Part II

Multiple Linear Regression

Chapter 6

Multiple Regression–I

We look at linear regression models with multiple predictor variables.

6.1 Multiple Regression Models

SAS program: att5-6-1-read-multreg

We look at special cases of the multiple general linear regression models, given by,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, \quad i = 1, \ldots, n$$

where
Y_i is the response for the ith trial
β_0, β_1, β_2, ..., β_{p-1} are parameters
X_{i1}, ..., $X_{i,p-1}$ are (known) values[1]
ε_i are independent $N(0, \sigma^2)$
$i = 1, \ldots, n$

We find out that this *general* linear model is able to fit a surprisingly large set of different data sets.

Exercise 6.1 (Multiple Regression Models)

1. *A First Look, Reading Versus Illumination.*

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

[1]These quantities are also known as predictor or independent variables.

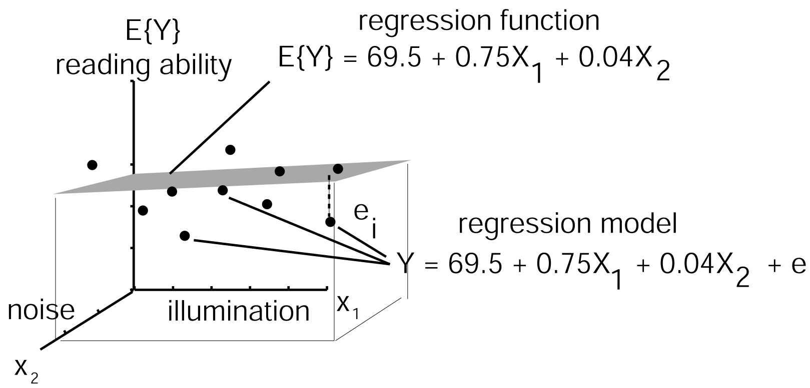

Figure 6.1 (Scatter Plot and Multiple Regression)

(a) *Number of Predictors.*
There are (circle one) **one** / **two** / **three** predictors in the regression model used to describe this data.

(b) *Dimension of Scatter Plot and Number of Predictors.*
True / **False** A regression model with *two* predictors appears as a *surface* on a scatter plot in *three* dimensions, (X_1, X_2, Y).

(c) *Selecting The Functional Form Of The Regression Relation.*
The regression is (circle one) **linear** / **quadratic** in the X_i.

(d) *Regression and Causality.*
True / **False** The regression does *not* say that different illumination and/or light levels *causes* different reading abilities.

(e) There are (circle one) **10** / **20** / **30** data points.

(f) One particular data point is (circle one) $\mathbf{(9, 15)}$ / $\mathbf{(9, 15, 70)}$ / $\mathbf{(15, 70)}$.

(g) The data point (9,15,70) means (circle one)

i. for a level of illumination of 9, the reading ability is 70.
ii. for a noise level of 15, the reading ability is 70.
iii. for a level of illumination of 9 and a noise level of 15, the reading ability is 70.

(h) The *multiple linear regression function* is given by the equation,

$$E\{Y\} = 69.5 + 0.75X_1 + 0.04X_2$$

The Y–intercept of this line, β_0, is (circle one) **69.5** / **0.75** / **0.04**.
The *slope* in the X_1 direction, β_1, is (circle one) **69.5** / **0.75** / **0.04**.
The *slope* in the X_2 direction, β_2, is (circle one) **69.5** / **0.75** / **0.04**.

(i) **True** / **False** The parameter $\beta_1 = 0.75$, is interpreted to mean, on average, the reading ability increases 0.75 units for an increase of *one* unit of the level of illumination, *while the noise level is held constant.*

(j) **True** / **False** The parameter $\beta_2 = 0.04$, is interpreted to mean, on average, the reading ability increases 0.04 units for an increase of *one* unit of the noise level, *while the level of illumination is held constant.*

(k) The *mean* value of the reading ability at $(X_1, X_2) = (19, 26)$, is
$E\{Y\} = 69.5 + 0.75(19) + 0.04(26) \approx$ (circle one) **83.52** / **84.79** / **87.23**.

(l) **True** / **False** Draw a *vertical* line which passes through (19,26) on the "(X_1, X_2)" plane. Now draw an *horizontal* line which passes through the point where the solid regression plane and the previously drawn vertical line intersect. This horizontal line will intersect the "reading ability" axis at 84.79.

(m) *Error, ε_i.*
At level $(X_1, X_2) = (19, 26)$, $E\{Y\} = 84.8$. The difference between this value and the *observed* value, $Y = 94$ (look at the table of the data above) is called the *error* and is given by
$\varepsilon_i = Y_i - E\{Y\} = 94 - 84.8 =$ (circle one) **9.1** / **9.2** / **9.3**.

(n) If we were to draw the error for $(X, Y_i) = (10, 5.88)$ on the scatter plot, we would (circle one)

i. draw line parallel to the regression line
ii. draw a line connecting the point (10,5.88) to the point (20,5.88)
iii. draw a line connecting the point (19,26,84.8) to (19,26,94)
iv. draw a *vertical* line connecting (10,5.88) to the regression line at the point (10,5.92)
v. draw a *horizontal* line connecting (10,5.88) to the regression line

(o) There are (circle one) **1** / **5** / **10** errors on the scatter plot.

(p) At level $(X_1, X_2) = (2, 3)$, $E\{Y\} = 69.5 + 0.75(2) + 0.04(3) = 71.12$. In this case, since $(X_1, X_2) = (2, 3)$ is outside the range of data, $(X_1, X_2) \in (9, 33) \times (15, 37)$, the predicted value, $E\{Y\} \approx 71.12$, is most likely a (circle one) **poor** / **good** estimate of reading ability.

(q) The range of data most likely covers an area (circle one) **smaller** / **exactly equal to** / **bigger** than the rectangular area given by $(9, 33) \times (15, 37)$.

(r) *Some Terminology.*
True / **False** The model $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$ is *multiple* because there is two predictors, X_{i1} and X_{i2}. This model is *linear in the parameters* because parameters β_0, β_1, β_2 do not appear in an exponent and are not multiplied or divided together. This model is *linear in the predictors* because predictors X_{i1}, X_{i2} do not appear in an exponent and are not multiplied or divided together.

(s) *Properties of Error and Responses.*
True / **False** The errors, ε_i, are random variables and are *assumed*[2] to have zero mean, have constant variance, σ^2, and be uncorrelated. This implies, since $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$, that the responses, Y_i are also random variables. More than this, Y_i have means equal to $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$, with constant variance, σ^2, and are uncorrelated.

(t) *Regression Model Versus Regression Function.*
True / **False** The *regression model*,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i,$$

describes all of the *points* on the scatter plot, whereas the *regression function*,

$$E\{Y_i\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2},$$

describes the *plane* superimposed on the scatter plot.

(u) *Additive Effects.*
True / **False**. In this model, we *assume*[3] the effect of X_1 on $E\{Y\}$ does not depend on X_2. This is also true of X_2. In other words, X_1 and X_2 do not interact with one another. The model is said to be *additive*.

(v) *Hyperplane.*
True / **False**. On the one hand, in this reading data example, the response surface of the regression function is a plane. On the other hand, the model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

is a *three–dimensional hyperplane* response surface.

(w) If we sampled at random another ten individuals, we would get (circle one) **the same** / **different** scatter plot of points than those given in the diagram above. The data in the diagram above is a example of a (circle one) **sample** / **population**. This data is said to be (circle one) **observed** / **not observed**.

2. *General Linear Regression Models.*
General linear regression models do *not* have to have *linear* response surfaces, such as for the reading data given above. As long as the regression function can be written in the following way,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, i = 1, \ldots, n$$

the regression is "general" linear.

[2]This assumption may or may not be true. We need to investigate if the zero mean, constant variance, uncorrelated errors assumption is, in fact, appropriate for a given data set.

[3]This assumption may or may not be true. We need to investigate if an additive model is, in fact, appropriate for a given data set.

(a) *Example of General Linear.*
Since the following regression function,

$$Y_i = \beta_0 + \beta_1 X_{i1}^2 + \beta_2 X_{i2}^2 + \varepsilon_i,$$

where $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$ and which appears in the figure below

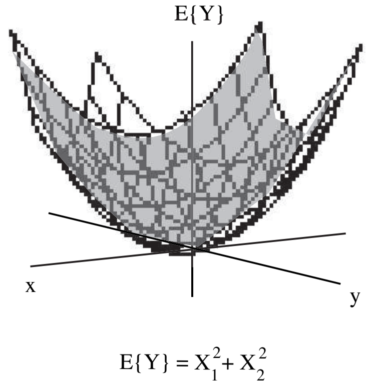

Figure 6.2 (General Linear Regression)

can be rewritten as

$$Y_i = \beta_0 + \beta_1 X'_{i1} + \beta_2 X'_{i2} + \varepsilon_i$$

where $X'_{i1} = X_{i1}^2$ and $X'_{i2} = X_{i2}^2$,
this regression function is (circle one) **general linear** / **nonlinear**.

(b) *Polynomials Are General Linear.*
Since

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i$$

which can be rewritten as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

where $X_{i1} = X_i$, $X_{i2} = X_i^2$ and $X_{i3} = X_i^3$
this regression function is (circle one) **general linear** / **nonlinear**

(c) *Transformed Variables Can Be Part Of General Linear.*
Since

$$\ln Y_i = \beta_0 + \beta_1 e^{X_{i1}} + \beta_2 X_{i2}^2 + \varepsilon_i$$

can be rewritten as

$$Y'_i = \beta_0 + \beta_1 X'_{i1} + \beta_2 X'_{i2} + \varepsilon_i$$

where $Y'_i = \ln Y_i$, $X'_{i1} = e^{X_{i1}}$ and $X'_{i2} = X_{i2}^2$,
this regression function is (circle one) **general linear** / **nonlinear**

(d) *Interactions Can Be Part Of General Linear.*
Since

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

can be rewritten as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

where $X_{i3} = X_{i1} X_{i2}$, an interaction term,
this regression function is (circle one) **general linear** / **nonlinear**
Notice, too, in this case, that since X_{i3} "depends" on X_{i1} and X_{i2},
this model (choose one) **is** / **is not** additive;
in other words, general linear models do not need to be additive.

(e) *Nonlinear.*
Since

$$Y_i = \beta_0 e^{\beta_1 X_{i1}} + \varepsilon_i$$

can*not* be rewritten in the following way

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

this regression function is (circle one) **general linear** / **nonlinear**

(f) *Qualitative Predictor Variables Can Be Part Of General Linear.*
General linear models can also deal with qualitative variables, in addition to quantitative variables. For example, since

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

where X_{i1} is gender and

$$X_{i1} = \begin{cases} 1 & \text{if male} \\ 0 & \text{if female,} \end{cases}$$

and X_{i2} is nationality where

$$X_{i2} = \begin{cases} 1 & \text{if Australian} \\ 0 & \text{if New Zealander,} \end{cases}$$

this regression function is (circle one) **general linear** / **nonlinear**.
In general, c classes can be defined by $c - 1$ qualitative predictor variables.

6.2 General Linear Model in Matrix Terms

The general linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, i = 1, \ldots, n$$

can be written using in matrix notation in the following way,

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

where

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \underset{p\times 1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}, \quad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

and

$$\underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{bmatrix}$$

Also,

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \mathbf{X}\beta, \quad \underset{n\times n}{\boldsymbol{\sigma^2}\{\mathbf{Y}\}} = \sigma^2 \mathbf{I} \quad \underset{n\times n}{\boldsymbol{\sigma^2}\{\varepsilon\}} = \sigma^2 \mathbf{I},$$

Exercise 6.2 (General Linear Model in Matrix Terms)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

The matrix version of the multiple regression model in this case is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{10} \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{10,1} \end{bmatrix} + \beta_2 \begin{bmatrix} X_{12} \\ X_{22} \\ \vdots \\ X_{10,2} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{10} \end{bmatrix}$$

or,

$$\begin{bmatrix} 70 \\ 70 \\ \vdots \\ 85 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} 9 \\ 7 \\ \vdots \\ 33 \end{bmatrix} + \beta_2 \begin{bmatrix} 15 \\ 20 \\ \vdots \\ 37 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{10} \end{bmatrix}$$

or

$$\begin{bmatrix} 70 \\ 70 \\ \vdots \\ 85 \end{bmatrix} = \begin{bmatrix} 1 & 9 & 15 \\ 1 & 7 & 20 \\ \vdots & & \vdots \\ 1 & 33 & 37 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{10} \end{bmatrix}$$

1. $\underset{10\times 1}{\mathbf{Y}}$ = (circle one)

$$\begin{bmatrix} 1 \\ 2 \\ \vdots \\ 10 \end{bmatrix}, \quad \begin{bmatrix} 9 \\ 7 \\ \vdots \\ 33 \end{bmatrix}, \quad \begin{bmatrix} 70 \\ 70 \\ \vdots \\ 85 \end{bmatrix}$$

   Store this matrix in the [B] matrix in your calculator.

2. $\underset{10\times 3}{\mathbf{X}}$ = (circle one)

$$\begin{bmatrix} 1 & 9 & 15 \\ 1 & 7 & 20 \\ \vdots & & \vdots \\ 1 & 33 & 37 \end{bmatrix}, \quad \begin{bmatrix} 9 & 1 & 15 \\ 7 & 1 & 20 \\ \vdots & & \vdots \\ 33 & 1 & 37 \end{bmatrix}, \quad \begin{bmatrix} 70 \\ 70 \\ \vdots \\ 85 \end{bmatrix}$$

   Store this matrix in the [A] matrix in your calculator.

3. And so,

   (a) $\underset{10\times 1}{\mathbf{Y}} = \underset{10\times 3}{\mathbf{X}} \underset{3\times 1}{\beta} + \underset{10\times 1}{\varepsilon}$

   (b) $\underset{3\times 1}{\mathbf{Y}} = \underset{3\times 1}{\beta} \underset{10\times 3}{\mathbf{X}} + \underset{10\times 1}{\varepsilon}$

   (c) $\underset{3\times 1}{\mathbf{Y}} = \underset{3\times 10}{\mathbf{X}} \underset{10\times 1}{\beta} + \underset{3\times 1}{\varepsilon}$

6.3 Estimation of Regression Coefficients

SAS program: att5-6-3-read-b-matrices

Using the least squares method, the estimators for the coefficients, β_0, β_1, …, β_{p-1}, in the general linear regression model,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, i = 1, \ldots, n$$

are

$$\underset{p\times 1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} = \underset{p\times p}{(\mathbf{X}'\mathbf{X})}^{-1}\underset{p\times 1}{\mathbf{X}'\mathbf{Y}}$$

Exercise 6.3 (Estimation of Regression Coefficients)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

1. $\underset{3\times 3}{(\mathbf{X}'\mathbf{X})}^{-\mathbf{1}}$ = (circle one)

$$\begin{bmatrix} 3.086 & 0.116 & -0.213 \\ 1.116 & 0.009 & -0.012 \\ -0.213 & -0.012 & 0.018 \end{bmatrix}, \begin{bmatrix} 3.086 & 0.116 & -0.213 \\ 0.116 & 0.009 & -0.012 \\ 0.213 & -0.012 & 0.018 \end{bmatrix}, \begin{bmatrix} 2.086 & 0.116 & -0.213 \\ 0.116 & 0.009 & -0.012 \\ -0.213 & -0.012 & 0.018 \end{bmatrix}$$

2. $\underset{3\times 1}{\mathbf{XY}}$ = (circle one)

$$\begin{bmatrix} 865 \\ 17,613 \\ 21,604 \end{bmatrix}, \begin{bmatrix} 85,500 \\ 17,613 \\ 21,604 \end{bmatrix}, \begin{bmatrix} 855 \\ 17,613 \\ 21,604 \end{bmatrix}$$

3. $\underset{3\times 1}{\mathbf{b}}$ = (circle one)

$$\begin{bmatrix} 69.498 \\ 1.749 \\ 0.044 \end{bmatrix}, \begin{bmatrix} 69.498 \\ 0.749 \\ 1.044 \end{bmatrix}, \begin{bmatrix} 69.498 \\ 0.749 \\ 0.044 \end{bmatrix}$$

6.4 Fitted Values and Residuals

SAS program: att5-6-4-read-e-matrices

Recall, an important matrix in regression analysis is the *hat* matrix, given by,

$$\underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$$

The fitted values, residuals and estimated variance–covariance of the residuals in the general linear regression model are given by, respectively,

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \mathbf{Xb} = \mathbf{HY}$$
$$\underset{n\times 1}{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$
$$\underset{n\times n}{s^2\{\mathbf{e}\}} = \left(\frac{\mathbf{e}'\mathbf{e}}{n-p}\right)(\mathbf{I} - \mathbf{H}) = MSE(\mathbf{I} - \mathbf{H})$$

These quantities, as we have already seen and will continue to see, are useful in assessing whether or not the simple linear model is a good fit to the data.

Exercise 6.4 (Fitted Values and Residuals)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

1. $\underset{10\times 10}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ = (circle one)

$$\begin{bmatrix} 0.375 & \cdots & 0.251 \\ \vdots & & \vdots \\ -0.251 & \cdots & 0.549 \end{bmatrix}, \quad \begin{bmatrix} 0.475 & \cdots & -0.251 \\ \vdots & & \vdots \\ -0.251 & \cdots & 0.549 \end{bmatrix}, \quad \begin{bmatrix} 0.375 & \cdots & -0.251 \\ \vdots & & \vdots \\ -0.251 & \cdots & 0.549 \end{bmatrix}$$

2. $\underset{10\times 1}{\hat{\mathbf{Y}}} = \underset{10\times 10}{\mathbf{H}} \; \underset{10\times 1}{\mathbf{Y}}$ = (circle one)

$$\begin{bmatrix} 74.902 \\ 75.626 \\ \vdots \\ 95.852 \end{bmatrix}, \quad \begin{bmatrix} 73.902 \\ 75.626 \\ \vdots \\ 95.852 \end{bmatrix}, \quad \begin{bmatrix} 76.902 \\ 75.626 \\ \vdots \\ 95.852 \end{bmatrix}$$

3. $\underset{10\times 1}{\mathbf{e}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ = (circle one)

$$\begin{bmatrix} -6.902 \\ -5.626 \\ \vdots - 10.852 \end{bmatrix}, \quad \begin{bmatrix} -7.902 \\ -5.626 \\ \vdots - 10.852 \end{bmatrix}, \quad \begin{bmatrix} -6.902 \\ -4.626 \\ \vdots - 10.852 \end{bmatrix}$$

4. $MSE = \mathbf{e}'\mathbf{e} \div (10 - 3)$ = (circle one) **506.531** / **411.56** / **72.361**

5. $\underset{10\times 10}{s^2\{\mathbf{e}\}} = MSE(\mathbf{I} - \mathbf{H})$ = (circle one)

$$\begin{bmatrix} 9.561 & \cdots & 15.869 \\ \vdots & & \vdots \\ 15.869 & \cdots & 28.581 \end{bmatrix}, \quad \begin{bmatrix} 35.165 & \cdots & 14.106 \\ \vdots & & \vdots \\ 14.106 & \cdots & 25.405 \end{bmatrix}, \quad \begin{bmatrix} 45.212 & \cdots & 18.136 \\ \vdots & & \vdots \\ 18.136 & \cdots & 32.664 \end{bmatrix}$$

6.5 Analysis of Variance Results

SAS programs: att5-6-5-read-anova-matrices

We test the regression (if $\beta_1 = \beta_2 = \cdots = 0$ or not) using the following ANOVA table.

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	$SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{\mathbf{1}}{\mathbf{n}}\right)\mathbf{J}\right]\mathbf{Y}$	$p-1$	$MSR = \frac{SSR}{p-1}$
Error	$SSE = \mathbf{Y}'\left[\mathbf{I} - \mathbf{H}\right]\mathbf{Y}$	$n-p$	$MSE = \frac{SSE}{n-p}$
Total	$SSTO = \mathbf{Y}'\left[\mathbf{I} - \mathbf{H}\right]\mathbf{Y}$	$n-1$	

where, recall, the matrix $\underset{n\times n}{J}$ is a matrix of ones (1s) and $\mathbf{H}$ is the hat matrix.

Exercise 6.5 (Analysis of Variance Results)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

1. The regression sums of squares is given by:
   $SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{\mathbf{1}}{\mathbf{n}}\right)\mathbf{J}\right]\mathbf{Y} =$ (circle one) **482.4** / **465.97** / **972.5**

2. The error sums of squares is given by:
   $SSE = \mathbf{Y}'\left[\mathbf{I} - \mathbf{H}\right]\mathbf{Y} =$ (circle one) **482.4** / **506.5** / **972.5**

3. The total sums of squares is given by:
   $SSTO = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{\mathbf{1}}{\mathbf{n}}\right)\mathbf{J}\right]\mathbf{Y} =$ (circle one) **482.4** / **490.1** / **972.5**

4. **True** / **False**. The ANOVA table is given by

Source	Sum Of Squares	Degrees of Freedom	Mean Squares
Regression	466.0	2	233
Error	506.5	7	72.4
Total	972.5	9	

5. *Test Of Regression Relation.*
   Test if all β_i, $i = 1, 2$ are *not* zero at $\alpha = 0.05$.

   (a) *Statement.*
   The statement of the test is (check none, one or more):

   i. $H_0 : \beta_1 = \beta_2 = 0$ versus $H_a :\ \beta_1 > 0, \beta_2 \neq 0$.
   ii. $H_0 : \beta_1 = \beta_2 = 0$ versus $H_a :$ not all β_i equal to zero.
   iii. $H_0 : \beta_1 = \beta_2 = 0$ versus $H_a :\ \beta_1 \neq 0, \beta_2 \neq 0$.

(b) *Test.*
The test statistic is

$$F^* \text{ test statistic } = \frac{233}{72.4} =$$

(circle one) **3.22** / **7.88** / **8.88**.
The upper critical value at $\alpha = 0.05$, with $(2, 7)$ degrees of freedom is (circle one) **4.74** / **6.32** / **7.32**
(Use PRGM INVF ENTER 2 ENTER 7 ENTER 0.95 ENTER)

(c) *Conclusion.*
Since the test statistic, 3.22, is smaller than the critical value, 4.74, we (circle one) **accept** / **reject** the null hypothesis that β_1 and β_2 are both zero.

6. *Coefficient of Multiple Determination.*

$$R^2 = \frac{SSR}{SSTO} = \frac{466}{972.5}$$

(circle one) **0.48** / **0.51** / **0.56**
and so the coefficient of multiple correlation is $\sqrt{R^2} = R = 0.69$

6.6 Inference about Regression Parameters

SAS programs: att5-6-6-read-parCI,test

The estimated variance–covariance matrix for the regression coefficients is given by

$$\underset{p\times p}{\mathbf{s^2}\{\mathbf{b}\}} = MSE(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$$

The various elements of this matrix, such as $s^2\{b_1\}$ or $s^2\{b_2\}$, are used as part of the formulas used to perform tests and calculate both statement and family (simultaneous) confidence intervals for the β_i, $i = 1, \ldots, p$ parameters.

$$\begin{aligned} t^* &= \frac{b_k}{s\{b_k\}}, \quad \text{using} \quad t(1-\alpha/2, n-p) \\ b_k &\pm t(1-\alpha/2, n-p)s\{b_k\} \quad \text{statement CI} \\ b_k &\pm Bs\{b_k\}, \quad \text{where} \quad B = t(1-\alpha/2g, n-p) \quad \text{family CI} \end{aligned}$$

Exercise 6.6 (Inference about Regression Parameters)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

1. *Variance–Covariance Matrix.*
   $\mathbf{s^2}\{\mathbf{b}\} = MSE(\mathbf{X'X})^{-1}$ = (circle one)

$$\begin{bmatrix} 9.94 & 0.37 & -0.69 \\ 0.37 & 0.53 & -0.04 \\ -0.69 & -0.04 & 0.06 \end{bmatrix}, \quad \begin{bmatrix} 223.44 & 8.39 & -15.45 \\ 8.39 & 0.68 & -0.89 \\ -15.45 & -0.89 & 1.33 \end{bmatrix}, \quad \begin{bmatrix} 7.94 & 0.37 & -0.69 \\ 0.37 & 0.03 & -0.04 \\ -0.69 & -0.04 & 0.06 \end{bmatrix}$$

2. A 95% CI for the parameter β_1 is given by
   $b_1 \pm t(1-\alpha/2, n-p)s\{b_1\} = 0.749 \pm 2.36\sqrt{0.68}$
   (circle one) **(69.03, 105.84)** / **(−1.197, 2.695)** / **(0.6782, 0.8198)**.

3. A 95% CI for the parameter β_2 is given by
   $b_1 \pm t(1-\alpha/2, n-p)s\{b_1\} = 0.044 \pm 2.36\sqrt{1.33}$
   (circle one) **(1.248, 15.128)** / **(−2.678, 2.766)** / **(−0.0976, 0.1856)**.

4. A 95% *family* CI for the parameter β_1
   (also accounting for β_2; that is, $g = 2$) is given by
   $b_1 \pm t(1-\alpha/2g, n-p)s\{b_1\} = 0.749 \pm 2.84(0.82622)$
   (circle one) **(1.248, 15.128)** / **(−1.593, 3.091)** / **(0.6638, 0.8342)**.

5. A 95% *family* CI for the parameter β_2
   (also accounting for β_1; that is, $g = 2$) is given by
   $b_2 \pm t(1-\alpha/2g, n-p)s\{b_2\} = 0.044 \pm 2.84(1.15479)$
   (circle one) **(1.248, 15.128)** / **(−3.231, 3.319)** / **(−0.2144, 0.2144)**.

6. *One–At–A–Time Versus Simultaneously.*
   The one–at–a–time CIs for the parameters β_1 and β_0 are (choose one) **wider** / **narrower** than the simultaneous (family) CIs for the parameters β_1 and β_2.

7. *Test of β_2.*
   Test if β_2 (only β_2 and not β_1) is *not* zero at $\alpha = 0.05$.

   (a) *Statement.*
   The statement of the test is (check none, one or more):

   i. $H_0: \beta_1 = \beta_2 = 0$ versus $H_a: \beta_1 > 0, \beta_2 \neq 0$.
   ii. $H_0: \beta_1 = 0$ versus $H_a: \beta_1 \neq 0$
   iii. $H_0: \beta_2 = 0$ versus $H_a: \beta_2 \neq 0$.

   (b) *Test.*
   The p–value is given by (circle one) **0.00** / **0.022** / **0.97**.
   The level of significance is 0.05.

(c) *Conclusion.*
Since the p–value, 0.97, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that β_2 is zero.

6.7 Estimation of Mean Response and Prediction of New Observations

SAS programs: att5-6-7-read-respCI

Various statement and family (simultaneous) confidence intervals for the mean response[4], $\hat{Y}$, for a new observation or for a mean of m new observations are given by:

(Statement CI $E\{Y_h\}$, One X_h)

$$\begin{aligned} \hat{Y}_h &\pm ts\{\hat{Y}_h\} \\ t &= t(1-\alpha/2; n-p) \end{aligned}$$

(Simultaneous (Family) CI $E\{Y_h\}$, g X_h, Bonferroni or Working–Hotelling)

$$\begin{aligned} \hat{Y}_h &\pm Bs\{\hat{Y}_h\} \\ B &= t(1-\alpha/2g; p, n-p) \end{aligned}$$

or

$$\begin{aligned} \hat{Y}_h &\pm Ws\{\hat{Y}_h\} \\ W &= \sqrt{pF(1-\alpha; p, n-p)} \end{aligned}$$

(Statement CI of Mean of m New Observations $\hat{Y}_{h(new)}$, One X_h)

$$\begin{aligned} \hat{Y}_h &\pm ts\{predmean\} \\ t &= t(1-\alpha/2; p, n-p) \\ s\{predmean\} &= \sqrt{MSE(\frac{1}{m} + \mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h})} \end{aligned}$$

(Simultaneous (Family) CIs of m New Observations $\hat{Y}_{h(new)}$, g Different X_h, Bonferroni or Scheffe)

$$\begin{aligned} \hat{Y}_h &\pm Bs\{pred\} \\ B &= t(1-\alpha/2g; p, n-p) \end{aligned}$$

[4]Previously, we looked at the regression parameters, β_1 and β_2.

or

$$\begin{aligned} \hat{Y}_h &\pm Bs\{pred\} \\ S &= \sqrt{gF(1-\alpha; g, n-p)} \end{aligned}$$

where

$$s\{\hat{Y}_h\} = \sqrt{\mathbf{X_h' s^2\{b\} X_h}} = \sqrt{MSE(1 + \mathbf{X_h'(X'X)^{-1}X_h})}$$

Exercise 6.7 (Estimation of Mean Response and Prediction of New Observations)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

Let two new observations be

$$\mathbf{X}_{h1} = \begin{bmatrix} 1 \\ X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 25 \\ 29 \end{bmatrix}, \quad \mathbf{X}_{h2} = \begin{bmatrix} 1 \\ X_{12} \\ X_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 15 \end{bmatrix}$$

1. *Statement CI $E\{Y_h\}$, One X_h.*
   Determine a 95% CI for $E\{Y_h\}$ at $\mathbf{X}_{h1}$ and $\mathbf{X}_{h2}$.
   From TI–83,
   $t(1-\alpha/2; n-p) = t(1-0.05/2; 10-3) = t(0.975; 7) = 2.365$
   (INVT 7 ENTER 0.975 ENTER)
   From SAS,

   (a) $X_{11} = 25$, $X_{12} = 29$:
   $\hat{Y}_1 =$ (circle one) **89.5057** / **91.5832** / **98.9833**.
   and $s\{\hat{Y}_1\} =$ (circle one) **3.25009** / **4.56676** / **5.77882**.
   and so $\hat{Y}_1 \pm ts\{\hat{Y}_1\} = 89.5057 \pm 2.365(3.25009) = (81.82, 97.19)$

   (b) $X_{21} = 9$, $X_{22} = 15$:
   $\hat{Y}_2 =$ (circle one) **76.9024** / **91.5832** / **98.9833**.
   and $s\{\hat{Y}_2\} =$ (circle one) **3.25009** / **4.56676** / **5.21048**.
   and so $\hat{Y}_2 \pm ts\{\hat{Y}_2\} = 76.9024 \pm 2.365(5.21048) = (64.58, 89.23)$

2. *Simultaneous (Family) CI $E\{Y_h\}$, g X_h, Bonferroni and Working–Hotelling.*
   Determine the *most efficient* 95% family CI for $E\{Y_h\}$ at $\mathbf{X}_{h1}$ and $\mathbf{X}_{h2}$.
   From TI–83,
   Bonferroni:
   $B = t(1-\alpha/2g; n-p) = t(1-0.05/2(2); 10-3) = t(0.9875; 7) = 2.841$
   (INVT 7 ENTER 0.9875 ENTER)

Working–Hotelling:
$W = \sqrt{pF(1-\alpha; p, n-p)} = \sqrt{3F(1-0.05; 3, 10-3)} = 3.611$
(INVF 3 ENTER 7 ENTER 0.95 ENTER,
then multiply by 3 and find the square root)
Since $W = 3.611 > B = 2.841$, use B because Bonferroni gives narrower (*more* efficient) CIs than Working–Hotelling.
From SAS,

(a) $X_{11} = 25$, $X_{12} = 29$:
$\hat{Y}_1 =$ (circle one) **89.5057** / **91.5832** / **98.9833**.
and (STDP) $s\{\hat{Y}_1\} =$ (circle one) **3.25009** / **4.56676** / **5.77882**.
and so $\hat{Y}_1 \pm Bs\{\hat{Y}_1\} = 89.5057 \pm 2.841(3.25009) = (80.27, 98.74)$

(b) $X_{21} = 9$, $X_{22} = 15$:
$\hat{Y}_2 =$ (circle one) **76.9024** / **91.5832** / **98.9833**.
and (STDP) $s\{\hat{Y}_2\} =$ (circle one) **3.25009** / **4.56676** / **5.21048**.
and so $\hat{Y}_2 \pm Bs\{\hat{Y}_2\} = 76.9024 \pm 2.841(5.21048) = (62.10, 91.71)$

3. *Statement PI of the Mean of m New Observations , One X_h.*
Determine a 95% CI of the mean of $m = 3$ new observations, all at $\mathbf{X}_{h1}$.
From TI–83,
$t(1-\alpha/2; n-p) = t(1-0.05/2; 10-3) = t(0.975; 7) = 2.365$
(INVT 7 ENTER 0.975 ENTER)
From SAS,
at $X_{11} = 25$, $X_{12} = 29$:
$\hat{Y}_1 =$ (circle one) **89.5057** / **91.5832** / **98.9833**.
and $s\{predmean\} =$ (circle one) **3.25009** / **4.56676** / **5.8892781**.
and so $\hat{Y}_1 \pm ts\{predmean\} = 89.5057 \pm 2.365(5.8892781) = (80.31, 98.70)$

4. *Simultaneous (Family) PIs of m New Observations, g Different X_h, Bonferroni and Scheffe.*
Determine the *most efficient* 95% family CIs for $g = 2$ *new observations*[5] at $\mathbf{X}_{h1}$ and $\mathbf{X}_{h2}$.
From TI–83,
Bonferroni:
$B = t(1-\alpha/2g; n-p) = t(1-0.05/2(2); 10-3) = t(0.9875; 7) = 2.841$
(INVT 7 ENTER 0.9875 ENTER)
Scheffe:
$S = \sqrt{gF(1-\alpha; g, n-p)} = \sqrt{2F(1-0.05; 2, 10-3)} = 3.078$
(INVF 2 ENTER 7 ENTER 0.95 ENTER,
then multiply by 2 and find the square root)
Since $S = 3.078 > B = 2.841$, use B because Bonferroni gives narrower (*more*

[5]Previously, $\mathbf{X}_{h1}$ and $\mathbf{X}_{h2}$ where treated as current observations, rather than new observations, as they are here.

efficient) CIs than Scheffe.
From SAS,

(a) $X_{11} = 25$, $X_{12} = 29$:
$\hat{Y}_1 =$ (circle one) **89.5057** / **91.5832** / **98.9833**.
and (STDI) $s\{pred\} =$ (circle one) **3.25009** / **4.56676** / **9.10629**.
and so $\hat{Y}_1 \pm Bs\{pred\} = 89.5057 \pm 2.841(9.10629) = (63.63, 115.38)$

(b) $X_{21} = 9$, $X_{22} = 15$:
$\hat{Y}_2 =$ (circle one) **76.9024** / **91.5832** / **98.9833**.
and (STDI) $s\{pred\} =$ (circle one) **3.25009** / **4.56676** / **9.97550**.
and so $\hat{Y}_2 \pm Bs\{pred\} = 76.9024 \pm 2.841(9.97550) = (48.56, 105.24)$

6.8 Diagnostics and Remedial Measures

SAS programs: att5-6-8-read-diagnos

The diagnostics and remedial measures used for simple linear regression also apply in the multiple linear regression case. Some diagnostic measures are given below.

- *Univariate information* on the predictor variables and response variable, such as histograms, stem and leaf plots, dot plots, percentiles and the box plots can provide useful preliminary information.
- *Scatter plots and the correlation matrix* between all predictors and the response variable identify if a linear relationship (scatter plot) or an association (correlation) between response and predictor exists or not.
- *Residual plots* for all predictors and the response variable check for linearity, constant variance (normality, independent) of the error terms, outliers and any missing predictors (such as interaction predictors).
- *Tests* of normality, error variance (modified Levene and Breusch–Pagan) and lack of fit are also possible.

Remedial measures, again, consist of transforming either X or Y or both to straighten out nonlinear data or to correct for non constant error variance and non–normality.

Exercise 6.8 (Diagnostics and Remedial Measures)

illumination, X_{i1}	9	7	11	16	21	19	23	29	31	33
noise, X_{i2}	15	20	17	22	24	26	28	30	29	37
ability to read, Y	70	70	75	88	91	94	100	92	90	85

- *Univariate Information.*
  True / **False**
  The stem and leaf plots indicate a fairly even distribution of X_{i1} and X_{i2} and that there are no outliers.

- *Scatter plots and the Correlation Matrix.*
  True / **False**
  The various scatter plots indicate that the response variable is fairly weakly associated with the two predictor variables (which is "bad" because we want the response to depend on the predictor variables) and also, the two predictor variables are fairly strongly associated with one another (which is also "bad" because we do not want the two predictor variables to depend on one another, but to be independent or one another, or additive).

- *Residual plots*
  True / **False**
  The various residual plots indicate there is a pattern in the residual versus illumination plot (which is "bad" because it indicates variable variance in the error) although there does not appear to be any outliers (which is "good").

- *Tests of normality.*
  True / **False**
  The normal probability plot is linear (which is "good" because this indicates the error variance is normally distributed.)
  The Levene test of error variance confirms the normal probability plot because of the insignificant p–value of 0.8690.

6.9 An Example–Multiple Regression With Two Predictor Variables

This is an example which ties together a number of the techniques discussed in this chapter about multiple regression with two predictor variables.