Chapter 14

Multiple linear regression: Theory

14.1 The Regression Model

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times k}{\mathbf{X}}\ \underset{k\times 1}{\mathbf{b}} + \underset{n\times 1}{\varepsilon} \quad \text{where} \quad \varepsilon \sim \mathrm{N}\left(\mathbf{0}, \sigma^2\,\mathbf{I}\right)$$

y: is the vector of the response variable

X: the matrix of the k independent/explanatory variables (usually the first column is a column of ones for the constant term).

b: is a k x 1 vector of unknown parameters.

ε : is a vector of randomly distributed errors.

We assume that

$$\mathrm{E}(\varepsilon) = \mathbf{0} \quad \mathrm{Var}(\varepsilon) = \sigma^2\,\mathbf{I}$$

where

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Note that

$$E(y) = Xb \qquad Var(y) = \sigma^2\,\mathbf{I}$$

and that if

$$\varepsilon \sim N\left(\mathbf{0}, \sigma^2\,\mathbf{I}\right) \quad \Rightarrow \quad y \sim \mathrm{N}\left(Xb, \sigma^2\,\mathbf{I}\right)$$

14.2 Estimation of the parameters

14.2.1 Least Square Estimation

Minimise W with respect to **b** where

$$W=\varepsilon'\varepsilon=(y-Xb)'(y-Xb)$$

Now W can be written as

$$\begin{aligned} W &= (y'-b'X')(y-Xb) \\ &= y'y-y'Xb-b'X'y+b'X'Xb \\ &= y'y-2b'X'y+b'X'Xb \end{aligned}$$

(since $y'Xb=b'Xy$ a scalar). Differentiate with respect to b

$$\frac{\partial W}{\partial b}=-2X'y+2X'Xb$$

We set $\frac{\partial W}{\partial b}$ to zero to give the **Normal** equations

$$\boxed{X'X\hat{b}=X'y}$$

If now **X** is of full rank $\Rightarrow$ **X′X** is of full rank and $(\mathbf{X}'\mathbf{X})^{-1}$ exist so

$$\boxed{\hat{b}=(X'X)^{-1}X'y}$$

$\hat{b}$ is the least squares estimator of b.

14.2.2 Maximum Likelihood Estimation

We have that

$$y \sim N(Xb, \sigma^2 \mathbf{I})$$

so the likelihood function will be

$$L(b,\sigma^2)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}(y-Xb)'(y-Xb)\right\}$$

with log-likelihood

$$\ell(b,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - Xb)'(y - Xb)$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}W$$

Note: Maximising the log likelihood for **b** is equivalent of minimising the least squares quantity $W = (y - Xb)'(y - Xb)$. So in this case MLE and LSE for **b** are identical.

Differentiating the log likelihood with respect to **b** and σ^2 we have

$$\frac{\partial \ell}{\partial b} = \frac{X'y - X'Xb}{\sigma^2}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(y - X\hat{b})'(y - X\hat{b})}{2\sigma^4}$$

By setting the above equations equal to zero and solving them for **b** and σ^2 we have that the maximum likelihood estimator (MLE) for **b** and σ^2 are

$$\boxed{\begin{aligned}\hat{b} &= (X'X)^{-1}X'y \\ \hat{\sigma}^2 &= \frac{(y - X\hat{b})'(y - X\hat{b})}{n}\end{aligned}}$$

The quantity $(y - X\hat{b})'(y - X\hat{b})$ is called the Residual Sum of Squares (RSS) or Deviance (in GLIM).

The MLE for σ^2, $\hat{\sigma}^2$ is a biased estimate so we generally use

$$s^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})}{n - k}$$

where k is the rank of the matrix **X.**

14.3 The mean and variance for the least square estimators

The mean

$$
\begin{aligned}
\mathrm{E}(\hat{b}) &= \mathrm{E}\left((X'X)^{-1}X'y\right)\\
&= (X'X)^{-1}X'\mathrm{E}(\mathbf{y})\\
&= (X'X)^{-1}X'Xb\\
&= b
\end{aligned}
$$

so $\boxed{\mathrm{E}(\hat{b}) = b}$ that is $\hat{\mathbf{b}}$ is unbiased for **b.**

The variance

$$
\mathrm{Var}(\hat{b}) = \mathrm{E}\left[(\hat{b}-b)(\hat{b}-b)'\right]
$$

since $\mathrm{E}(\hat{b}) = b$.

Now $\hat{b} - b = (X'X)^{-1}X'\varepsilon$ Thus

$$
\begin{aligned}
\mathrm{Var}(\hat{b}) &= \mathrm{E}\left[\left((X'X)^{-1}X'\varepsilon\right)\left((X'X)^{-1}X'\varepsilon\right)'\right]\\
&= \mathrm{E}\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right]\\
&= (X'X)^{-1}X'\mathrm{E}(\varepsilon\varepsilon')X(X'X)^{-1}\\
&= (X'X)^{-1}X'\mathrm{Var}(\varepsilon)X(X'X)^{-1}\\
&= (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1}\\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

Note that $\mathrm{Var}(\varepsilon) = \mathrm{E}\left[(\varepsilon - \mathrm{E}(\varepsilon))(\varepsilon - \mathrm{E}(\varepsilon))'\right] = \mathrm{E}(\varepsilon\varepsilon')$ since $\mathrm{E}(\varepsilon) = \mathbf{0}$

14.4 Fitted Values and Residuals

Definitions

Fitted values $\hat{y} = X\hat{b}$

Residuals $\hat{\varepsilon} = y - \hat{y} = y - X\hat{b}$

$\hat{\mathbf{y}}$ is an estimate of $E(y) = Xb$ the mean of **y** and $\hat{\boldsymbol{\varepsilon}}$ is an estimate of $\boldsymbol{\varepsilon}$ the error term.

The Hat Matrix

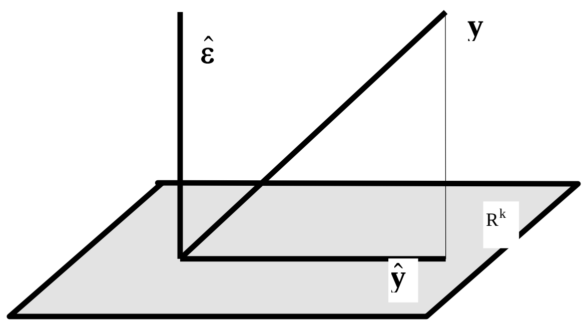
$$\hat{y} = X\hat{b} = X(X'X)^{-1}X'y = Hy$$

$$\boxed{H = X(X'X)^{-1}X'}$$ is called the Hat matrix.

H is symmetric and idempotent, so it is an orthogonal projection matrix. It projects any n-dimensional vector into the subspace generated by X. Also

$$\hat{\varepsilon} = y - X\hat{b} = y - X(X'X)^{-1}X'y$$
$$= (I - H)y$$

(I - H) is symmetric and idempotent so is an orthogonal projection matrix. **(I - H)** projects any n-dimensional vector into the orthogonal complement of **X** (see the figure below).



$\hat{y}$: is the orthogonal projection of y into the subspace generated by **X**. The hat matrix **H** projects **y** into the linear subspace generated by the columns of **X,** $\hat{y} = Hy$.

$\hat{\boldsymbol{\varepsilon}}$: is the orthogonal projection of **y** into the subspace generated by the orthogonal complement of **X**, $\hat{\varepsilon} = (I - H)y$.

14.5 Properties of the residuals and the fitted values

14.5.1 The mean of the fitted values

$$\boxed{E(\hat{y}) = Xb}$$

proof:

$$E(\hat{y}) = \mathrm{E}(Hy) = H\mathrm{E}(y) = HXb$$

$$= X(X'X)^{-1}X'Xb$$

$$= X\underline{b}$$

14.5.2 The variance of the fitted values

$$\boxed{\mathrm{var}(\hat{y}) = \sigma^2 H}$$

proof:

$$\mathrm{Var}(\hat{y}) = \mathrm{Var}(Hy) = H\mathrm{var}(\underline{y})H'$$

$$= H\sigma^2 IH' = \sigma^2 HH'$$

$$= \sigma^2 H$$

because **H** is symmetric and idempotent.

14.5.3 The expected value of the residuals

$$\boxed{E(\hat{\varepsilon}) = \mathbf{0}}$$

See exercise 14.1.

14.5.4 The variance of the residuals

$$\boxed{\mathrm{Var}(\varepsilon) = \sigma^2(I - H)}$$

See exercise 14.1.

14.5.5 The sum of the residuals

The following result holds for the sum of the residuals provided that the matrix of the explanatory variables **X** contains the constant term vector i.e. the first vector in **X** is **1**.

$$\boxed{\sum_{i=1}^{n} \hat{\boldsymbol{\varepsilon}}_i = \hat{\boldsymbol{\varepsilon}}'\mathbf{1} = \mathbf{0}}$$

proof: From the Normal equation we have

$$
\begin{aligned}
& X'X\hat{b} = X'y \\
\Rightarrow\ & X'\left(y - X\hat{b}\right) = 0 \\
\Rightarrow\ & X'\hat{\varepsilon} = 0 \quad \text{so if } \mathbf{X} \text{ contains the 1's.} \\
& \mathbf{1}'\varepsilon = 0
\end{aligned}
$$

14.5.6 The fitted values and the residuals are uncorrelated

The above statement is equivalent to the statement that

$$\boxed{\text{Cov}(\hat{\varepsilon}, \hat{y}) = \mathbf{0}}$$

proof:

$$
\begin{aligned}
\text{Cov}(\hat{\varepsilon}, \hat{y}) &= \text{E}\left[\left(y - X\hat{b}\right)\left(X\hat{b} - \text{E}\left(X\hat{b}\right)\right)'\right] \\
&= \text{E}\left[(y - Hy)(Hy - Xb)'\right] \\
&= \text{E}\left[(I - H)y\left(H(y - Xb)\right)'\right] \text{since } \mathbf{HX = X} \\
&= \text{E}\left[(I - H)(Xb + \varepsilon)\left(H(y - Xb)\right)'\right] \\
&= \text{E}\left[(I - H)\varepsilon\left(H\varepsilon\right)'\right] \text{since } (I - H)X = 0 \\
&= \text{E}\left[(I - H)\varepsilon\varepsilon' H\right] \\
&= (I - H)\text{E}(\varepsilon\varepsilon')H \\
&= (I - H)\text{Var}(\underline{\varepsilon})H \\
&= (I - H)\sigma^2 I H \\
&= \sigma^2 (I - H)H \\
&= \mathbf{0} \text{ since } (I - H)H = \mathbf{0}
\end{aligned}
$$

14.5.7 The covariance of the residuals and the y-variable

$$\boxed{\text{Cov}(\hat{\varepsilon}, y) = \sigma^2 (I - H)}$$

See exercise 14.1.

14.5.8 The covariance of the residuals and the estimator of the parameter b

$$\boxed{\text{Cov}(\hat{b}, \hat{\varepsilon}) = \mathbf{0}}$$

proof:

$$\begin{aligned}
\text{Cov}(\hat{b}, \hat{\varepsilon}) &= \text{E}\left[(\hat{b} - b)(\hat{\varepsilon} - \mathbf{0})' \right] \\
&= \text{E}\left[\left((X'X)^{-1} X' y - b\right)\left((I - H)\varepsilon\right)' \right] \\
&= \text{E}\left[\left((X'X)^{-1} X'(Xb + \varepsilon) - b\right)\left((I - H)\varepsilon\right)' \right] \\
&= \text{E}\left[(X'X)^{-1} X' \varepsilon\varepsilon' (I - H) \right] \\
&= (X'X)^{-1} X' \sigma^2 I (I - H) \\
&= \sigma^2 (X'X)^{-1} X' \left(I - X(X'X)^{-1} X'\right) \\
&= 0
\end{aligned}$$

14.6 The Residual Sum of Squares & other Sum of Squares

The residual sum of squares (RSS) or Deviance is defined as

$$\text{RSS} = \hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{b})'(y - X\hat{b})$$

$$= y'y - \hat{b}'X'y - y'X b + \hat{b}'X'X\hat{b}$$

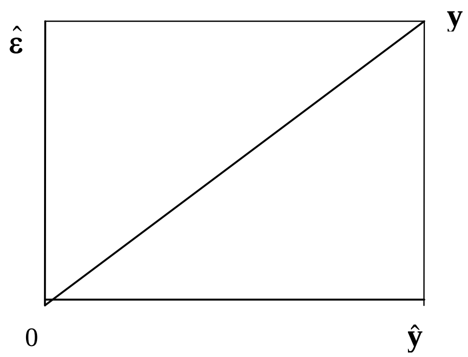
$$= y'y - 2\hat{b}'X'y + \hat{b}'(X'X)(X'X)^{-1} X \underline{y}$$

$$= y'y - \hat{b}'X'y$$

or

$$\boxed{\mathbf{y}'\,\mathbf{y} = \hat{\mathbf{b}}'\,\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\varepsilon}}'\,\hat{\boldsymbol{\varepsilon}}}$$

Total Sum of Squares = Regression sum of squares + Residual Sum of Squares



Note that

$$\hat{b}'X'y = \left((X'X)^{-1}X'y\right)'X'y = y'X(X'X)^{-1}X'y = y'Hy = \hat{y}'\hat{y}$$

so $\quad \mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \varepsilon'\varepsilon$

also $\quad \hat{\varepsilon}'\hat{\varepsilon} = y'y - \hat{b}X'y = y'y - yHy = y'(I-H)y$

So we have

$$\boxed{y'y = y'Hy + y'(I-H)y}$$

TSS = Regression SS + RSS

14.7 The R-square

How good the model is depends on how close the fitted values $\hat{\mathbf{y}}$ are to the actual values $\mathbf{y}$. The quantity

$$\hat{R}^2 = \frac{\hat{y}'\hat{y}}{y'y} = \frac{y'Hy}{y'y}$$

could be used as a measure of goodness of fit but the problem is that **X** usually contains the vector of ones. In order to eliminate the contribution of the constant term we use the quantity.

$$R^2 = \frac{\hat{y}'\hat{y} - n\bar{y}^2}{y'y - n\bar{y}^2} = \frac{\text{Regression SS (Adjusted)}}{\text{Total SS (Adjusted)}}$$

Note: $R^2 \times 100 = \%$ of variation explained by the fitted model.

Another measure which takes into consideration the number of x-variables used in the model is

$$\bar{R}^2 = R^2(\text{Adjusted}) = \frac{\dfrac{\text{Regression SS (Adj)}}{n-k-1}}{\dfrac{\text{TSS (Adj)}}{n-1}}$$

14.8 Statistical Hypotheses

14.8.1 The expected value of RSS.

$$E(\hat{\varepsilon}'\hat{\varepsilon}) = \sigma^2(n-k)$$
$$\Rightarrow E\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}\right) = \sigma^2$$

so $s^2 = \dfrac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k}$ is unbiased estimator for σ^2

14.8.2 The distribution and C.I. for $\hat{\mathbf{b}}$

Note that

$$\hat{b} \sim N\left(b, \sigma^2 (X'X)^{-1}\right)$$

So

$$\frac{\hat{b}_i - b_i}{\sqrt{\sigma^2 a_{ii}}} \sim N(0,1) \text{ where } a_{ii} \text{ are the diagonal elements of } (X'X)^{-1}.$$

Also it can be proved that.

$$\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} \sim \chi^2(n-k)$$

and that $\hat{\mathbf{b}}$ and $\dfrac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\sigma^2}$ are independent.

From section 2.4.1, we note that if $Z \sim N(1, 0)$ and $W \sim \chi^2(v)$ and Z and W are independent then

$$T = \frac{Z}{\sqrt{\frac{W}{\nu}}}.$$

hence we have that

$$t=\frac{\dfrac{\hat{b}_i-b_i}{\left(\sigma^2\, a_{ii}\right)^{\frac{1}{2}}}}{\left(\dfrac{\underline{\hat{\varepsilon}}'\,\underline{\hat{\varepsilon}}/\sigma^2}{n-u}\right)^{\frac{1}{2}}}=\frac{\hat{b}_i-b_i}{\left(\hat{\sigma}^2\, a_{ii}\right)^{\frac{1}{2}}}=\frac{\hat{b}_i-b_i}{se(\hat{b}_i)}\sim t(n-u)$$

where

$$\hat{\sigma}^2=\frac{\underline{\hat{\varepsilon}}'\,\underline{\varepsilon}}{n-u}$$

Using the above result we can define 100 (1 - α)% confidence intervals (C.I.) for the $b_i'\,s$.

$$\hat{b}_i \pm t_{n-k,\frac{\alpha}{2}}\; se\left(\hat{b}_i\right)$$

Note: The variance covariance matrix for $\hat{\mathrm{b}}$ is $\mathrm{var}\left(\hat{\mathbf{b}}\right)=\sigma^2\left(X'X\right)^{-1}$. Since we do not know σ^2 we use $\hat{\sigma}^2\left(X'X\right)^{-1}$ as an estimate. This matrix in general does not have the off-diagonal elements equal to zero so the $\hat{b}$ are correlated.

14.8.3 t - test for the $\mathbf{b}'$s.

We can use the quantity

$$t=\frac{b_i-b_0}{se\left(\hat{b}\right)}$$

to test whether the coefficient is equal to some specific value b_0 or not. The most common hypothesis is $b_i = 0$. i.e.

$H_o : b_i = 0$ the other $b'\,s$ are unconstrained.

H_1 : All the $b'\,s$ are unconstrained.

Reject H_o if

$$|t| = \left|\frac{\hat{\mathrm{b}}_{\mathrm{i}}}{\mathrm{se}(\hat{\mathrm{b}}_{\mathrm{i}})}\right| > t_{n-k,\frac{\alpha}{2}}$$

(for two tail test).

Note:

(i) This test is appropriate when we want to test the coefficient of a term given that all the other term are included in the model.

(ii) This test is not appropriate to test $b_1 = b_2 = 0$ simultaneously.

14.8.4 The F - test

This is relevant for testing whether a subset of the x-variables contributes significantly in explaining the variation in the y-variable.

The model

$$\underset{nx1}{\mathbf{y}} = \underset{nxk}{\mathbf{X}}\underset{kx1}{\mathbf{b}} = \underset{nxk}{\left[\underset{nx\ell}{\mathbf{X}_1}\ \underset{n-(k-\ell)}{\mathbf{X}_2}\right]}\overset{kx1}{\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix}}\begin{matrix}\}\ \ \ell\\ \}\ \ n-\ell\end{matrix}$$

The test

(i) $H_o : \mathbf{b}_2 = \mathbf{0}$ and $\mathbf{b}_1$ unconstrained (that is the true model is $\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \boldsymbol{\varepsilon}$)

$H_1 : \mathbf{b}_2 \neq \mathbf{0}$ $\mathbf{b}_1$ and $\mathbf{b}_2$ unconstrained (the true model is $\mathbf{y} = \mathbf{X}\ \mathbf{y} + \boldsymbol{\varepsilon}$)

(ii) Fit the H_o model and obtain its RSS= Deviance = D_o and its degrees of freedom df_0.

(iii) Fit the H_1 model and obtain its Deviance = D_1 and its degrees of freedom df_1.

(iv) It can be shown that if H_o is true $\frac{D_o - D_1}{\sigma^2} \sim \chi^2(df_0 - df_1)$ independently of $\frac{D_1}{\sigma^2} \sim \chi^2(df_1)$ so the ratio

$$\boxed{F = \frac{\frac{D_o - D_1}{(df_0 - df_1)}}{\frac{D_1}{df_1}} \sim F_{(df_0 - df_1), df_1}}$$

(If H_o is true then both $D_o - D_1$ and D_1 measure the random error, but if H_o is false then we would expect $D_o - D_1$ which measures the variation explained by $\mathbf{X}_1$ to be significantly bigger than D_1).

This test is a one-sided F test, that is, reject H_o at $100\alpha\%$ level if the observed

$$F = \frac{\frac{D_o - D_1}{(df_0 - df_1)}}{\frac{D_1}{df_0}} > F_{(df_o - df_1), df_1, \alpha}$$

Special cases.

(a) Testing whether jointly all the x-variables explain a significant part of the variation in $\mathbf{y}$. (Note that b_1 represent the constant term).

$H_o : b_2 = b_3 = .. = b_k = 0$ $\quad$ b_1 : unconstrained. (that is the true model is $y_i = \mu + \varepsilon_i$).

$H_1 : b_1, b_2, .. b_k$ $\quad$ unconstrained. ($\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$ is the true model).

The F test will be

$$F = \frac{\frac{D_o - D_1}{(df_0 - df_1)}}{\frac{D_1}{df_1}}$$

(b) Testing whether only one of the x's explains a significant part of the variation given the rest.

$H_o : b_j = 0$ $\quad$ the test of b_i unconstrained

$H_1 : b_1, b_2 \ldots b_k$ $\quad$ all the b's unconstrained.

Consider the case k=4 for illustration

with $$F = \frac{\frac{D_o - D_1}{(df_0 - df_1)}}{\frac{D_1}{df_1}} \sim F_{1, df_1} .$$

Note that $t^2_{n-k} = F_{1,(n-k)}$ or $t^2_k = F_{1,k}$ so this is equivalent to a t - test.

Exercise 14.1

1. If $\mathbf{X}$ is a (n x k) matrix of rank k.

   i) show that $\mathbf{X'X}$ and $\mathbf{XX'}$ are symmetric matrices.

   ii) show that $\mathbf{H} = \mathbf{X(X'X)}^{-1}\mathbf{X'}$ and $\mathbf{I} - \mathbf{H}$ are both symmetric and idempotent matrices. (That is, both matrices are orthogonal projections)

2. Shown that $\hat{\mathbf{b}} = \mathbf{b} + \mathbf{(X'X)}^{-1}\mathbf{X'\varepsilon}$ where $\hat{\mathbf{b}}$ is the least squares estimator and $\boldsymbol{\varepsilon}$ is the error term

3. Show that the expected value of the residuals is zero i.e. $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}) = \mathbf{0}$ and that $\mathrm{Var}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H})$.

4. Show that $\mathrm{Cov}(\hat{\boldsymbol{\varepsilon}}, \mathbf{y}) = \sigma^2(\mathbf{I} - \mathbf{H})$, and deduce that in general the residual estimates are correlated with the observations.

5. Show that the residual sum of squares is given by $\mathbf{y'(I - H)y}$ where $\mathbf{H} = \mathbf{X(X'X)}^{-1}\mathbf{X'}$.

Exercise 14.2

1) Let the design matrix $\mathbf{X} = [\mathbf{x_1} \quad \mathbf{x_2} \quad \mathbf{x_3} \quad \mathbf{x_4}]$ where $\mathbf{x_1}$ represents the constant term in the linear model. Give the model equation corresponding to the null and alternative hypothesis in the following test.

H_0: $b_2 = b_3 = b_4 = 0$: b_1 unconstrained.

H_1: $b_4 = 0$: b_1, b_2, b_3 unconstrained

2) With the same design matrix as above give the model equation corresponding to the null and alternative hypothesis in the following test.

H_0: $b_3 = 2b_4$: $b_1, b_2 = 0$ unconstrained.

H_1:: b_1, b_2, b_3, b_4 unconstrained

3) Write down the GLIM commands in order to test the hypothesis in 1) and 2) and indicate how to use the resulting Deviances to test the relevant hypotheses.

4) From the anaerobic Threshold output in section 10.2 test whether the quadratic model X<2> is better than the X<6> model.

Exercise 14.3

Assume the multiple linear regression model of the form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

where **y** is an (n x 1) vector of observations, **X** is an (n x k) design matrix of rank (k < n), **b** is a (k x 1) vector of parameters and $\boldsymbol{\varepsilon}$ is an (n x 1) vector of random variables (error term) such that

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

i) Show that $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the Least Squares Estimator of **b,** and that $\hat{\mathbf{b}}$ is an unbiased estimator for **b**.

ii) Show that the likelihood function for **b** and σ^2 is given by

$$L(\mathbf{b}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{(\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b})}{2\sigma^2}\right\}$$

iii) State briefly the reason why the Least Squares estimator and the Maximum Likelihood estimator for **b** are identical.

iv) Use the result $\hat{\mathbf{b}} - \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$ to show that the variance-covariance matrix of $\hat{\mathbf{b}}$ is equal to $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

v) Show that $\hat{\boldsymbol{\varepsilon}} = (\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the hat matrix, $\hat{\boldsymbol{\varepsilon}}$ is the vector of residuals and $\boldsymbol{\varepsilon}$ the error term. Use this result to find the expected value, and the variance-covariance matrix of $\hat{\varepsilon}$.