Imperial College London

Problem Sheet 6 Solutions

MATH50011
Statistical Modelling 1

Week 8

Lecture 13: Properties of Least Squares

1. Consider an "error in the variables" model in which there is a true relationship between Y and w given by $Y_i = \beta_1 + \beta_2 w_i + \epsilon_i$ with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$ independent for $i = 1, \dots, n$. Suppose that rather than observe w_i, we have X_i, an imprecise measurement given by $X_i = \alpha_1 + \alpha_2 w_i + \delta_i$ with $E(\delta_i) = 0$ and $Var(\delta_i) = \tau^2$ which are independent and independent of the ϵ_is. The parameters α_1 and α_2 are unknown.

   We fit a regression model based on $E(Y_i|X_i) = \gamma_1 + \gamma_2 X_i$. Show that the least squares estimator $\hat{\gamma}$ is biased for estimating β, even when $\alpha_1 = 0$ and $\alpha_2 = 1$.

   Solution. First, we compute the least square estimator $\hat{\gamma}$. As mentioned in the question we observe $X_1, \dots, X_n$. So let $X_1, \dots, X_n$ be given. Then, the regression model that we fit becomes $E[Y_i] = \gamma_1 + \gamma_2 x_i$. Thus, as usual the least squares estimator $\hat{\gamma}$ is $\hat{\gamma} = (X^T X)^{-1} X^T Y$. Notice that we have both $E[Y_i] = \beta_1 + \beta_2 w_i$ and $E[Y_i] = \gamma_1 + \gamma_2 x_i$. However, only the second model is observed. This question ask if the least squares estimator of the model we observe is an unbiased estimator for the parameters of the model that we do not observe.

   By $\hat{\gamma} = (X^T X)^{-1} X^T Y$ we have that $E[\hat{\gamma}] = (X^T X)^{-1} X^T E[Y] = (X^T X)^{-1} X^T W\beta$, where X and W are the respective design matrices. Hence, $\hat{\gamma}$ is unbiased if $(X^T X)^{-1} X^T W = I$. This can only happen if $X^T W = X^T X$. However, using that $x_i = \alpha_1 + \alpha_2 w_i + \delta_i$ we have

$$X^T W = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n w_i & \sum_{i=1}^n x_i w_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n (\alpha_1 + \alpha_2 w_i + \delta_i) \\ \sum_{i=1}^n w_i & \sum_{i=1}^n w_i(\alpha_1 + \alpha_2 w_i + \delta_i) \end{pmatrix} \neq \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

   except in very unusual situations where sums involving δ_i equal zero, even if $\alpha_1 = 0$ and $\alpha_2 = 1$, because we need to have $\sum_{i=1}^n w_i = \sum_{i=1}^n \alpha_1 + \alpha_2 w_i + \delta_i$.

2. Consider a linear model with a p-dimensional parameter vector β. For a deterministic vector $c \in \mathbb{R}^p$ we know that $c^T \hat{\beta}$, where $\hat{\beta}$ is the least squares estimator, is a linear unbiased estimator for $c^T \beta$.

   In linear models of your choice and for vectors c of your choice, give examples for other unbiased linear estimators for $c^T \beta$. Quantify the loss in precision (measured by increase in MSE) of some of those estimators (what assumptions do you need to make for this?).

Solution. We consider simple linear regression $E(Y) = \beta_0 + \beta_1 x$ based on (Y_i, x_i) for $i = 1, \dots, n$ and saw that the slopes

$$\frac{Y_i - Y_j}{x_i - x_j}, \quad i \neq j$$

are linear unbiased estimators of β_1. We can compare the variance of these to the least squares estimator $\hat{\beta}_1$. Assuming the errors in the linear model are uncorrelated with variance σ^2:

$$Var\left(\frac{Y_i - Y_j}{x_i - x_j}\right) = \frac{2\sigma^2}{(x_i - x_j)^2}$$

The Gauss-Markov theorem tells us that this variance will be at least $\sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2$.

Many variations on this theme are possible. Consider, e.g., dropping a single observation i from the data and computing the least squares estimator based on the remaining $n - 1$ samples.

3. (a) Compute the projection matrix onto $span((1,1,1,1)^T, (0,0,1,1)^T)$ in $\mathbb{R}^4$.
   (b) Compute the projection matrix onto $span((1,0,0)^T, (1,1,1)^T, (0,0,2)^T)$ in $\mathbb{R}^3$.
   (c) Compute the projection matrix onto $span((1,\dots,1)^T)$ in $\mathbb{R}^n$.
   (d) Compute the projection matrix onto $span((0,\dots,0)^T)$ in $\mathbb{R}^n$.

   What is the rank of these matrices? What are their eigenvalues (including their multiplicities)?

Solution. *(This is not the full solution, only a sketch.)* Using results and methods from the lectures, one obtains the following solutions:

(a)

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

(b) The 3×3 identity matrix.

(c) The $n \times n$ matrix with all elements equal to $\frac{1}{n}$.

(d) The $n \times n$ matrix with all elements equal to 0.

The rank of these matrices is equal to their trace.

Alternatively, one could determine the space these matrices project onto to get the rank.

The eigenvalues are also immediate (1 (twice) and 0 (twice), 1 (three times), 1 (once) and 0 (n-1 times), 0 (n times)).

Lecture 14: Fitted Values, Residuals

4. Suppose we are interested in the relationship between the height Y of tomato plants one month after being potted in soil a, b, or c.

Plants $1, \dots, n$ are potted in soil a. Plants $n+1, \dots, 2n$ are potted in soil b. Plants $2n+1, \dots, 3n$ are potted in soil c.

(a) Write a linear model for this setting of the form

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

where x_{i1} and x_{i2} are binary variables taking on the values 0 and 1.

(b) Express the parameter vector for your model in terms of the mean heights for a tomato plant grown in soil a and/or b and/or c.

(c) Find the least squares estimate of β for your model.

(d) Express the fitted values for your model in terms of appropriate sample means based on the Y_i.

(Harder: how many solutions are there to (a)? How do the fitted values change in each case?)

Solution.

(a) Let $x_{i1} = 1$ for $i = 1, \dots, n$ and $x_{i1} = 0$ otherwise. Similarly, let $x_{i2} = 1$ for $i = n+1, \dots, 2n$. Hence, x_{i1} is an indicator that the ith plant was potted in soil a and x_{i2} is an indicator that the ith plant was potted in soil b. If $x_{i1} = x_{i2} = 0$, as is the case for $i = 2n+1, \dots, 3n$, then the ith plant was potted in soil c.

This can be expressed in matrix form as

$$E(Y) = X\beta = \begin{pmatrix} \vec{1}_n & \vec{1}_n & \vec{0}_n \\ \vec{1}_n & \vec{0}_n & \vec{1}_n \\ \vec{1}_n & \vec{0}_n & \vec{0}_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

where $\vec{0}_n$ is a vector of n 0s and $\vec{1}_n$ is a vector of n 1s.

(b) We have $\mu_a = E(Y|\text{soil } a) = \beta_0 + \beta_1$, $\mu_b = E(Y|\text{soil } b) = \beta_0 + \beta_2$, and $\mu_c = E(Y|\text{soil } c) = \beta_0$.

(c) Evaluating $(X^T X)^{-1} X^T Y$, we find $\hat{\beta} = (\hat{Y}_c, \hat{Y}_a - \hat{Y}_c, \hat{Y}_b - \hat{Y}_c)$ where notation such as $\hat{Y}_a$ denotes the respective group mean. These calculations can be simplified by changing to a parametrisation in terms of

$$\begin{pmatrix} \vec{1}_n & \vec{0}_n & \vec{0}_n \\ \vec{0}_n & \vec{1}_n & \vec{0}_n \\ \vec{0}_n & \vec{0}_n & \vec{1}_n \end{pmatrix} \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix} = X \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \beta = (XH)(H^{-1}\beta) = X\beta.$$

See the argument below (d) for details.

(d) The fitted values of the model are a vector of n copies of $\bar{Y}_a$, followed by n copies of $\bar{Y}_b$, and n copies of $\bar{Y}_c$.

Note that soil a or soil b could have been made to correspond to $x_{i1} = x_{i2} = 0$, and so on. In all instances, we can relate these reparametrisations by considering $X\beta = (XH)(H^{-1}\beta) = W\gamma$ for some nonsingular H. Then

$$\hat{\gamma} = (W^T W)^{-1} W^T Y = H^{-1}(X^T X)^{-1} X^T Y = H^{-1}\hat{\beta}$$

and hence the fitted values are equal:

$$W(W^TW)^{-1}W^TY = X(X^TX)^{-1}X^TY.$$

To summarise, regardless of parametrisation, we wind up projecting Y onto the same space and the projection is unique. This leads to equality of the fitted values (and, indeed, the projection matrices).

5. Consider a simple linear regression model, $EY_i = \beta_1 + \beta_2 x_i$ ($i = 1, \dots, n$), where β_1 and β_2 are unknown, the second order assumptions hold and $Var(Y_i) = \sigma^2 > 0$. Suppose moreover that at least two of the x_i are distinct. Let $e = (e_1, \dots, e_n)^T$ be the vector of residuals. Compute $Cov(e)$.

Solution.

$$\begin{aligned} Cov(e) &= Cov((I-P)Y) = (I-P)Cov(Y)(I-P)^T = (I-P)\sigma^2 I(I-P)^T \\ &= (I-P)^2\sigma^2 = (I-P)\sigma^2 \end{aligned}$$

Hence it suffices to derive P.

$$\begin{aligned} P &= (1,x)\left((1,x)^T(1,x)\right)^{-1}(1,x)^T \\ &= (1,x)\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}(1,x)^T \\ &= (1,x)\frac{1}{n\sum x_i^2 - (\sum x_i)^2}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}(1,x)^T \end{aligned}$$

Note that $S_{xx} = \sum x_i^2 - \frac{1}{n}(\sum x_i)^2$. Hence,

$$\begin{aligned} P &= \frac{1}{nS_{xx}}(1,x)\begin{pmatrix} \sum x_i^2 - x_j\sum x_i \\ -\sum x_i + x_j n \end{pmatrix}^T_{j=1\dots,n} \\ &= \frac{1}{nS_{xx}}\left(\sum x_i^2 \underbrace{-x_j\sum x_i + (-\sum x_i + x_j n)x_\mu}_{=-(\sum x_i - x_j n)(x_\mu - \sum x_i/n) + (\sum x_i)^2/n}\right)_{\mu,j=1\dots,n} \\ &= \left(\frac{1}{n} + \frac{(x_j-\bar{x})(x_\mu-\bar{x})}{S_{xx}}\right)_{\mu,j=1\dots,n} \end{aligned}$$

6. In a linear model satisfying the second order assumptions, $E(Y) = \beta_1 x_1 + \cdots + \beta_p x_p$, where $x_1, \dots, x_p$ are the p columns of the design matrix. For each of the two statements below, state whether it is true or false, justifying your answer in each case.

   (a) If the vectors $x_1, \dots, x_p$ are mutually orthogonal, then the residuals are uncorrelated.

   (b) If a is a constant vector which is orthogonal to each x_i, then a is orthogonal to the vector of residuals.

Solution.

(a) False. Consider a simple linear regression problem $E(Y_i) = \beta_1 + \beta_2 x_i$, in which $\sum x_i = 0$. The 2 columns of the design matrix are orthogonal but

$$Cov(R_i, R_j) = -\sigma^2 \left(\frac{1}{n} + \frac{x_i x_j}{\sum_k x_k^2} \right)$$

which in general is non-zero.

(b) False. $a^T R = a^T (Y - X\hat{\beta}) = a^T Y$ because $a^T X = 0$. In general, $a^T Y \neq 0$.

7. In a study of the effect of thermal pollution on fish, the proportion of a certain variety of sunfish surviving a fixed level of thermal pollution was determined by Matis and Wehrly (1979) for various exposure times. The following paired data were reported on scaled time (x) versus proportion surviving (y).

x	0.10	0.15	0.2	0.25	0.30	0.35	0.40	0.45	0.5	0.55
y	1.00	0.95	0.95	0.9	0.85	0.7	0.65	0.60	0.55	0.40

(a) Plot the paired data as points in an x-y coordinate system.

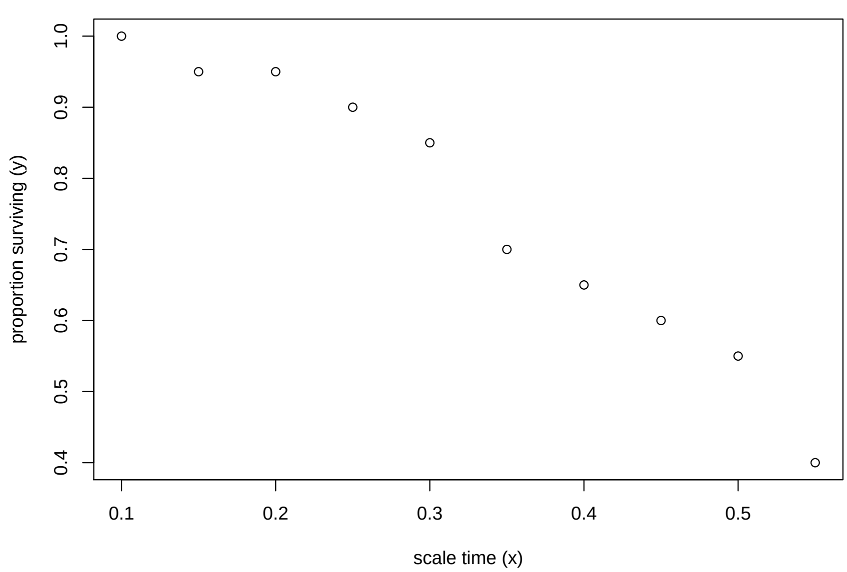
(b) Assuming a straight line regression compute the least squares estimates of β_0 (the intercept) and β_1 (the slope).

(c) Estimate $E(Y) = \beta_0 + \beta_1 x$ if the exposure time is $x = 0.325$ units.

(d) Compute the residual sum of squares and give an unbiased estimate of $\sigma^2 = Var(Y)$.

Solution.

(a)



There seems to be a linear relationship.

(b) Using the equations given in class we find

$$\hat{\beta}_0 = 1.1824 \text{ and } \hat{\beta}_1 = -1.3152$$

(c) $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \times 0.325 = 0.755.$

(d) The residual sum of squares are given by 0.015515 and an unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{0.015515}{10 - 2} = 0.001939$$