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]

Solutions:
See Q1 questions on myuni

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]

Solutions:

(a)

$$\begin{aligned}
E[\bar{Y}] &= E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[Y_i] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu \\
&= \mu
\end{aligned}$$

(b)

$$\begin{aligned}
var[\bar{Y}] &= var\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} var[Y_i] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2
\end{aligned}$$

(c)

$$\begin{aligned}
X &= \mathbf{1}_{n\times 1} \\
\boldsymbol{\beta} &= [\mu] \\
E[\boldsymbol{\epsilon}] &= \mathbf{0} \\
V = var[\boldsymbol{\epsilon}] &= \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}
\end{aligned}$$

(d)

$$V^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_n^2 \end{bmatrix}$$

$$\begin{aligned}
\hat{\beta}_{GLS} &= (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbf{Y} \\
&= \left([11\dots1] \begin{bmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_n^2 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} [11\dots1] \begin{bmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_n^2 \end{bmatrix} \begin{bmatrix} y \\ y \\ \vdots \\ y \end{bmatrix} \\
&= \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}
\end{aligned}$$

Hence

$$\hat{\mu} = \hat{\beta}_{GLS} = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}$$

(e)

$$\begin{aligned}
E[\hat{\mu}] &= E\left[\left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^{n} \frac{Y_i}{\sigma_i^2} \right] \\
&= \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^{n} \frac{E[Y_i]}{\sigma_i^2} \\
&= \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^{n} \frac{\mu}{\sigma_i^2} \\
&= \mu \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \\
&= \mu.
\end{aligned}$$

(f)

$$
\begin{aligned}
var[\hat{\mu}] = var\left[\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-1}\sum_{i=1}^{n}\frac{Y_i}{\sigma_i^2}\right] &\\
&= \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-2}\sum_{i=1}^{n}\frac{var[Y_i]}{(\sigma_i^2)^2}\\
&= \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-2}\sum_{i=1}^{n}\frac{\sigma_i^2}{(\sigma_i^2)^2}\\
&= \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-2}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\\
&= \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-1}
\end{aligned}
$$

---

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]

Solutions:
See Q3_mice.html

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

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

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

(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]

Solutions:

(a)

Using the fact

$$\left[\frac{\partial \|\boldsymbol{u}\|^2}{\partial \boldsymbol{u}}\right] = 2\boldsymbol{u}^T$$

it follows from the chain rule that

$$\left[\frac{\partial \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}}\right] = -2(\boldsymbol{y} - X\boldsymbol{\beta})^T X = 2\boldsymbol{\beta}^T(X^T X) - 2\boldsymbol{y}^T X.$$

Also,

$$\left[\frac{\partial \|\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}}\right] = 2\boldsymbol{\beta}^T$$

Hence it follows that

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

(b)

Observe

$$\hat{\boldsymbol{\beta}}_\lambda$$

is the solution to

$$\left[\frac{\partial Q}{\partial \beta_j}\right] = \mathbf{0}^T$$

and

$$\begin{aligned}
\left[\frac{\partial Q}{\partial \beta_j}\right] = \mathbf{0}^T
&\Rightarrow 2\boldsymbol{\beta}^T(X^T X + \lambda I) - 2\boldsymbol{y}^T X = \mathbf{0}^T \\
&\Rightarrow (X^T X + \lambda I)\boldsymbol{\beta} = X^T \boldsymbol{y} \\
&\Rightarrow \boldsymbol{\beta} = (X^T X + \lambda I)^{-1} X^T \boldsymbol{y}
\end{aligned}$$

as required.

(c)

$$E(\hat{\boldsymbol{\beta}}_\lambda) = (X^T X + \lambda I)^{-1} X^T X \boldsymbol{\beta}.$$
$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_\lambda) = \sigma^2 (X^T X + \lambda I)^{-1} (X^T X)(X^T X + \lambda I)^{-1}.$$

(d)

$$\hat{\boldsymbol{\beta}}_\lambda \to \mathbf{0} \text{ as } \lambda \to +\infty.$$

(e)

$$\hat{\boldsymbol{\beta}}_\lambda \to \hat{\boldsymbol{\beta}}_{OLS} \text{ as } \lambda \to 0.$$

(f)

In practice, λ can be chosen to minimize the cross validated error.

(g)

Without centring, the ridge penalty shrinks the fitted values toward 0. In general it is more desirable to shrink the fitted values toward $\bar{y}$. This can be implemented by centring the response, y, and all of the predictor variables, $\boldsymbol{x}$, and then applying ridge regression to the centred data with no intercept.

FINAL PAGE

End of question.