Subject CS1

CMP Upgrade 2023/24

CMP Upgrade

This CMP Upgrade lists the changes to the Syllabus, Core Reading and the ActEd material since last year that might realistically affect your chance of success in the exam. It is produced so that you can manually amend your 2023 CMP to make it suitable for study for the 2024 exams. It includes replacement pages and additional pages where appropriate.

Alternatively, you can buy a full set of up-to-date Course Notes / CMP at a significantly reduced price if you have previously bought the full-price Course Notes / CMP in this subject. Please see our 2024 *Student Brochure* for more details.

We only accept the current version of assignments for marking, *ie* those published for the sessions leading to the 2024 exams. If you wish to submit your script for marking but only have an old version, then you can order the current assignments free of charge if you have purchased the same assignments in the same subject in a previous year, and have purchased marking for the 2024 session.

This CMP Upgrade contains:

- all significant changes to the Syllabus and Core Reading
- additional changes to the ActEd Course Notes and Assignments that will make them suitable for study for the 2024 exams.

1 Changes to the Syllabus

The syllabus objectives across all subjects have been reworded. More detail is given on the IFoA's website at **actuaries.org.uk/curriculum/**.

The structure of the syllabus objectives has changed. Data analysis is now topic 1 and Random variables and distributions is now topic 2. The objectives relating to Random sampling and sampling distributions have been moved out of the Data analysis topic and into the Random variables and distributions topic. They are now under syllabus objective 2.6.

The following syllabus objectives have also been removed:

1.2.5 Define the probability function/density function of the sum of two independent random variables as the convolution of two functions.

1.2.7 Use generating functions to establish the distribution of linear combinations of independent random variables.

2.2.1 Describe the purpose of exploratory data analysis.

5.1.4 Explain what is meant by a loss function.

2 Changes to the Core Reading and ActEd text

This section contains all the ***non-trivial*** changes to the Core Reading and ActEd text.

Chapter 13

Section 1

There have been various corrections on page 7. From the second paragraph on it should read:

We now need a link function. η here will not necessarily take a value in the interval (0, 1). Depending on the values of α_i, β_1 and β_2, η might take any value. If we use the link function $g(\mu)=\log\left(\frac{\mu}{1-\mu}\right)$ and set this equal to the linear predictor η, we have $\log\left(\frac{\mu}{1-\mu}\right)=\eta$. We invert this function to make μ the subject to give $\mu=\frac{e^\eta}{1+e^\eta}=\frac{1}{e^{-\eta}+1}=\left(1+e^{-\eta}\right)^{-1}$. We can now see that μ will lie in the range from zero to one, and so can be used as a pass rate.

We now use maximum likelihood estimation to estimate the four parameter values: α_Y, α_N (the α parameters corresponding to having attended tutorials and not having attended tutorials, respectively), β_1 (the parameter for the number of assignments) and β_2 (the parameter for the mock mark). To do this we need (ideally) the actual exam results of a large sample of students who fall into each of the categories.

Having done this for a set of data, we might come up with the following parameter values for the linear predictor:

$$\alpha_Y=-1.501 \qquad \alpha_N=-3.196 \qquad \beta_1=0.5459 \qquad \beta_2=0.0251$$

We can now use the linear predictor and link function to predict pass rates for groups of students with a particular characteristic. For example, for a student who attends tutorials, submits three assignments and scores 65% on the mock, we have:

$$\eta=-1.501+0.5459\times 3+0.0251\times 65=1.7682$$

We now use the inverse of the link function to calculate μ:

$$\mu=\left(1+e^{-1.7682}\right)^{-1}=0.8542$$

So the model predicts an 85% probability of passing for a student in this situation. So in this particular situation, the linear predictor is $\eta=\alpha_i+\beta_1 N+\beta_2 S$ and the link function is $g(\mu)=\log\left(\frac{\mu}{1-\mu}\right)$.

Section 5.3

The section after the solution on pages 39 to 40 on the saturated model has been updated. Replacement pages 39 to 40a are attached.

Section 5.5

The seventh paragraph on page 43 has been updated to the following:

What we are trying to do here is to decide whether the added complexity results in significant additional accuracy. If the result is significant we would reject H_0, and conclude that Model 2 is a significant improvement over Model 1. If not, then it would be preferable to use the model with fewer parameters.

The second paragraph after the solution at the top of page 44 has been updated to the following:

In the first case, the difference between the models is $\beta_2 x^2$, and so a significant difference between the models tells us that the quadratic term should be included. In the second case, the difference between the models is $\beta_3 \log x - \beta_2 x^2 - \beta_1 x$, and so a significant difference doesn't tell us *which* parameter is significant and the test cannot be used.

Practice Question 13.4

On page 60, the marks have been updated in part (v) to 3, and the total marks increased to 16.

Practice Question 13.9

On page 62, the marks have been updated in part (iv) to 6, and the total marks increased to 17.

Practice Solution 13.4

On page 66, the null and alternative hypotheses have been added to part (v) and the marks have been updated:

(v) ***Comparing A with B***

We are testing:

H_0 : Model A is not a significant improvement over Model B

H_1 : Model A is a significant improvement over Model B [½]

We can use the chi-squared distribution to compare Model A with Model B. We calculate the difference in the scaled deviances (which is just $2(\log L_A - \log L_B)$):

$$35.03 - 24.93 = 10.10$$ [½]

The number of degrees of freedom in the χ^2 distribution is the number of additional parameters in Model A. This is $3-1=2$. [½]

The upper 5% point of the χ_2^2 distribution is 5.991. [½]

Since $10.10 > 5.991$, there is sufficient evidence to reject the null hypothesis at the 5% significance level. So we conclude that Model A is a significant improvement over Model B. We prefer Model A here. [1]

Practice Solution 13.9

On page 72, the marks for (iv)(a) is now 2. In (b) the null and alternative hypotheses have been added and the marks have been updated.

(iv)(b) ***Compare models***

Comparing the constant model and Model 1

We are testing:

H_0 : Model 1 is not a significant improvement over the constant model

H_1 : Model 1 is a significant improvement over the constant model [½]

The difference in the scaled deviances is 40.

This is greater than 7.815, the upper 5% point of the χ_3^2 distribution. [½]

So we have sufficient evidence at the 5% significance level to reject the null hypothesis. We conclude that Model 1 **is** a significant improvement over the constant model. [½]

Alternatively, if we use the AIC to compare models, we find that since $\Delta(\text{deviance}) > 2 \times \Delta(\text{parameters})$ *, Model 1* ***is*** *a significant improvement over the constant model.*

Comparing Model 1 and Model 2

We are testing:

H_0 : Model 2 is not a significant improvement over Model 1

H_1 : Model 2 is a significant improvement over Model 1

The difference in the scaled deviances is 5.

This is greater than 3.841, the upper 5% point of the χ_1^2 distribution. [½]

So we have sufficient evidence at the 5% significance level to reject the null hypothesis. We conclude that Model 2 **is** a significant improvement over Model 1. [½]

Alternatively, if we use the AIC to compare models, we find that since $\Delta(\text{deviance}) > 2 \times \Delta(\text{parameters})$ *, Model 2* ***is*** *a significant improvement over Model 1.*

Comparing Model 2 and Model 3

We are testing:

H_0 : Model 3 is not a significant improvement over Model 2

H_1 : Model 3 is a significant improvement over Model 2

The difference in the scaled deviances is 5.

This is less than 7.815, the upper 5% point of the χ_3^2 distribution. [½]

So we have insufficient evidence at the 5% significance level to reject the null hypothesis. We conclude that Model 3 is **not** a significant improvement over Model 2. [½]

Alternatively, if we use the AIC to compare models, we find that since $\Delta(\text{deviance}) \not> 2 \times \Delta(\text{parameters})$ *, Model 3 is* ***not*** *a significant improvement over Model 2.*

So Model 2 is the most appropriate in this case. [½]

Chapter 14

Practice Solution 14.5

On page 39, part (ii)(b) the solution has been corrected.

$$\ln f_{post}(\mu) = \ln C + 6\ln\mu - 13\mu \qquad [½]$$

3 Changes to the X Assignments

Assignment X4

Solution X4.8

The solution to part (i) working includes an alternative method to calculate the number of parameters.

(i) ***Workings***

SA + PT + SA • PT = SA ∗ PT $=(\alpha+\beta x)*T_i=\alpha_i+\beta_i x$ $11+(2-1)(10-1)=20$ parameters

(or $2\times 10=20$ parameters)

4 Changes to the Y Assignments

There are no ***non-trivial*** changes to the Y assignments.

Corrections to the 2023 material have been incorporated into the 2024 text and can be found at **ActEd.co.uk/coreStudyMaterial**.

5 Other tuition services

In addition to the CMP you might find the following services helpful with your study.

5.1 Study material

We also offer the following study material in Subject CS1:

- Flashcards
- Revision Notes
- ASET (ActEd Solutions with Exam Technique) and Mini-ASET
- Mock Exam and AMP (Additional Mock Pack).

For further details on ActEd's study materials, please refer to the 2024 *Student Brochure*, which is available from the ActEd website at **ActEd.co.uk**.

5.2 Tutorials

We offer the following (face-to-face and/or online) tutorials in Subject CS1:

- a set of Regular Tutorials (four days or eight half days)
- a Block (or Split Block) Tutorial (lasting four full days)
- a Paper B preparation day
- a five-day Bundle (four days of regular tutorials plus a Paper B preparation day)
- an Online Classroom.

For further details on ActEd's tutorials, please refer to our latest *Tuition Bulletin*, which is available from the ActEd website at **ActEd.co.uk**.

5.3 Marking

You can have your attempts at any of our assignments or mock exams marked by ActEd. When marking your scripts, we aim to provide specific advice to improve your chances of success in the exam and to return your scripts as quickly as possible.

For further details on ActEd's marking services, please refer to the 2024 *Student Brochure*, which is available from the ActEd website at **ActEd.co.uk**.

5.4 Feedback on the study material

ActEd is always pleased to receive feedback from students about any aspect of our study programmes. Please let us know if you have any specific comments (*eg* about certain sections of the notes or particular questions) or general suggestions about how we can improve the study material. We will incorporate as many of your suggestions as we can when we update the course material each year.

If you have any comments on this course, please send them by email to **CS1@bpp.com**.

Solution

The log of the likelihood function is:

$$\log L(\mu_i) = -\sum \frac{y_i}{\mu_i} - \sum \log \mu_i$$

Setting the canonical link function for the exponential distribution to the linear predictor, $g(\mu_i) = 1/\mu_i = \eta_i$, gives:

$$\frac{1}{\mu_i} = \alpha_i \quad \Rightarrow \quad \mu_i = \frac{1}{\alpha_i}$$

This enables us to write the log-likelihood function in terms of α_i:

$$\log L(\alpha_i) = -\sum y_i \alpha_i + \sum \log(\alpha_i)$$

We can now differentiate this with respect to α_i:

$$\frac{\partial}{\partial \alpha_i} \log L(\alpha_i) = -y_i + \frac{1}{\alpha_i}$$

All the terms other than those involving the specific α_i we are looking at disappear.

So the equations satisfied by the MLEs of α_i are:

$$-y_i + \frac{1}{\hat{\alpha}_i} = 0 \quad \Rightarrow \quad \hat{\alpha}_i = \frac{1}{y_i}$$

Hence, the fitted values are:

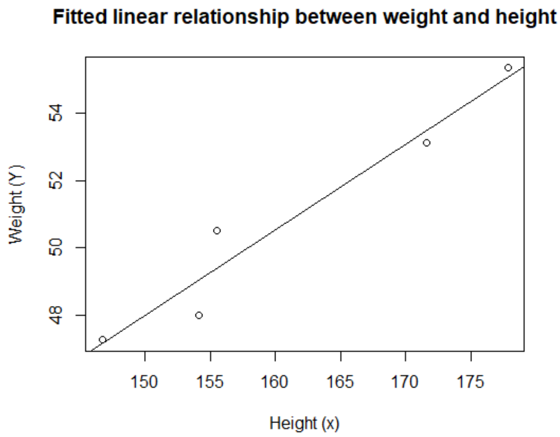
$$\hat{\mu}_i = \frac{1}{\hat{\alpha}_i} = y_i$$

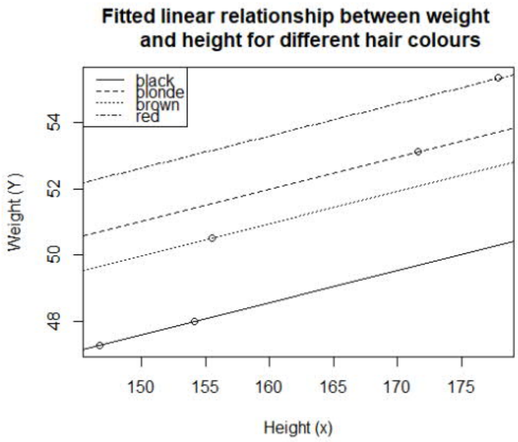
The fitted values, $\hat{\mu}_i$, are equal to the observed values, y_i.

However, a model that fits the data perfectly is not necessarily a satisfactory model.

Suppose we wish to model weight (kg) (Y) and we are considering two possible approaches. For the first model, we use the explanatory variable height (cm) (x) and therefore the linear predictor is $\alpha + \beta x$. A fitted regression line for this model based on a sample of five data points is shown below on the left.

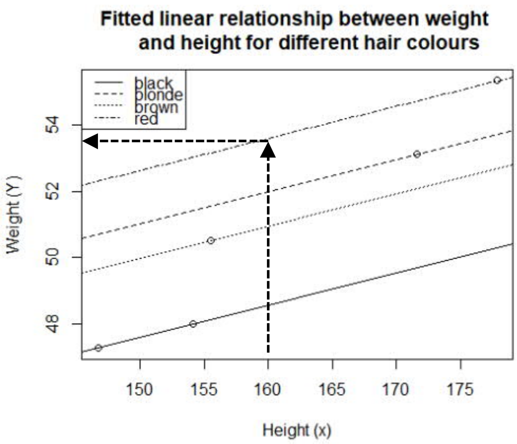
The second candidate model incorporates hair colour as well as height. For the purposes of this analysis, say that there are 4 different possible hair colours: black, blonde, brown and red. This model has a total of 5 parameters, which is the same as the number of data points and so this is a saturated model. The linear predictor is $\alpha_i + \beta x$ for $i \in \{black, blonde, brown, red\}$. A graphical representation of the fitted saturated model is shown below on the right.





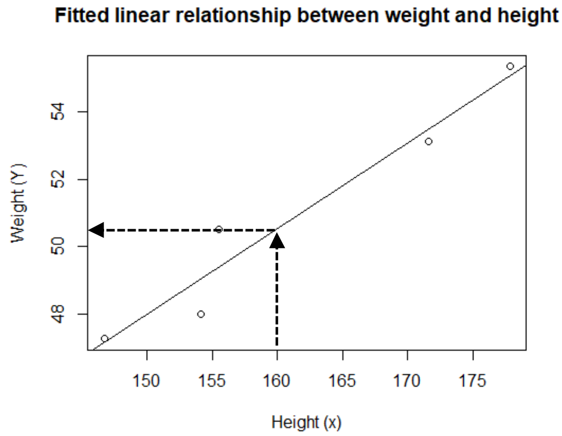
For this saturated model, each line represents the fitted linear relationship between weight and height for the different hair colours. The model is a perfect fit to the data (the fitted values are exactly the observed values). However, the model hasn't necessarily accurately identified the underlying relationships between the covariates and the response, rather it has focused in on the specifics of this particular data set.

We can use this saturated model for predicting the weight of a new data point, though we may not expect it to be particularly accurate. For example, consider a new individual with red hair and a height of 160cm. According to this model, the relationship between height and weight for a red-haired person is estimated by the top line in the graph above. We can see from the graph that the predicted weight in this case is roughly 53.5:



However, we only have one individual with red hair in our sample data set. So, we wouldn't place much confidence in this top line accurately defining the relationship between weight and height for people with red hair.

We can also predict the weight for the same individual using the first model, which only considers height. We can see from the graph that the predicted weight in this case is roughly 50.5:



We would likely be more confident in this second prediction being more accurate based on the data available. The first model appears to have better captured a relationship between weight and height, rather than focusing too much on the specifics of the data used to fit the model. Of course, the overall small sample size in this simple example means we may not place much reliance in this prediction either.

5.3 Scaled deviance (or likelihood ratio)

In order to assess the adequacy of a model for describing a set of data, we can compare the likelihood under this model with the likelihood under the saturated model.

The saturated model uses the same distribution and link function as the current model, but has as many parameters as there are data points. As such it fits the data perfectly. We can then compare our model to the saturated model to see how good a fit it is.

Suppose that L_S and L_M denote the likelihood functions of the saturated and current models, evaluated at their respective optimal parameter values. The likelihood ratio statistic is given by L_S / L_M. If the current model describes the data well then the value of L_M should be close to the value of L_S. If the model is poor then the value of L_M will be much smaller than the value of L_S and the likelihood ratio statistic will be large.

Alternatively, we could examine the natural log of the likelihood ratio statistic:

$$\log \frac{L_S}{L_M} = \ell_S - \ell_M$$

where $\ell_S = \log L_S$ and $\ell_M = \log L_M$.

The *scaled deviance* is defined as twice the difference between the log-likelihood of the model under consideration (known as the current model) and the saturated model.

This page has been left blank so that you can easily put in the replacement pages.