Imperial College London

Problem Sheet 6

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.

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

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

3. Show that

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

   where $\mathbf{A}$ and $\mathbf{B}$ 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.

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.

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 $(\hat{\beta}_1, \hat{\beta}_2)^T$.

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.

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.)

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 $Y = X\beta + \epsilon$.