Imperial College London

Problem Sheet 6 Solutions

MATH50011
Statistical Modelling 1

Week 6

Lecture 11 (Introduction to Linear Models)

1. Let $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1, \dots, n$ where $x_i = 0, 1$ and $\epsilon_1, \dots, \epsilon_n$ are iid $N(0, \sigma^2)$ random variables where $\sigma^2 > 0$ is known. We can think of the covariate x_i as defining two groups receiving a different treatment, as in a clinical trial.

   (a) What is the interpretation of β_0, β_1 and $\beta_0 + \beta_1$ in this model?

   (b) Based on your answer to part (a), propose estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ in terms of particular sample averages.

   (c) What is the distribution of $\hat{\beta}_1$?

   (d) Describe how to construct a 95% confidence interval for β_1 using the distribution identified in the previous question.

Solution.

(a) Since $E(Y_i) = \beta_0 + \beta_1 x_i$ we have that β_0 is the mean of Y_i when $x_i = 0$ and β_1 is the difference in $E(Y_i)$ when $x_i = 1$ and $x_i = 0$.

(b) Let $\bar{Y}_k = \frac{1}{n_k} \sum_{i:x_i=k} Y_i$ for $k = 0, 1$ with n_k the number of individuals having $x_i = k$. In particular, $\bar{Y}_0$ is the sample mean of the Y_is having $x_i = 0$ and $\bar{Y}_1$ is the sample mean of the Y_is having $x_i = 1$. Then $\hat{\beta}_0 = \bar{Y}_0$ and $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$ are reasonable candidate estimators, in view of part (a).

(c) Since the Y_is are independent, so too are $\bar{Y}_1$ and $\bar{Y}_0$. Hence $\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$ has a normal distribution with mean $E(\bar{Y}_1 - \bar{Y}_0) = \beta_1$ and variance $Var(\bar{Y}_1 - \bar{Y}_0) = \sigma^2(n_0^{-1} + n_1^{-1})$.

(d) In this case, we know that

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2(n_0^{-1} + n_1^{-1})}} \sim N(0, 1)$$

so we can use the familiar confidence interval construction with upper/lower limits given by

$$\hat{\beta}_1 \pm 1.96 \times \sqrt{\sigma^2(n_0^{-1} + n_1^{-1})}.$$

2. Which of the following matrices is positive definite? positive semidefinite?

$$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Solution. The first matrix is positive definite, the second positive semidefinite. The third matrix is not positive semidefinite.

3. Show that

$$Cov(\mathbf{A}X, \mathbf{B}Y) = \mathbf{A}Cov(X, Y)\mathbf{B}^T$$

where **A** and **B** are deterministic matrices of suitable dimensions. What does "suitable dimension" mean in this case?

   (a) Show that $Cov(X)$ is positive semidefinite.

   (b) Find an example where $Cov(X)$ is not positive definite.

   (c) Find an example where $Cov(X)$ is positive definite.

Solution.

(a)

$$\begin{aligned} Cov(AX, BY) =& E[(AX - EAX)(BY - EBY)^T] \\ =& E[(AX - AEX)(BY - BEY)^T] \\ =& AE[(X - EX)(B(Y - EY))^T] \\ =& AE[(X - EX)(Y - EY)^T B^T] \\ =& AE[(X - EX)(Y - EY)^T]B^T \\ =& ACov(X, Y)B^T \end{aligned}$$

Suitable dimension means that A must have n columns and B m columns.

(b) For any vector $c \in \mathbb{R}^n$,

$$c^T Cov(X)c = E[c^T(X - EX)(X - EX)^T c] = E[\{c^T(X - EX)\}^2] \geq 0$$

(note that $c^T(X - EX)$ is one-dimensional)

(c) Let $X = c \in \mathbb{R}$, i.e. the one-dimensional random vector that is constant. Then $Cov(X) = Var(c) = 0$. $Cov(X)$ is not positive definite since e.g. $1 \cdot Cov(X) \cdot 1 = 0$.

(d) Let Y be a random variable with $\infty > VarY > 0$. let $X = (Y)$, i.e. X is a one-dimensional random vector. Then for any $c \in \mathbb{R} \setminus \{0\}$,

$$cCov(X)c = cVar(Y)c = c^2 Var(Y) > 0.$$

4. Suppose $X, Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$ independent. Let $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$, $Y = (Y_1, \dots, Y_n)^T$. Let $Z = \sqrt{\rho}X\mathbf{1} + \sqrt{1 - \rho}Y$ for some $\rho \in [0, 1]$.

   Find $Cov(Z)$ using rules for manipulation of Cov.

Solution.

$$
\begin{aligned}
Cov(Z) =& Cov(\sqrt{\rho}X\mathbf{1} + \sqrt{1-\rho}Y, \sqrt{\rho}X\mathbf{1} + \sqrt{1-\rho}Y) \\
=& Cov(\sqrt{\rho}X\mathbf{1}, \sqrt{\rho}X\mathbf{1}) + 2Cov(\sqrt{\rho}X\mathbf{1}, \sqrt{1-\rho}Y) + Cov(\sqrt{1-\rho}Y, \sqrt{1-\rho}Y) \\
=& \rho\mathbf{1}Cov(X,X)\mathbf{1}^T + 0 + (1-\rho)Cov(Y,Y) \\
=& \rho\sigma^2\mathbf{1}\mathbf{1}^T + (1-\rho)\sigma^2 I \\
=& \sigma^2(\rho\mathbf{1}\mathbf{1}^T + (1-\rho)I) = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}
\end{aligned}
$$

Lecture 12 (Linear Models)

5. For a simple linear regression model, $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$ for $i = 1, \dots, n$ where $E(\epsilon_i) = 0$ and $Cov(\epsilon) = \sigma^2 I_n$.

   (a) Derive the least squares estimators of β_1 and β_2 based on the above sample.

   (b) How do the least squares estimators change if they are computed in terms of $Z_i = Y_i - \bar{Y}$ and $w_i = x_i - \bar{x}$ instead?

   (c) What is the expected value of the least squares estimators?

   (d) Using properties of covariances for random vectors, derive the covariance matrix of the least squares estimators $(\hat{\beta}_1, \hat{\beta}_2)^T$.

Solution. See the lecture notes for details of this exercise.

6. In a study on childhood development, the following data about the height and weight of 11 children was collected.

Height	135	146	153	154	139	131	149	137	143	146	141
Weight	26	33	55	50	32	25	44	31	36	35	28

Formulate a linear regression model with response variable height and explanatory variable weight.

Compute the least squares estimates and sketch both the data and the estimated regression curve.

Solution. Model: $Y_i = \beta_1 + x_i\beta_2 + \epsilon_i$ with ϵ_i iid N(0,σ^2), where, for the ith individual, Y_i =height, x_i=weight.

Letting $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{10} \end{pmatrix}$ we get

$$\begin{aligned} \hat{\beta} =& (X^T X)^{-1} X^T Y = \begin{pmatrix} 11 & 395 \\ 395 & 15141 \end{pmatrix}^{-1} \begin{pmatrix} 1574 \\ 57175 \end{pmatrix} \\ =& \begin{pmatrix} 1.438 & -0.038 \\ -0.038 & 0.001 \end{pmatrix} \begin{pmatrix} 1574 \\ 57175 \end{pmatrix} = \begin{pmatrix} 118.5 \\ 0.68 \end{pmatrix} \end{aligned}$$

7. In the Forbes and Mammals data examples in Chapter 9 of the notes, we transform variables by taking the natural logarithm. This impacts our interpretation of the coefficients in our linear model.

   (a) Consider a simple linear model $E(Y) = \beta_0 + \beta_1 x$. Interpret β_1 by comparing two groups that differ in x by 1 unit.

   (b) Consider a simple linear model $E(\log Y) = \beta_0 + \beta_1 x$. Interpret β_1 by comparing two groups that differ in x by 1 unit.

   (c) Consider a simple linear model $E(\log Y) = \beta_0 + \beta_1 \log x$. Interpret β_1 by comparing two groups that differ in x by 1 unit.
   (Hint: $\exp(E(\log Y))$ is called the *geometric mean* of Y.)

Solution.

(a) We will use the notation $E(Y|X = x)$, though we are not treating the x as realisations of a random variable at this time. For this model, we have

$$E(Y|X = x + 1) - E(Y|X = x) = \beta_0 + \beta_1(x + 1) - (\beta_0 + \beta_1 x) = \beta_1.$$

Hence β_1 is the difference in the mean of the response Y associated with a unit difference in the predictor x.

(b) For the log-transformed outcome, we have

$$E(\log Y|X = x + 1) - E(\log Y|X = x) = \beta_0 + \beta_1(x + 1) - (\beta_0 + \beta_1 x) = \beta_1.$$

Hence $e^{E(\log Y|X=x+1)-E(\log Y|X=x)} = GM(Y|X = x + 1)/GM(Y|X = x) = e^{\beta_1}$, where $GM(Y|X = x)$ is the geometric mean of Y for a given value of the predictor x. Hence, e^{β_1} denotes the ratio of geometric means associated with a unit difference in the predictor x.

(c) By similar logic to part (b), we see that

$$\begin{aligned} GM(Y|X = x + 1) &= \exp E(\log Y|X = x + 1) = \exp(\beta_0 + \beta_1 \log(x + 1)) = e^{\beta_0}(x + 1)^{\beta_1} \\ GM(Y|X = x) &= \exp E(\log Y|X = x) = \exp(\beta_0 + \beta_1 \log x) = e^{\beta_0} x^{\beta_1} \end{aligned}$$

The interpretation of β_1 for a unit increase in x is not meaningful in this context. However, if we instead multiply x by $e \equiv exp(1)$, we find that

$$GM(Y|X = ex)/GM(Y|X = x) = \exp(\beta_1)$$

so that we can interpret $\exp(\beta_1)$ as the ratio in geometric mean outcomes associated with an e-fold difference in the predictor.

8. Let $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i$ for $i = 1, 2, 3, 4$ and $x_i = i$. Write the above polynomial model in matrix form such that $Y = X\beta + \epsilon$.

Solution. We have

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$