Faculty of Science, Technology, Engineering and Mathematics
M347 Mathematical statistics

M347

TMA 04

Covers Block 4 **See module website for the cut-off date.**

Submitting your assignment

You can submit your TMA either by post or electronically using the online TMA/EMA service. Please read the guidance in the 'Assessment' area of the M347 website.

This TMA

This TMA covers Block 4 of M347. You will probably find it best to answer the questions associated with each unit soon after completing that unit, rather than waiting until you have completed the block.

Each TMA is marked out of 25, with the marks allocated to each part of each question indicated in brackets in the margin. Your overall score for the TMA will be the sum of your marks for all questions in the TMA, which is then converted into a percentage score. Your tutor's reply will give you feedback on how well you answer the TMA. Being a formative assessment, your mark on this TMA counts only towards achieving the required threshold for completion of assignments to be able to pass the module, but does not contribute to your final mark for the module.

Your work should only include your answers to the questions: do not include material which was not asked for in the questions. You should, however, include your working. If you don't, and you make a mistake, then your tutor will not be able to provide any feedback about where you went wrong. Leaving out your working may also cost you marks if it is specifically asked for.

WEB 05349 8
4.1

This assignment covers Block 4.

Question 1 (Unit 11) – 9 marks

Consider the problem of using the linear regression model with one explanatory variable to estimate the value of $\alpha + \beta x_0$. In both classical analysis (Subsection 3.11) and Bayesian analysis with the improper prior used in Section 5, the point estimator of $\alpha + \beta x_0$ is $\widehat{\alpha} + \widehat{\beta} x_0$. Also write, as usual, $S^2 = R/(n-2)$.

(a) In the classical case (Subsection 3.11), the width of the confidence interval for $\alpha + \beta x_0$ is based on the variance term which arises in the following distribution:

$$\widehat{\alpha} + \widehat{\beta} x_0 \sim N\left(\alpha + \beta x_0, \sigma^2 \left\{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right\}\right).$$

(i) Explain how this distribution arises, and specify the statuses of the parameters and their estimated values, that is, which are fixed quantities and which are random variables. *[3]*

(ii) The width of the confidence interval for $\alpha + \beta x_0$ also depends on S and a percentile of the $t(n-2)$ distribution (and not the normal distribution). Explain briefly how this arises. *[2]*

(b) In the Bayesian case with improper prior $f(\alpha, \beta, \tau) \propto \tau^{-1}$, the width of the credible interval for $\alpha + \beta x_0$ can be shown to be based on the variance term which arises in the following distribution:

$$\alpha + \beta x_0 \,|\, \boldsymbol{y} \sim t\left(n-2; \widehat{\alpha} + \widehat{\beta} x_0, S^2 \left\{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right\}\right).$$

Explain how this distribution arises, and specify the statuses of the parameters and their estimated values, that is, which are fixed quantities and which are random variables. *[4]*

Question 2 (Unit 13) – 16 marks

In the agricultural experiment introduced in Subsection 1.3 and forming a running example through Sections 1 and 2, suppose that it is of interest to estimate the difference in effects between Treatment B and Treatment C, that is, $\alpha_2 - \alpha_3$. Write $\widehat{\alpha_2 - \alpha_3}$ for its estimator, and write this as a linear function of the responses:

$$\widehat{\alpha_2 - \alpha_3} = a_1 Y_1 + a_2 Y_2 + a_3 Y_3 + a_4 Y_4 + a_5 Y_5 + a_6 Y_6,$$

for some constants $a_1, a_2, \ldots, a_6$. In part (a) of Exercise 13.8 in Subsection 1.5, it was shown that the expectation of the right-hand side, and hence of what is now being called $\widehat{\alpha_2 - \alpha_3}$, is

$$\begin{aligned} E(\widehat{\alpha_2 - \alpha_3}) =& (a_1 + a_4)\alpha_1 + (a_3 + a_5)\alpha_2 + (a_2 + a_6)\alpha_3 \\ &+ (a_1 + a_2)\beta_1 + (a_3 + a_4)\beta_2 + (a_5 + a_6)\beta_3. \end{aligned}$$

(a) What are the requirements on $a_1, a_2, \ldots, a_6$ so that $\widehat{\alpha_2 - \alpha_3}$ is an unbiased estimator of $\alpha_2 - \alpha_3$? [3]

(b) Write $a_1 = c$. Solve the requirements in part (a) in terms of c, and hence show that, for any value of c,

$$\widehat{\alpha_2 - \alpha_3} = cY_1 - cY_2 + cY_3 - cY_4 + (1 - c)Y_5 - (1 - c)Y_6$$

is an unbiased estimator of $\alpha_2 - \alpha_3$. [4]

(c) Explain why $Y_5 - Y_6$ is an unbiased estimator of $\alpha_2 - \alpha_3$. By reference to the experimental design in Figure 13.2 in Subsection 1.3, does this estimator make sense as an estimator of the difference between the effects of Treatment B and Treatment C, and what disadvantage might it have? [3]

(d) Apply the general formula for uncorrelated random variables,

$$V\left(\sum_{i=1}^{d} a_i X_i\right) = \sum_{i=1}^{d} a_i^2\, V(X_i),$$

to the unbiased estimator that was obtained in part (b) to show that

$$V(\widehat{\alpha_2 - \alpha_3}) = 2(3c^2 - 2c + 1)\sigma^2.$$ [3]

(e) Hence find the value of c that minimises $V(\widehat{\alpha_2 - \alpha_3})$, and give the formula for the estimator corresponding to this value. [3]