STATS 2107
Statistical Modelling and Inference II
Tutorial 4

Sharon Lee, Matt Ryan

Semester 2 2022

1. (a) Consider regression data
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$
and let
$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2.$$
Prove that
$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}.$$

   (b) Consider independent random variables $Y_1, Y_2, \ldots, Y_n$ with
$$E(Y_i) = \beta_0 + \beta_1 x_i \quad \text{and} \quad \mathrm{Var}(Y_i) = \sigma^2.$$
Let
$$\hat{E}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i).$$
Prove that
$$E\left[\sum_{i=1}^{n}\{Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\}^2\right] = (n-2)\sigma^2.$$
Hence deduce that S_e^2 is an unbiased estimator for σ^2.

2. Suppose X is an $n \times p$ matrix with linearly independent columns and let
$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top.$$
In lectures, we have stated that the matrices $\boldsymbol{H}$ and $(\boldsymbol{I} - \boldsymbol{H})$ are symmetric and idempotent. We will prove these properties here.
   (a) Show that $\boldsymbol{H}$ is symmetric, *i.e.*, $\boldsymbol{H} = \boldsymbol{H}^\top$.
   (b) Show that $\boldsymbol{H}$ is idempotent, that is, $\boldsymbol{H}^2 = \boldsymbol{H}$.
   (c) Show that $(\boldsymbol{I} - \boldsymbol{H})$ is symmetric and idempotent, where $\boldsymbol{I}$ is the $n \times n$ identity matrix.

3. Consider the multiple regression model
$$M: \ \boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$
where $Y_i \sim N(\eta_i, \sigma^2)$ independently for $i = 1, 2, \ldots, n$ and $E[\boldsymbol{Y}] = \boldsymbol{\eta} = X\boldsymbol{\beta}$. The vector of residuals is defined by $\hat{\boldsymbol{E}} = \boldsymbol{Y} - X\hat{\boldsymbol{\beta}}$. In lectures, we have stated some properties of $\hat{\boldsymbol{E}}$. We will look at these here.
   (a) Prove that $\hat{\boldsymbol{E}} = (I - H)\boldsymbol{Y}$, where $H = X(X^T X)^{-1} X^T$.

(b) Prove that $E(\hat{\boldsymbol{E}}) = \mathbf{0}$ (and hence $E(\hat{E}_i) = 0$ for $i = 1, 2, \ldots, n$).

(c) Prove that: $\text{Var}(\hat{E}_i) = \sigma^2(1 - h_{ii})$, where h_{ii} is the (i, i)th element of H. **Hint: Let $(I - H)$ have rows $\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \ldots, \boldsymbol{a}_n^T$.**

4. Linearise the following equations:

(a) $Y = \alpha\beta^x$

(b) $Y = \alpha e^{\frac{\beta}{x}}$

(c) $Y = \alpha + \frac{\beta}{x}$

(d) $Y = \frac{\alpha}{\beta + x}$

(e) $Y = \alpha + \beta x^n$

(f) $Y = \frac{1}{\alpha + \beta e^{-x}}$

(g) $Y = e^{-\alpha x_1 e^{-\frac{\beta}{x_2}}}$

The following question is optional:

5. In the proof of Theorem 11, we have used the result that if $\boldsymbol{X}$ is a random variable with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, then

$$E(\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}) = tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}.$$

Prove this result.