INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2020

Subject CS1 Paper A – Actuarial Statistics Core Principles

Introduction

The Examiners' Report is written by the Chief 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 pertaining to the date that the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

Mike Hammer
Chair of the Board of Examiners
December 2020

A. General comments on the *aims of this subject and how it is marked*

1. The aim of the Actuarial Statistics subject is to provide a grounding in mathematical and statistical techniques that are of particular relevance to actuarial work.

2. Some of the questions in the examination 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.

3. Rounding errors were not penalised, but candidates lost marks where excessive rounding led to significantly different answers.

4. In cases where the same error was carried forward to later parts of the answer, candidates were given appropriate credit for the later parts.

5. In questions where comments were required, valid comments that were different from those provided in the solutions also received full credit where appropriate.

6. The paper included a number of multiple choice questions, where showing working was not required as part of the answer.
   In all multiple choice questions, the details provided in the answers below (e.g. calculations) are for information. Candidates were not be required to show working.

7. In all numerical questions that were not multiple-choice, full credit was given for correct answers that also included appropriate workings.

8. Standard keyboard typing was accepted for mathematical notation.

B. Comments on *candidate' performance in this diet of the examination.*

1. Performance was very satisfactory in general, with most candidates showing very good understanding of the topics in this subject. Well prepared candidates were able to score highly.

2. A smaller number of candidates appeared to be inadequately prepared, in terms of not having covered sufficiently the entire breadth of the subject.

3. Topics that were not particularly well answered in this paper include moment generating functions (Q4), GLMs (Q6) and non-standard CIs (Q9(iv), (vii)).

4. Questions that required higher order skills and comments were generally not well answered (e.g. Q8(iii)(b), Q9(iii), Q10(v)).

5. Questions corresponding to parts of the syllabus that had not been recently examined were generally poorly answered (e.g. Q4). This highlights the need for candidates to cover the whole syllabus when they revise for the exam and not only rely on themes appearing in past papers.

C. Pass Mark

The Pass Mark for this exam was 60.
1189 presented themselves and 823 passed.

Solutions for Subject CS1 Paper A September 2020

Q1
(i)
From the Central Limit Theorem, approximately

$T \sim N(81 \text{ x } 5\ , 81 \text{ x } 4), i.e. N(405, 18^2), or N(405, 324)$ [2]

(ii)
Standardising, we get:
$P(T > 369) = P\left(\frac{T-405}{18} > \frac{369-405}{18}\right)$ [1]

$\approx P(Z > -2)$ [½]

$= 0.97725$ [½]

using tables.

[Total 4]

Generally very well answered. In part (ii) some candidates applied a continuity correction, which was not needed.

Q2
(i) **Ans**: **A2** [2]

$P[X + Y = 7] = P[X = 1, Y = 6] + P[X = 2, Y = 5] + P[X = 3, Y = 4] + P[X = 4, Y = 3]P + [X = 5, Y = 2] + P[X = 6, Y = 1]$

$$= P[X = 1]P[Y = 6] + \cdots + P[X = 6]P[Y = 1] = 6 \times \frac{1}{36} = \frac{1}{6}$$

(ii) **Ans: A3** [2]

$P[X = 3, Y = 3] + P[X = 3, Y \neq 3] + P[X \neq 3, Y = 3] =$
$1 - P[X \neq 3, Y \neq 3] = 1 - \frac{5}{6} \times \frac{5}{6} = \frac{11}{36}$

(iii) **Ans: A2** [2]

$$1 - P[X \in \{2,4,6\}]P[Y \in \{2,4,6\}] = 1 - \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}$$

(iv) (a) **Ans: A4** [1]

$E[X + Y|X = 4] = E[X|X = 4] + E[Y|X = 4] = 4 + E[Y]$

(b) We assume that X and Y are independent (pair of fair dice). [1]

(c) $E[X+Y|X=4] = 4 + \frac{1}{6}(1+2+\cdots+6) = 4 + \frac{21}{6} = 7.5$ [2]

[Total 10]

The question was very well answered by candidates.

Q3

(i) **Ans: A2** [2]

The required probability is: P(TV made in Factory B | defective)

Using Bayes' theorem:

$$= \frac{\text{P(defective|made in factory B)} \times \text{P(made in factory B)}}{\text{P(defective|factory A)P(factory A)} + \text{P(defective|factory B)P(factory B)} + \text{P(defective|factory C)P(factory C)}}$$

(ii) P(TV made in Factory B | defective)

$$= \frac{0.015 \times 0.4}{0.02 \times 0.35 + 0.015 \times 0.4 + 0.01 \times 0.25} = 0.38710$$

[2]

[Total 4]

Part (i) was well answered. In part (ii), a number of candidates despite identifying the correct answer in (i), went on to calculate incorrect probabilities. In some cases this was due to misinterpreting the probabilities in the table.

Q4

(i) Integrate from b to plus infinity. [1]

(ii) **Ans: A4** [2]

Moment generating function of Y is:

$$M_Y(t) = E(e^{tY}) = \int_b^\infty e^{ty}\, ae^{-5y} dy$$

$$= a\int_b^\infty e^{-(5-t)y} dy$$

$$= a\left[-\frac{e^{-(5-t)y}}{5-t}\right]_b^{\infty} = \frac{ae^{-(5-t)b}}{5-t}$$

(iii) $t < 5$ [1]

(iv) Evaluating the function at $t = 0$ gives 1. [1]

We obtain a = 5e^(5b) [2]

[Total 7]

This question was not answerd well in general – particulalrly parts (iii) and (iv). This was a type of question that is not examined very often. Candidates are advised to cover the whole syllabus when they revise for the exam and not only rely on themes appearing in past papers.

Q5

(i) $S_{xx} = 2014 - \frac{141^2}{10} = 25.9$ [1]

$S_{yy} = 1629 - \frac{127^2}{10} = 16.1$ [1]

$S_{xy} = 1810 - \frac{141\times127}{10} = 19.3$ [1]

(ii) $r = \frac{19.3}{\sqrt{25.9\times16.1}} = 0.9451364$ [1]

(iii) $\hat{b} = \frac{19.3}{25.9} = 0.745$ [1]

$\hat{a} = \frac{127}{10} - 0.745 \times \frac{141}{10} = 2.193$ [1]

[Total 6]

This is a typical regression/correlation question and was answered very well.

Q6

(i) A distribution of the response variable Y. [1]
A "linear predictor" η [1]
A "link function" g [1]

(ii) The distribution of the response D_x is a Poisson distribution. [1]
The linear predictor $\eta_x = a + bx$. [1]
The link function is the logarithm since $log(E[D_x]) = \eta_x$. [1]

(iii) **Ans: A1** [2]

(iv) **Ans: A3** [2]

[Total 10]

Parts (i) and (ii) were well answered, whereas many candidates gave wrong answers in parts (iii) and (iv). These concerned a direct application of likelihood estimation in a less typical scenario, as compared to the setting usually appearing in estimation questions.

Q7

(i) **Ans: A2** [2]

(ii) **Ans: A4** [3]

(iii) The expectation of *X* is correct.
This is obtained by taking the derivative of b(theta). [1]

The standard deviation is not correct. In fact it is the variance that is s squared. [1]
It is obtained by taking the second derivative of b(theta) and multiply by a(phi). [1]

(iv) A factor takes a categorical value and for a factor with k levels, there are generally k parameters. [1]

For a numerical variable, the value is included as such in the linear predictor and there is a single parameter in the model for each numerical variable. [1]

[Total 10]

Parts (i) and (ii) were well answered. Part (iii) was overall answered well, with a common problem of failing to identify that the standard deviation is incorrect, or not making any comments. Part (iv) was poorly answered, often with no mention regarding parameters and levels.

Q8

(i) In this case, n = 2 and N = 4. Therefore the estimates are:

(a) $E[m(\theta)] \;=\; \bar{x} = \frac{1}{4}\sum_{i=1}^{4}\bar{x}_i = \frac{1}{4}(36+40+20+62) = 39.5$ [1]

(b) $E[s^2(\theta)] = \frac{1}{4}\sum_{i=1}^{4}\frac{1}{1}\sum_{j=1}^{2}(x_{ij} - \bar{x}_i)^2 = (98 + 8 + 8 + 72)/4 = 46.5$ [1]

(c) $Var[m(\theta)] = \frac{1}{3}\sum_{i=1}^{4}(\bar{x}_i - \bar{x})^2 - \frac{1}{2}\left[\frac{1}{4}\sum_{i=1}^{4}\frac{1}{1}\sum_{j=1}^{2}(x_{ij} - \bar{x}_i{}^2)\right]$

$= \frac{1}{3}[(36 - 39.5)^2 + (40 - 39.5)^2 + (20 - 39.5)^2 + (62 - 39.5)^2] - \frac{1}{2}(46.5)$
$= 299\frac{2}{3} - 23\frac{1}{4}$
$= 276.42$ [2]

(ii) The credibility factor is:

$$\frac{2}{2+\frac{46.5}{276.41}} = 0.92241$$ [1]

And the estimate of X_{13} is (0.92241 x 36) + (1 – 0.92241) x 39.5 = 36.272 [1]

(iii)(a) Assumption 1
The distribution of each X_{ij} (i = 1, 2, 3, 4 and j = 1, 2) depends on the value of a parameter θ_i, whose value is fixed, unknown, and the same for each value of j. [1]

Assumption 2
Given θ_i (i = 1, 2, 3, 4), X_{ij} (j = 1, 2) are independent and identically distributed. [1]

(iii)(b) For the given data, the assumptions can be interpreted as saying:
- The number of calls received follows a distribution with a parameter that varies according to the time of year, but that is constant between years. [2]

[Total 10]

Parts (i) and (ii) were well answerd – except from (i)(c) where the calculation of the variance was often incorrect. Part (iii)(b) was poorly answered, with the interpretation in the context of the question scenario being handled poorly. Note that alternative assumptions (as in the Core Reading) were given credit as appropriate.

Q9

(i) Z_X has a chi-squared distribution [1]
with $n - 1 = 299$ degrees of freedom [1]

(ii) $E[Z_X] = 299$ [1]
and $Var(Z_X) = 598$ [1]

(iii) A chi-squared distribution with 299 degrees of freedom is the distribution of a sum of 299 independent random variable that are all squared standard normally distributed. [1]

It follows from the CLT that a chi-squared distribution with a large number of degrees of freedom can be approximated with a normal distribution. [1]

(iv) $P[Z_x \le q] = P\left[\frac{Z_x - 299}{\sqrt{598}} \le \frac{q-299}{\sqrt{598}}\right] = P\left[Z \le \frac{q-299}{\sqrt{598}}\right]$

$\frac{q_{97.5}-299}{\sqrt{598}} = 1.96$ and $q_{97.5} = 299 + 1.96 \times \sqrt{598} = 346.93$ [1½]

$\frac{q_{2.5}-299}{\sqrt{598}} = -1.96$ and $q_{2.5} = 299 - 1.96 \times \sqrt{598} = 251.07$ [1½]

(v) 95% confidence interval (using normal approximation of t-distribution):

Income: $\left[1838 - 1.96\frac{211}{\sqrt{300}}, 1838 + 1.96\frac{211}{\sqrt{300}}\right]$ [1]

$= [1814.12, 1861.88]$ [1]

(vi) 95% confidence interval (using normal approximation of t-distribution):

Rent: $\left[608 - 1.96\frac{275}{\sqrt{300}}, 608 + 1.96\frac{275}{\sqrt{300}}\right]$ [1]

$= [576.88,639.12]$ [1]

(vii) $\left[\frac{299\times 211^2}{\chi^2_{0.975}}, \frac{299\times 211^2}{\chi^2_{0.025}}\right] \approx \left[\frac{299\times 211^2}{346.93}, \frac{299\times 211^2}{251.07}\right]$ [1]

$= [38370.22,\ 53020.19]$ [1]

(viii) **Ans: A4** [2]

(ix) $S_{xy} = \sum x_i y_i - (\sum x_i)(\sum y_i)/n = 348 \times 10^6 - 1838 \times 300 \times 608$
$= 12{,}748{,}800$

[1]

$S_{xx} = 299 \times 211^2 = 13{,}311{,}779$

[1]

$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{12{,}748{,}800}{13{,}311{,}779} = 0.9577082$

[1]

$\hat{a} = \bar{y} - \hat{b}\bar{x} = 608 - 0.9577082 \times 1838 = -1152.268$

[1]

[Total 21]

Parts (i) and (ii) were well answered. In part (iii) the reasoning was often inadequate. Parts (iv) and (vii) were poorly answered or unattempted, with many candidates failing to calculate the quantiles required. Parts (viii) and (ix) were reasonably well answered.

Q10

(i) In bivariate data, the response variable is a random variable whose value is influenced by the explanatory variable. [1]

(ii) There is an increasing and relatively linear relationship. [1]
However the trend and linearity are not very clear around values x = 5, 6. [1]

(iii) (a) **Ans: A1** [2]

$W = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$ is normally distributed with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ and standard deviation $1/\sqrt{n-3}$. $W = 0.8673$ and $W \sim N(0,\ 1/7)$.
Test statistic = $0.867/\left(\frac{1}{7}\right)^{0.5} = 2.295$.

(b) This is a two-sided test with the 2.5% critical values being -1.96 and 1.96 [2]
So we reject H_0 at 5% significance level and conclude that Pearson's correlation coefficient is significantly different from zero. [1]
[Alternatively, use p-value = 0.022 for same conclusion.]

(iv) (a) **Ans: A3** [2]

$S_{xx} = 462 - \frac{66^2}{10} = 26.4$
$S_{yy} = 335975 - \frac{1825^2}{10} = 2912.5$

$S_{xy} = 12240 - \frac{66\times1825}{10} = 195$

(b)
$\hat{\sigma}^2 = \frac{1}{8}\left(2912.5 - \frac{195^2}{26.4}\right) = 184.02$
s.e.$(\hat{\beta}) = (\hat{\sigma}^2/S_{xx})^{1/2} = (184.02/26.4)^{1/2} = 2.64$

$$\hat{\beta} = \frac{195}{26.4} = 7.386$$

Test statistic $= 7.386/2.64 = 2.80$ [2]

(c)
The test statistic follows a t-distribution with 8 df under the null hypothesis. [1]

(d)
This is a two-sided test with the 2.5% critical values being -2.306 and 2.306. [2]
We have evidence at 5% significance level to reject the null hypothesis that $\beta = 0$. [1]

(v) The two tests are actually similar therefore it is not surprising that they yield to the same conclusion that there is a linear relationship between house prices and school indices. [2]

[Total 18]

There were no particular isues with part (i). In part (ii), many candidates failed to make any comment regarding the unclear trend in part of the data. A common error in parts (iii) and (iv) was to not use a two sided test. Part (v) was poorly answered, often with no mention of the two tests being similar.

END OF EXAMINERS' REPORT