_i - \mu)^2$

$\dfrac{\partial}{\partial\mu} \log L = \dfrac{1}{\sigma^2}\Sigma(x_i - \mu)$

equate to zero $\Rightarrow \hat{\mu} = \bar{x}$

$\dfrac{\partial}{\partial\sigma^2} \log L = -\dfrac{n}{2} \cdot \dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4} \Sigma(x_i - \mu)^2$

equate to zero, substituting $\hat{\mu} = \bar{x}$

$\Rightarrow \hat{\sigma}^2 = \dfrac{1}{n}\Sigma(x_i - \bar{x})^2$ as required.

(ii) Given result $\Rightarrow$ $\left.\begin{aligned} E\{\Sigma(X_i - \bar{X})^2\} &= (n-1)\sigma^2 \\ \text{Var}\{\Sigma(X_i - \bar{X})^2\} &= 2(n-1)\,\sigma^4 \end{aligned}\right\}$

$\therefore \; E(\hat{\sigma}^2) = \dfrac{1}{n}(n-1)\,\sigma^2 \neq \sigma^2 \quad \therefore$ biased

$\text{bias}(\hat{\sigma}^2) = \dfrac{n-1}{n}\,\sigma^2 - \sigma^2 = -\dfrac{1}{n}\,\sigma^2$

$\text{Var}(\hat{\sigma}^2) = \dfrac{1}{n^2} \,.\, 2(n-1)\,\sigma^4 = \dfrac{2(n-1)}{n^2}\,\sigma^4$

(iii) $E(S^2) \quad = \dfrac{1}{n-1}(n-1)\,\sigma^2 = \sigma^2 \qquad \therefore$ unbiased

$\text{Var}(S^2) \quad = \dfrac{1}{(n-1)^2}\,.\,2(n-1)\,\sigma^4 = \dfrac{2}{n-1}\sigma^4$

(iv) $\text{MSE}(\hat{\sigma}^2) = \dfrac{2(n-1)}{n^2}\,\sigma^4 + \left(-\dfrac{1}{n}\sigma^2\right)^2 = \dfrac{(2n-1)}{n^2}\,\sigma^4$

$\text{MSE}(S^2) = \dfrac{2}{n-1}\sigma^4$

$$\frac{\text{MSE}(\hat{\sigma}^2)}{\text{MSE}(S^2)} = \frac{(2n-1)}{n^2}\cdot\frac{n-1}{2} = \frac{2n-1}{2n}\cdot\frac{n-1}{n} < 1$$

$\therefore\ \text{MSE}(\hat{\sigma}^2) < \text{MSE}(S^2)$ for all n (≥ 2)

(v) Label it as $\hat{\sigma}^2_+$.

$$\frac{\text{MSE}(\hat{\sigma}^2_+)}{\text{MSE}(\hat{\sigma}^2)} = \frac{2}{n+1}\cdot\frac{n^2}{2n-1} = \frac{2n^2}{2n^2+n-1} < 1 \ \text{ for } n \geq 2$$

$\therefore\ \hat{\sigma}^2_+$ is better than $\hat{\sigma}^2$ (and S^2)

17 (i) (a) Plot:

```
      .         .     . . :.     .   .          .
-----+---------+---------+---------+---------+---------  A
.
      . .        . . . . .     .
-----+---------+---------+---------+---------+---------  B
.   10        20        30        40        50        60
```

Normally assumption seems reasonable in both cases.

(b) $s^2_A = 143.14,\ s^2_B = 74.48$

$H_0 : \sigma^2_A = \sigma^2_B \quad \text{v} \quad H_1 : \sigma^2_A \neq \sigma^2_B$

$F = 143.14 / 74.48 = 1.9$ on 9, 7 d.f.

Upper 5% point is 3.68, so P-value greatly exceeds 10%.

H_0 can stand: analyst's assumption is justified.

(ii) (a) A: $\Sigma(x-\bar{x})^2 = 1288.256$; B: $\Sigma(y-\bar{y})^2 = 521.359$

$\therefore$ Pooled $s^2 = (1288.256 + 521.359) / 16 = 113.10$

$\bar{x}_A - \bar{x}_B = 32.38 - 24.463 = 7.917$

Upper 2½% point of $t_{16} = 2.12$

$\therefore$ 95% CI is $7.917 \pm 2.12 \times [113.10 \times (1/10 + 1/8)]^{1/2}$

i.e. 7.917 ± 10.694 i.e. $(-2.8, 18.6)$

CI contains zero, so we cannot conclude that the population mean claim amounts differ.

(i.e. we have not been able to detect a difference.)

(b) $H_0 : \mu_A = \mu_B$ v $H_1 : \mu_A > \mu_B$

$t = 7.917 / [113.10 \times (1/10 + 1/8)]^{1/2} = 1.57$ on 16 d.f.

P-value is approximately 0.07. H_0 can stand.

We do not have strong enough evidence to conclude that the mean claim amount for A exceeds that for B.