Examination in the School of Mathematical Sciences
Semester 1, 2021

STATS 3001 Statistical Modelling III
STATS 4101 Statistical Modelling - Honours

Instructions:

- Refer to the Instructions page in the Exam module for instructions.

1. **Submission instructions**

   In the exam module you should find a section with a link to the quiz called

   `Part A: Quiz`

   A single attempt is allowed for each question. The quiz will be available for the entirety of the exam.

   [60 marks]

2. **Submission instructions**

Your answers may be hand-written and scanned as a pdf. Your pdf can then be uploaded in the section of the exam module that states

`Part 2: Exam Question 2`

Consider n independent random variables $Y_1, Y_2, \ldots, Y_n$ such that

$$\mathrm{E}[Y_i] = \mu$$

and

$$\mathrm{var}(Y_i) = \sigma_i^2$$

Let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

(a) Calculate $\mathrm{E}[\bar{Y}]$
(b) Calculate $\mathrm{var}[\bar{Y}]$
(c) We will find a new estimator of μ using the generalised least squares framework. First write the model as

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

give the form of X, $\boldsymbol{\beta}$, $\mathrm{E}[\boldsymbol{\epsilon}]$, and $\mathrm{Var}[\boldsymbol{\epsilon}]$.
(d) Calculate $\hat{\boldsymbol{\beta}}$, and hence $\hat{\mu}$.
(e) Find $\mathrm{E}[\hat{\mu}]$
(f) Find $\mathrm{Var}[\hat{\mu}]$

[30 marks]

3. In a certain experiment, the lung weights of two strains of mice were compared. One strain of mice was normal, `C57`, and the other was a mutant strain, `mdx`, that develops a condition similar to muscular dystrophy in humans.

   An analysis of the dataset is given in `Q3_mice-analysis.html`. Please read the analysis and then answer the questions in the quiz.

   Submission instructions

   The analysis is given in

   `Part C: Mice Analysis`

   in the exam module.

   As well, there is a link to a quiz with questions about the interpretation of this analysis called

   `Part C: Mice Analysis Quiz`

   A single attempt is allowed for each question. The quiz will be available for the entirety of the exam.

   [30 marks]

4. **Submission instructions**

Your answers may be hand-written and scanned as a pdf. Your pdf can then be uploaded in the section of the exam module that states

`Part D: Exam Question 4`

Consider the ridge regression objective function

$$Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2.$$

(a) Show that the vector of partial derivatives

$$\left[\frac{\partial Q}{\partial \beta_j}\right] = 2\boldsymbol{\beta}^T(X^TX + \lambda I) - 2\boldsymbol{y}^TX.$$

(b) Show for fixed $\lambda \geq 0$ that the ridge regression estimator is given by

$$\hat{\boldsymbol{\beta}}_\lambda = (X^TX + \lambda I)^{-1}X^T\boldsymbol{y}.$$

(c) Assuming

$$E(\boldsymbol{Y}) = X\boldsymbol{\beta} \text{ and } \mathrm{Var}(\boldsymbol{Y}) = \sigma^2 I$$

find $E(\hat{\boldsymbol{\beta}}_\lambda)$ and $\mathrm{Var}(\hat{\boldsymbol{\beta}}_\lambda)$.

(d) Describe the behaviour of the ridge regression estimate as $\lambda \to +\infty$.

(e) Describe the behaviour of the ridge regression estimate as $\lambda \to 0$.

(f) Give an example of how a suitable value for λ can be obtained in practice.

(g) Explain the role of centring the variables in ridge regression.

[30 marks]

FINAL PAGE

End of question.