Institute
and Faculty
of Actuaries

EXAMINERS' REPORT

CS1 - Actuarial Statistics
Core Principles
Paper B

September 2023

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.

For some candidates, this may be their first attempt at answering an examination using open books and online. The Examiners expect all candidates to have a good level of knowledge and understanding of the topics and therefore candidates should not be overly dependent on open book materials. In our experience, candidates that spend too long researching answers in their materials will not be successful either because of time management issues or because they do not properly answer the questions.

Many candidates rely on past exam papers and examiner reports. Great caution must be exercised in doing so because each exam question is unique. As with all professional examinations, it is insufficient to repeat points of principle, formula or other text book works. The examinations are designed to test "higher order" thinking including candidates' ability to apply their knowledge to the facts presented in detail, synthesise and analyse their findings, and present conclusions or advice. Successful candidates concentrate on answering the questions asked rather than repeating their knowledge without application.

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.

Sarah Hutchinson
Chair of the Board of Examiners
November 2023

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

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

In particular, the CS1B paper is a problem-based examination and focuses on the assessment of computer-based data analysis and statistical modelling skills.

For the CS1B exam candidates are expected to include the R code that they have used to obtain the answers, together with the main R output produced, such as charts or tables.

When a question requires a particular numerical answer or conclusion, this should be explicitly and clearly stated, separately from, and in addition to the R output that may contain the relevant numerical information.

Some of the questions in the examination paper accept alternative solutions from those presented in this report, or different ways in which the provided answer can be determined. In particular, there are variations of the R code presented here, which are valid and can produce the correct output. All mathematically and computationally valid solutions or answers received credit as appropriate.

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

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

In cases where a question is based on simulations, and no seed was specified, all numerical answers provided in this document are examples of possible results. The numerical values presented here will be different if the simulations are repeated.

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

Overall performance in CS1B was satisfactory. Well prepared candidates were able to score highly.

Comments given alongside the R output were not always clear or adequate. Some candidates failed to provide clear numerical answers as required in the questions. Also, the main R output was missing in some cases.

A number of candidates provided their answers in a layout that was not satisfactory, with input and output being separated and presented with large gaps, and with significant duplications.

C. Pass Mark

The Pass Mark for this exam was 60.
1508 presented themselves and 730 passed.

Solutions for Subject CS1B - September 2023

Q1

(i)

R code:

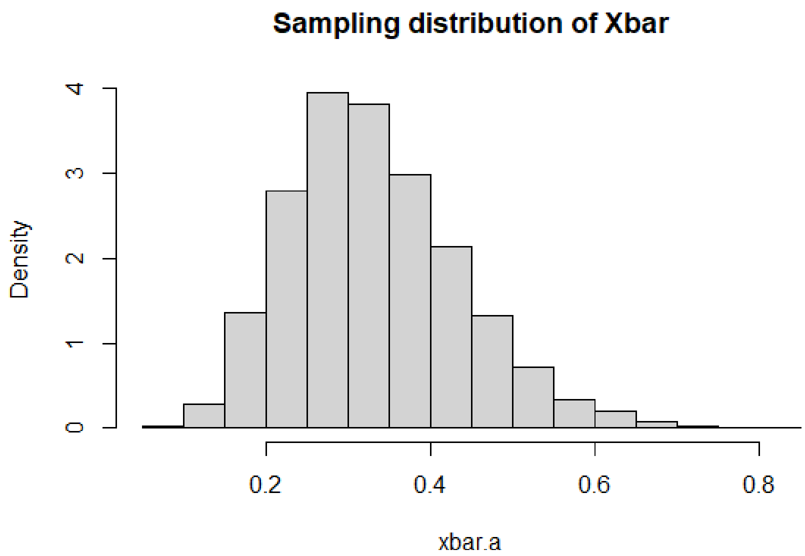
```
set.seed(3202)
n = 10; theta = 3
m1 = matrix(0,nrow=10^4,ncol=n)  # Initialise matrix, each
sample in a row
for (i in 1:10^4){                                      [½]
u = runif(n,0,1)                                        [1]
m= - log(1-u)/theta    # Generate samples in rows       [1]
m1[i,] = m                                              [½]
}
```

(ii)

R Code:

```
xbar.a = apply(m1,1,mean)                               [1]

hist(xbar.a,freq=F,main="Sampling distribution of Xbar")  [1]
```



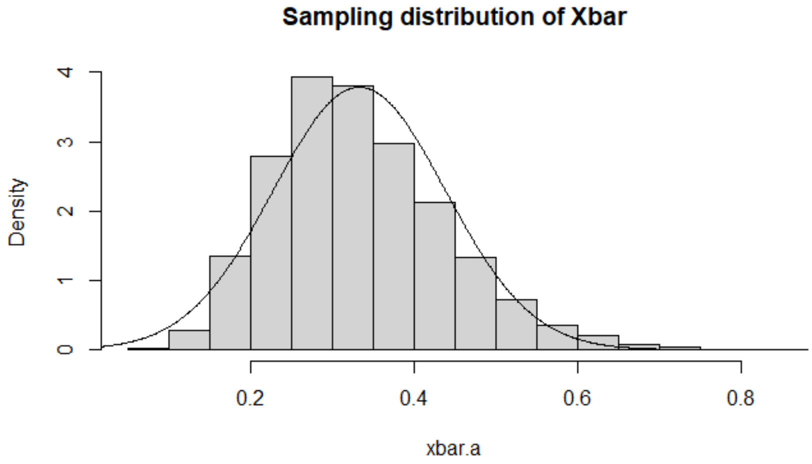
[1]

(iii)

Using the CLT, the sampling distribution should be N(1/3, 1/90). [2]

R code:

```
x.grid = seq(0,1,length=1000)                           [1]
lines(x.grid,dnorm(x.grid,1/3,sqrt(1/90)))              [2]
```



[1]

The curve shows the exact pdf of the sampling distribution. [1]

(iv)
The sampling distribution is relatively asymmetrical compared to the normal. [1]
The CLT does not work very well here [1]
as the sample size is small (n = 10). [1]

[Total 16]

Parts (i)-(iii) were well answered in general. A common error in part (i) was not using properly the inverse transform method as the question asked.

Part (iv) was not well answered, with comments often being vague or not consistent with the graphs plotted earlier.

Q2

(i)
Simulate samples:

```
set.seed(12345)
data.A = rbinom(900,1,0.020)
data.B = rbinom(1200,1,0.025)
#data.A; data.B
```

[2]

(ii)

```
n.A = length(data.A); n.B = length(data.B)
thetahat.A = sum(data.A)/n.A
thetahat.B = sum(data.B)/n.B
thetahat.A; thetahat.B
# 0.02111111
# 0.02916667
```

[2]


```
ese.diff = sqrt( thetahat.A*(1-thetahat.A)/n.A +
thetahat.B*(1-thetahat.B)/n.B )
ese.diff
# 0.006823359
```
[1]

```
ci.theta.diff = (thetahat.A-thetahat.B) + c(-
1,1)*qnorm(0.995)*ese.diff
ci.theta.diff
#[1] -0.025631365  0.009520254
```
[1]

The 99% CI for the difference in proportions is [-0.025631365, 0.009520254] [1]

(iii)
Hypotheses:
```
H0: theta_A = theta_B v. H1: theta_A < theta_B
```
[1]

Statistic:
```
thetahat.common = (sum(data.A)+sum(data.B))/(n.A+n.B)
thetahat.common
# 0.02571429
```
[1]

```
ese.diff.H0 = sqrt( thetahat.common*(1-thetahat.common) *
(1/n.A + 1/n.B))
ese.diff.H0
# 0.006979562
```
[1]

```
stat.diff = (thetahat.A-thetahat.B)/ese.diff.H0
stat.diff
# -1.154163
```
The value of the statistic is -1.154163 [1]

P-value:
```
pvalue.diff =[ pnorm(stat.diff)
pvalue.diff
# 0.1242166
```
The p-value of the test is 0.1242166 [2]

Conclusion:
We do not have evidence against the null hypothesis of equal proportions [1]
at level 12.4% [1]
so we do not reject it. [1]

[Total 16]

Part (i) was well answered in general. A common error was using rbinom(1, n, p), therefore not producing the detailed sample.

In part (ii) there were mixed answers with various errors.

Answers in part (iii) were also mixed, with common errors including not stating correctly, or at all, the hypotheses, parameter errors and not providing a clear conclusion.

Q3

#Initialise the dataset.
```
> data("cars")
> attach(cars)
```

(i)
```
> model = lm(dist~speed, data = cars)                    [1]

> model
```

Call:
```
lm(formula = dist ~ speed, data = cars)
```

Coefficients:
```
(Intercept)    speed
-17.579        3.932                                      [1]
```

(ii)
```
> data_predict <- data.frame(speed =
c(6,7,16,22,28,33,40,42,57,64))                           [2]

> predict(model,newdata=data_predict)                     [2]
       1         2         3         4         5          6
6.015358  9.947766 45.339445 68.933898 92.528350 112.190394
         7          8          9         10
139.717255 147.582073 206.568204 234.095066               [1]
```

(iii)
```
> predict(model,newdata=data_predict,
interval="confidence",level=0.90)                         [3]

        fit        lwr       upr
1   6.015358  -1.482803  13.51352
2   9.947766   3.050136  16.84540
3  45.339445  41.667593  49.01130
4  68.933898  63.063300  74.80450
5  92.528350  83.019704 102.03700
6 112.190394  99.393826 124.98696
```

```
7  139.717255 122.189498 157.24501
8  147.582073 128.688793 166.47535
9  206.568204 177.348219 235.78819
10 234.095066 200.029446 268.16069
```
[1]

(iv)(a)
The CI is constructed such that with repeated sampling, 90% of the obtained intervals will contain the true unknown mean stopping distance of cars travelling at 64 mph. [1]

(b)
For speed 6mph the model gives a CI with negative lower endpoint. [½]
A linear model, predicting negative stopping distance values, may not be appropriate. [½]

[Total 13]

Parts (i)-(iii) were very well answered in general.

The answers in part (iv) varied, with a number of candidates giving unclear comments.

Q4
(i)(a)
The probability of heads or tail for a fair coin is 0.5. [½]

```
round(1 - pbinom(115,200,0.5) + pbinom(84,200,0.5),2)
```
[4]

Answer is 0.03 [½]

The probability of rejecting the hypothesis of fair coin when it is actually correct is 0.03.

(b)
As the sample size n is large, the binomial distribution can be approximated by the normal distribution with mean np and variance $np(1-p)$ using CLT. [1]

Applying continuity correction, the probability of rejecting the hypothesis of fair coin when it is actually correct, is calculated as:

```
round(1 - pnorm((115.5 - 100)/sqrt(50)) + pnorm((84.5-
100)/sqrt(50)), 2)
```
[3½]

Answer is: 0.03 [½]

(c)
We obtain the same answer and conclude the approximation works well. [1]

(ii)
A coin is judged fair if in a single sample of 200 tosses, between 85 and 115 (inclusive) heads occur, with $p = 0.7$ here:

```
round(pbinom(115, 200, 0.7) - pbinom(84, 200, 0.7),4)
```
[3]

```
# 0.0001
```
[1]

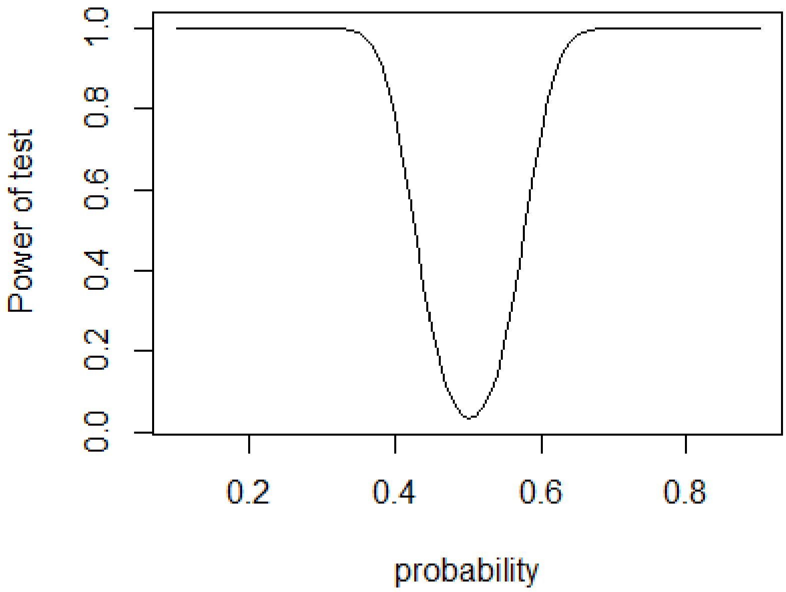
(iii)

```
p = seq(0.1, 0.9, 0.01)
```
[1]

```
pow = round(1 - (pbinom(115, 200, p) - pbinom(84, 200, p)),4)
```
[3]

(iv)

```
plot(p,pow, type="l", ylab = "Power of test", xlab = "probability", main="Power of test")
```
[2]



[1]

(v)
The power of the test increases as we get away from the fair coin probability. [1]
This is consistent with the definition of the power of a test since the ability to reject the null hypothesis of coin being fair should be higher when the hypothesis should be rejected. [1]

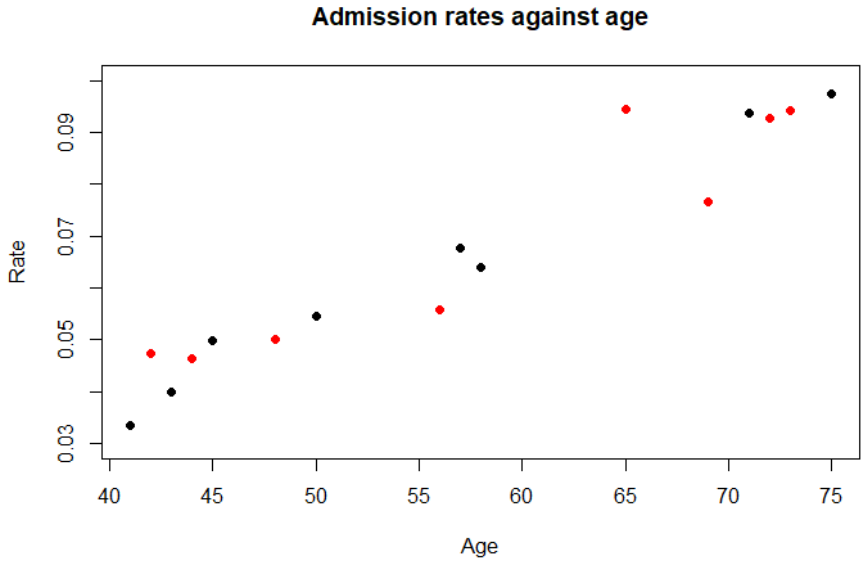
[Total 24]

Parts (i)-(iv) were answered reasonably well, with common errors including incorrect parameterisations for the binomial distributions involved.

Part (v) was not well answered with a number of candidates not attempting it.

Q5
(i)
R code:


```
plot(Age[History==0],Rate[History==0],pch=19,xlim=c(41,75),y
lim=c(0.03,0.10),xlab="Age",ylab="Rate",  main="Admission
rates against age")                                         [2]
points(Age[History==1],Rate[History==1],pch=19, col =
"red")                                                      [1]
```

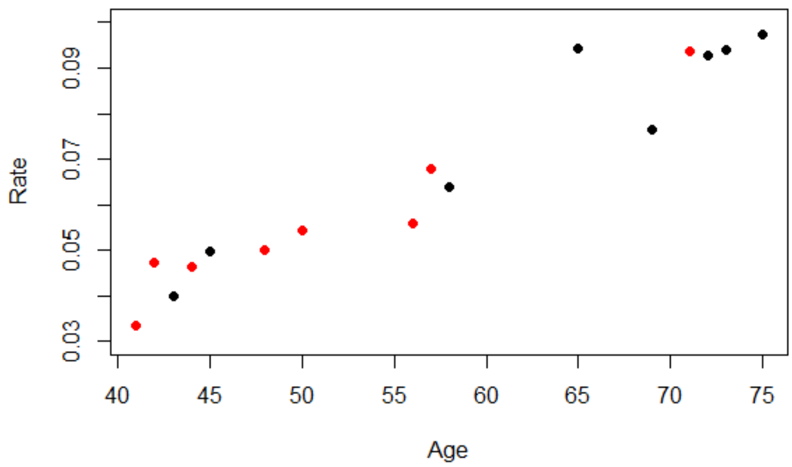


Black bullets: History = 0; red bullets: History = 1. [2]

(ii)
R code:

```
plot(Age[Operation==0],Rate[Operation==0],pch=19,xlim=c(41,
75),ylim=c(0.03, 0.10),xlab="Age",ylab="Rate")              [2]

points(Age[Operation==1],Rate[Operation==1],pch=19, col =
"red")                                                      [1]
```



`Black bullets: Operation = 0; red bullets: operation = 1.` [2]

(iii)
Age seems to have a clear increasing effect on rates. [1]
The effects of History and Operation are not clear. [1]

(iv)
R code:
```
m1 = glm(Rate ~ History + Operation + Age, family =
gaussian(link="log"))
```
[2]

```
summary(m1)

#Call:
#glm(formula = Rate ~ History + Operation + Age, family =
gaussian(link = "log"))
#
#Deviance Residuals:
#    Min          1Q      Median          3Q         Max
#-0.0096278  -0.0027215  -0.0004823   0.0030534   0.0165070
#
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) -4.232037   0.184121 -22.985 2.73e-11 ***
#History1     0.008975   0.051115   0.176    0.864
#Operation1   0.010137   0.061309   0.165    0.871
#Age          0.025690   0.002616   9.822 4.35e-07 ***
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
#
#(Dispersion parameter for gaussian family taken to be
4.6878e-05)
#
#    Null deviance: 0.00738638  on 15  degrees of freedom
#Residual deviance: 0.00056253  on 12  degrees of freedom
#AIC: -108.68
```
[1]

(v)
History does not have a significant impact since p-value is large (0.864). [½]
Operation does not have a significant impact since p-value is large (0.871). [½]

Age has a highly significant [1]
positive effect since p-value is very small, 4.35e-07. [1]

(vi)
R code:
```
m2 = glm(Rate ~ History + Operation + Age + Comorbidity,
family = gaussian(link="log"))
```
[2]
```
summary(m2)
```

```
#Call:
#glm(formula = Rate ~ History + Operation + Age +
Comorbidity,
#    family = gaussian(link = "log"))
##
#Deviance Residuals:
#      Min         1Q     Median         3Q        Max
#-0.005923  -0.002551  -0.001152   0.002753   0.006639
#
#Coefficients:
#              Estimate Std. Error t value Pr(>|t|)
#(Intercept)  -4.203456   0.151308 -27.781 4.93e-10 ***
#History1      0.005103   0.041287   0.124   0.9044
#Operation1    0.006746   0.050478   0.134   0.8966
#Age           0.027170   0.002302  11.803 8.86e-07 ***
#Comorbidity2 -0.182506   0.058996  -3.094   0.0129 *
#Comorbidity3 -0.149206   0.059872  -2.492   0.0343 *
#Comorbidity4 -0.125286   0.059462  -2.107   0.0644 .
#---
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1
#
#(Dispersion parameter for gaussian family taken to be
2.927809e-05)
#
#    Null deviance: 0.0073864  on 15  degrees of freedom
#Residual deviance: 0.0002635  on  9  degrees of freedom
#AIC: -114.82
```
[1]

(vii)
As before, History and Operation do not have a significant impact, while age still has a highly significant positive effect. [1½]
Comorbidity has significant impact, [1]
with levels 2-3 (comorbidities B-C) having a significant reducing effect as compared to level 1 (baseline, comorbidity A) [1]
and level 4 (comorbidity D) having a marginally (non-)significant negative impact. [½]

(viii)
The model with the 4 covariates in part (vi) should be preferred, [1]
as it has smaller AIC. [1]

(ix)
Predicted link function value is:
```
- 4.203456 - 0.182506 + 0.027170 * 62
=  -2.701422
```
[1]

and predicted value is:
```
exp(-2.701422) = 0.06711001
```
[1]

(x)
The used GLM has a logarithmic link,
which implies that predicted rates can be above 1. [2]

[Total 31]

In parts (i)-(iii) answers were mixed, with many candidates failing to distinguish between the different history and operation levels.

Parts (iv)-(viii) were generally well answered. Common errors included failing to use a logarithmic function as specified in the question and comments often being limited or vague.

Answers in part (ix) were mixed, with a number of candidates failing to use the correct link function.

Part (x) was not well answered, with many candidates not attempting it.

[Paper Total 100]

END OF EXAMINERS' REPORT

Institute and Faculty of Actuaries

Beijing

14F China World Office 1 · 1 Jianwai Avenue · Beijing · China 100004
Tel: +86 (10) 6535 0248

Edinburgh

Level 2 · Exchange Crescent · 7 Conference Square · Edinburgh · EH3 8RA
Tel: +44 (0) 131 240 1300

Hong Kong

1803 Tower One · Lippo Centre · 89 Queensway · Hong Kong
Tel: +852 2147 9418

London (registered office)

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100

Oxford

1st Floor · Belsyre Court · 57 Woodstock Road · Oxford · OX2 6HJ
Tel: +44 (0) 1865 268 200

Singapore

5 Shenton Way · UIC Building · #10-01 · Singapore 068808
Tel: +65 8778 1784