INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2015 examinations

Subject CT3 – Probability and Mathematical Statistics Core Technical

Introduction

The Examiners' Report is written by the Principal Examiner with the aim of helping candidates, both those who are sitting the examination for the first time and using past papers as a revision aid and also those who have previously failed the subject.

The Examiners are charged by Council with examining the published syllabus. The Examiners have access to the Core Reading, which is designed to interpret the syllabus, and will generally base questions around it but are not required to examine the content of Core Reading specifically or exclusively.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report; other valid approaches are given appropriate credit. For essay-style questions, particularly the open-ended questions in the later subjects, the report may contain more points than the Examiners will expect from a solution that scores full marks.

The report is written based on the legislative and regulatory context at the date the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

F Layton
Chairman of the Board of Examiners

June 2015

General comments on Subject CT3

Some of the questions in this paper admit alternative solutions from these presented in this report, or different ways in which the provided answer can be determined. All mathematically correct and valid alternative solutions or answers received credit as appropriate. Rounding errors were not penalised, unless excessive rounding led to significantly different answers. In cases where the same error was carried forward to later parts of the answer, candidates were only penalised once. In questions where comments were required, reasonable comments that were different from those provided in the solutions also received full credit where appropriate.

Comments on the April 2015 paper

Candidates performed generally well and the pass rate was in line with previous sessions. There was a number of excellent scripts achieving very high scores.

In general, questions that tested concepts that had not appeared in recent exams were not particularly well answered.

Candidates are also reminded that the Examiners test the syllabus which is supported by the Core Reading provided by the Institute and Faculty of Actuaries. For all topics in the syllabus candidates may be tested on the understanding and application of a given result, which might involve working through a proof that is not explicitly presented in the Core Reading.

The comments on individual questions that follow cover important frequent errors, and specific parts that were not answered well.

1 $\sum_i x_{i1} = n_1\bar{x}_1 = 3328$ and $\sum_i x_{i2} = n_2\bar{x}_2 = 1890$ giving $\sum_i x_i = 5218$

$$\bar{x} = \frac{5218}{106} = 49.23$$

$$\sum_i x_{i1}^2 = (n_1 - 1)s_1^2 + \left(\sum_i x_{i1}\right)^2 / n_1 = 178159$$

$$\sum_i x_{i2}^2 = (n_2 - 1)s_2^2 + \left(\sum_i x_{i2}\right)^2 / n_2 = 87674 \text{ giving} \sum_i x_i^2 = 265833$$

$$s^2 = \frac{\left(265833 - \frac{5218^2}{106}\right)}{105} = 85.4244 \text{ and } s = 9.243$$

Generally well answered, although some problems were encountered with the variance.

2 $s_x^2 = \sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2 / \left(n-1\right)$

and $x_i - \bar{x} = x_i - \bar{x} - A + A = (x_i - A) - \left(\bar{x} - A\right) = d_i - \bar{d}$

since $\bar{d} = \frac{\sum f_i d_i}{n} = \frac{\sum f_i (x_i - A)}{n} = \bar{x} - A$

This gives $s_x^2 = \sum_{i=1}^{k} f_i \left(d_i - \bar{d}\right)^2 / \left(n-1\right) = \left(\sum_{i=1}^{k} f_i d_i^2 - \left(\sum_{i=1}^{k} f_i d_i\right)^2 / n\right) / \left(n-1\right)$

Overall performance was poor. Many answers completely ignored the involved frequencies. This is an example of a question that is not examined frequently and candidates found challenging.

3 Let X be the size of an individual claim, and N be the number of claims.

(i) Expected total amount is $E\left[X\right]E\left[N\right] = 1{,}000 \times 400 = 400{,}000$

(ii) $\text{Var}(\text{total amount}) = E[N]V[X] + V[N]E[X]^2$

A lower bound for the variance is then obtained by assuming $V[X] = 0$, that is,

$$\text{STD}(\text{total amount}) = \sqrt{V[N]E[X]^2} = 20 \times 1000 = 20{,}000$$

The first part was answered very well. In part (ii) many candidates failed to recognise that the answer relies on the variance being equal to zero.

4 No. claims ~Bin(900,p)

$\hat{p} = 290/900 = 0.322$

$\hat{p} \sim N(p, p(1-p)/n)$ approximately

$$\text{C.I.} = \hat{p} \pm Z_{0.95} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.322 \pm 1.6449 * \sqrt{\frac{0.322(1-0.322)}{900}} = (0.296, 0.348)$$

Very well answered.

5 (i) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

(ii) $\chi^2_{0.025;29} = 45.72, \chi^2_{0.975;29} = 16.05$

variance C.I. =

$$\left(\frac{(n-1)S^2}{\chi^2_{0.025;29}}, \frac{(n-1)S^2}{\chi^2_{0.975;29}}\right) = \left(29 * \frac{7.5^2}{45.72}, 29 * \frac{7.5^2}{16.05}\right) = (35.679, 101.64)$$

95% C.I. for S is $\left(\sqrt{35.679}, \sqrt{101.64}\right)$=(5.97,10.08)

Generally well answered.

6 (i) Using MGFs,

$$M_{\bar{X}}(t) = M_{X_1}\left(\frac{t}{n}\right) M_{X_2}\left(\frac{t}{n}\right) \cdots M_{X_6}\left(\frac{t}{n}\right) \text{ since } X_i \text{ are independent}$$

$$= \left\{ M_{X_1}\left(\frac{t}{6}\right) \right\}^6 \text{ since } X_i \text{ are identically distributed}$$

$$= \left\{ \left(1 - \frac{t}{6}\right)^{-2} \right\}^6 = \left(1 - \frac{t}{6}\right)^{-12}$$

i.e. a Gamma(12, 6) distribution.

(ii) $E\left(\hat{\theta}_1\right) = E\left(\bar{X}\right) = E\left(X_i\right) = 2$. So bias = 0.

$E\left(\hat{\theta}_2\right) = \frac{9}{30} 3E\left(X_i\right) + \frac{1}{30} 3E\left(X_i\right) = E\left(X_i\right) = 2$. So, again, bias = 0.

(iii) Since bias = 0 for both, we have:

$$\text{MSE}\left(\hat{\theta}_1\right) = V\left(\bar{X}\right) = \frac{V\left(X_i\right)}{6} = \frac{2}{6} = 0.333$$

$$\text{MSE}\left(\hat{\theta}_2\right) = V\left(\hat{\theta}_2\right) = \left(\frac{9}{30}\right)^2 3V\left(X_i\right) + \left(\frac{1}{30}\right)^2 3V\left(X_i\right) \quad \text{(independence)}$$

$$\frac{(81 \times 3) + 3}{900} \times 2 = 0.547$$

(iv) $\hat{\theta}_1$ has smaller MSE, and therefore is more efficient than $\hat{\theta}_2$.

Answers here were generally good. In part (ii) a number of candidates were confused with the estimators and the estimated parameter, but in part (iv) most candidates were familiar with the concept of efficiency.

7 (i) $f(x) = F'(x) = \frac{3}{8}x^2$ for $0 \le x \le 2$

(ii) $P[0.5 < X < 1] = F(1) - F(0.5) = \frac{1}{8}\left(1 - 0.5^3\right) = \frac{7}{64} = 0.109375$

(iii) Distribution function for $0 \le y \le \sqrt{2}$:

$$P[Y \le y] = P\left[X \le y^2\right] = F_X\left(y^2\right) = \frac{1}{8}y^6$$

Density function $0 \le y \le \sqrt{2}$:

$$f_Y(y) = \frac{6}{8}y^5$$

(iv) $E[X] = \frac{3}{8}\int_0^2 x^3 dx = \frac{3}{8}\left[\frac{1}{4}x^4\right]_0^2 = \frac{3}{32}2^4 = \frac{3}{32}16 = \frac{3}{2}$

$$E[Y] = \frac{6}{8}\int_0^{\sqrt{2}} y^6 dy = \frac{6}{8}\left[\frac{1}{7}y^7\right]_0^{\sqrt{2}} = \frac{6}{56}2^{3.5} = 1.212183$$

Most candidates did very well. However, candidates that were not very competent with differentiation and integration of functions made basic errors.

8 (i) $f_X(x) = \int_0^x 3x dy = [3xy]_{y=0}^{y=x} = 3x^2$ for $0 < x < 1$

$$f_Y(y) = \int_y^1 3x dx = \left[\frac{3}{2}x^2\right]_{x=y}^{x=1} = \frac{3}{2}\left(1 - y^2\right) \text{ for } 0 < y < 1$$

(ii) Not independent because $f_X(x) f_Y(y) \ne f_{XY}(x, y)$

(iii) $E[X] = \int_0^1 x f_X(x) dx = \left[\frac{3}{4}x^4\right]_0^1 = 0.75$

$$E[Y] = \int_0^1 y f_Y(y) dy = \left[\frac{3}{2}\left(\frac{y^2}{2} - \frac{y^4}{4}\right)\right]_0^1 = \frac{3}{8}$$

In part (i) many candidates could not identify correctly the range of integration, but part (ii) was well answered. In part (iii) again the wrong range of the variables was often used.

9 In thousands:

(i) $$\left[21 - t_{0.025,24}\frac{2.5}{5},\ 21 + t_{0.025,24}\frac{2.5}{5}\right] = \left[21 - 2.064\frac{1}{2},\ 21 + 2.064\frac{1}{2}\right]$$

$$[19.968,\ 22.032]$$

(ii) $H_0: \alpha \le 20$ v H_1: $\alpha > 20$
(or, H_0: $\alpha = 20$v H_1: $\alpha > 20$)

Test statistic: $$t = \frac{\bar{x} - \alpha_0}{s/\sqrt{n}} = \frac{21-20}{2.5 \times 0.2} = 2 > 1.711 = t_{0.05,24}$$

$$\frac{21-20}{2.5 \times 0.2} = 2 > 1.711 = t_{0.05,24}$$

We reject the null hypothesis.

(iii) The confidence interval in part (i) corresponds to a two-sided test. We found in part (i) that 20 is contained in the confidence interval, and we can therefore not reject the null hypothesis $H_0: \alpha = 20$ at a 5% significance level.
However, the one-sided test rejects $H_0: \alpha \le 20$ since only positive differences $\bar{X} - \alpha_0$ are considered. Answers are consistent.

(iv) $$\frac{21-\alpha_0}{2.5 \times 0.2} = 1.711, \quad 21 - \alpha_0 = 0.8555, \quad \alpha_0 = 20.1445$$

(v) Test H_0: $\lambda = 0.6$ v H_1: $\lambda \neq 0.6$

Test statistic (based on normal approximation to Poisson) is:

$$z = \frac{\bar{x} - \lambda_0}{\sqrt{\lambda_0 / n}} = \frac{0.5 - 0.6}{\sqrt{0.6/100}} = \frac{-0.1}{0.077} = -1.29 \in [-1.96, 1.96]$$

$$\frac{0.5 - 0.6}{\sqrt{0.6/100}} = \frac{-0.1}{0.077} = -1.29 \in [-1.96, 1.96]$$

(or, with continuity correction $z = \dfrac{0.5 + \frac{0.5}{100} - 0.6}{\sqrt{0.6/100}} = -1.226$)

The null hypothesis $H_0:\lambda = 0.6$ cannot be rejected for the year 2011.

(vi) Test $H_0:\lambda_{2012} \le \lambda_{2011}$ v $H_1:\lambda_{2012} > \lambda_{2011}$

(or, $H_0:\lambda_{2012} = \lambda_{2011}$ v $H_1:\lambda_{2012} > \lambda_{2011}$)

Overall sample mean $\hat{\lambda} = 0.55$

Test statistics now is: $z = \dfrac{\lambda_{2012} - \lambda_{2011}}{\sqrt{\dfrac{\hat{\lambda}}{n_1} + \dfrac{\hat{\lambda}}{n_2}}} = \dfrac{0.6 - 0.5}{\sqrt{1.1/100}} = \dfrac{0.1}{0.104} = 0.9535 < 1.64$

The null hypothesis $H_0:\lambda_{2012} \le \lambda_{2011}$ cannot be rejected at the 5% level. Therefore, we do not have empirical evidence to suggest that the alternative $\lambda_{2012} > \lambda_{2011}$ is true.

Generally well answered, although in part (ii) many candidates did not use the correct hypotheses for the test.

10 (i) $F = 29.12/36.62 = 0.795$ on 2,12df

$F_{2,12;10\%} = 2.807$

Therefore we cannot reject H_0 that there is no difference between the teams.

(ii) Clear outlier for team B, so not constant variance
(also other valid comments about non-normality etc.)

(iii) (a) Now $\sum y_B = 195.1, \sum y_B^2 = 9526.89$

So $\sum\sum y_{ij}^2 = 13873.06 + 9526.89 + 16305.82 = 39705.77$

$y_{..} = \sum\sum y_i = 263.2 + 195.1 + 285.4 = 743.7$

$$SS_T = 39705.77 - \frac{743.7^2}{14} = 199.36$$

$$SS_B = \left(\frac{263.2^2}{5} + \frac{195.1^2}{4} + \frac{285.4^2}{5}\right) - \frac{743.7^2}{14} = 155.08$$

$SS_R = 199.36 - 155.08 = 44.28$

	SS	df	MS
Between teams	155.08	2	77.538
Residual	44.28	11	4.025
Total	199.36	13	

(b) F = 77.538/4.025 = 19.26 on 2,11d.f.

$F_{2,11;1\%}$ = 7.206

So H_0 can be comfortably rejected at the 1% level – and there does seem to be a difference between teams

(iv) In (i) could not reject H_0 but in (iii) we did.

Outlier observation gave larger estimated variance which hid differences between groups.

Very well answered with only a few errors in calculations.

11 (i) (a) $$\hat{\beta} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{S_{xx}} = \frac{\sum(x_i - \overline{x})y_i}{S_{xx}}$$

$$E(\hat{\beta}) = \frac{\sum(x_i - \overline{x})}{S_{xx}} E(Y_i) = \alpha \frac{\sum(x_i - \overline{x})}{S_{xx}} + \beta \frac{\sum(x_i - \overline{x})x_i}{S_{xx}} = \beta$$

$$E(\hat{\beta}) = \frac{\sum(x_i - \overline{x})}{S_{xx}} E(Y_i) = \alpha \frac{\sum(x_i - \overline{x})}{S_{xx}} - \beta \frac{\sum(x_i - \overline{x})x_i}{S_{xx}} = \beta$$

(b) $$V(\hat{\beta}) = \frac{V\left(\sum(x_i - \overline{x})Y_i\right)}{(S_{xx})^2} = \frac{\sigma^2 \sum(x_i - \overline{x})^2}{(S_{xx})^2}$$ using independence of Y_i

and $$V(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

(c) With $z_i = x_i - \overline{x}$ we have $\overline{z} = 0$ and $\operatorname{cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\overline{z}\sigma^2}{S_{zz}} = 0$

i.e. the two estimators are uncorrelated which implies a better model for estimation.

(ii) There seems to be a decreasing relationship. However it is not clear if it is linear.

(iii) (a) H_0: $\beta = 0$ v. H_1: $\beta \neq 0$

$t = -0.2455/0.1015 = -2.419$ $t = -0.2455/0.1015 = 2.419$

with $t_8(2.5\%) = 2.306$ and $t_8(0.5\%) = 3.355$

At the 5% level we would reject the null in favour of the hypothesis that there is linear relationship between crawling age and average temperature (but we would not reject H_0 at the 1% level – or any level < 4.2%)

Alternative solutions:

Test H_0: $\rho = 0$ v H_1: $\rho \neq 0$

Use $R^2 = r^2$, giving $r = -0.65$.

` Under H_0: $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{-0.65\sqrt{8}}{\sqrt{1-0.4226}} = -2.419$

Then same as above.

Or, using Fisher's transformation:

$$w = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) = -0.7753$$

Under H_0: $W \sim N\left(0, \dfrac{1}{n-3}\right)$ or $\sqrt{7}W \sim N(0,1)$

$\sqrt{7}w = -2.051$, so conclusion is similar as before.

(b) The coefficient of determination R^2 is rather low, so the fit of the model does not seem good.

(iv) (a) Under the transformed data we have

$$\hat{y}_i = \hat{\gamma} + \hat{\delta} z_i = \hat{\gamma} + \hat{\delta}(x_i - \overline{x}) = (\hat{\gamma} - \hat{\delta}\overline{x}) + \hat{\delta}x_i$$

which is the same as $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ with $\hat{\alpha} = \hat{\gamma} - \hat{\delta}\overline{x} \Rightarrow \hat{\gamma} = \hat{\alpha} + \hat{\delta}\overline{x}$ and $\hat{\beta} = \hat{\delta}$.

Alternative solution:

With $z_i = x_i - \overline{x}$ we have

$$S_{zz} = \sum (z_i - \overline{z})^2 = \sum z_i^{\,2} = \sum \left(x_i - \overline{x}\right)^2 = S_{xx}$$

$$S_{zy} = \sum (z_i - \overline{z})(y_i - \overline{y}) = \sum (x_i - \overline{x})(y_i - \overline{y}) = S_{xy}$$

$$\hat{\delta} = \frac{S_{zy}}{S_{zz}} = \frac{S_{xy}}{S_{xx}} = \hat{\beta}$$

$$\hat{\gamma} = \overline{y} - \hat{\delta}\overline{z} = \overline{y} = \hat{\alpha} + \hat{\beta}\overline{x}$$

(b) Fitted model:

$$\hat{y}_i = 34.5501 - 0.2455 \times 11.2 - 0.2455\, z_i = 31.8005 - 0.2455\, z_i.$$
$$\hat{y}_i = 34.5501 - 0.2455 \times 11.2 - 0.2455\, x_i = 31.8005 - 0.2455\, x_i.$$

Part (i) was very poorly answered. The answers in this part can be derived using direct application of known results on statistics and probability that are explicitly given in the Core Reading, combined with basic algebraic skills. Parts (ii) and (iii) were well answered, while the performance in part (iv) was mixed.

END OF EXAMINERS' REPORT