MAS programmes - Statistical Analysis (Autumn Term)

Exercises 2

1. Note that the simple linear regression (with one explanatory variable)

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, n \tag{$*$}$$

with $\epsilon_i \sim$ i.i.d$(0, \sigma^2)$, may be reparameterized as

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \epsilon_i \tag{$**$}$$

where $\bar{x} = \sum_i x_i/n$. From the reparameterized model $(**)$, it follows that the parameters α and β in $(*)$ can be recovered, via

$$\alpha = \beta_0 - \beta_1\bar{x} \quad \text{and} \quad \beta = \beta_1$$

By formulating an appropriate functional, $\mathcal{L}$, that can be used in determining the residual sum of squares (SS_R) for the reparameterized model $(**)$, find the ordinary least squares estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, of β_0 and β_1. Check that these estimates are consistent with those for α and β given in the notes to Lecture 1.

2. Now consider the simple linear regression in $(**)$ as a *multiple* regression, given in matrix-vector form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

$$\underset{n\times 1}{\mathbf{y}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{n\times 2}{\mathbf{X}} = \begin{pmatrix} 1 & x_1 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}, \quad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \text{ and } \underset{n\times 1}{\boldsymbol{\epsilon}} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

with $\boldsymbol{\epsilon} \sim \text{MVN}(\mathbf{0}, \sigma^2\mathbf{I})$. Show that adopting

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

leads to the same ordinary least squares estimates of β_0 and β_1 obtained in question 1 above.

3. Let $\mathbf{P}$ be a symmetric idempotent matrix.

   (a) Prove that all the eigenvalues of $\mathbf{P}$ are either 0 or 1.

   (b) Deduce that rank($\mathbf{P}$)=tr($\mathbf{P}$).

Note that the result of question 3(b) was referred to in Section 2.4 of Lecture 2.

The exercise overleaf allows you to explore the fitting of a simple linear regression model using R.

4. Schultz (1933) examined the relationship between the cost of various commodities and the demand for them in the USA. In this question, we will investigate such a relationship between sugar consumption `consump` (measured in pounds per capita) and its price `price` (in cents per pound) over the period 1875 to 1929.

A text file `sugar.txt` containing the data is available on Moodle. (The data are also repeated at the end of this document for reference).

(a) In economic theory, a suggested model for the relationship between consumption and price is given by

$$\text{consump} = A \exp(\alpha \text{ price})$$

where A and α are unknown constants.

Show how a model for log(consump) can be considered as a simple linear regression.

(b) Enter the data into R as the data frame `sugar`, and obtain plots for `consump` versus `price` and `lconsump` versus `price`, where `lconsump` is the logarithm of the `consump` values:

```
> lconsump <- log(consump)
```

Verify that using `lconsump` as the response is more appropriate for a simple linear regression.

(c) Fit a simple linear regression of `lconsump` on `price`. What is the fitted model?

(d) On the basis of the fitted model, calculate the predicted consumption of sugar per capita for a year in which the price of sugar is 6 cents per pound.

(e) Recreate the plot of `lconsump` versus `price` from part (b) and add the fitted regression line to the plot.

Create a further plot showing the fitted relationship, suggested by this analysis, between `consump` and `price` directly. [You can use this plot to check that your predicted value for the consumption of sugar when the price is 6 cents from part (d) makes sense].

sugar dataset

price	consump
9.485	39.375
10.163	37.607
11.426	35.254
11.566	35.714
10.713	37.513
9.798	42.120
9.385	43.464
8.319	47.550
8.506	50.091
7.290	52.375
7.855	50.752
7.646	55.790
7.423	51.724
8.054	55.679
9.550	50.821
7.526	51.924
5.875	65.095
5.718	62.607
6.371	63.174
5.886	65.402
6.017	62.194
6.867	61.367
6.823	63.672
7.410	60.502
6.559	61.682
6.488	64.623
6.313	67.646
5.304	71.694
5.521	69.836
5.681	74.308
6.112	69.342
4.962	74.048
4.843	75.977
5.447	79.393
5.069	79.760
5.126	80.710
5.626	79.542
5.092	81.937
4.278	86.237
4.779	85.406
5.504	85.117
5.403	80.777
4.329	80.210
4.017	75.075
4.349	86.197
5.018	85.215
4.222	84.373
3.962	103.049
5.481	95.292
5.161	95.357
3.448	106.672
3.625	108.244
4.047	99.690
3.743	102.872
3.442	106.376