AI6123 Time Series Analysis

Chapter 2. Multiple Regression Model

The objective of regression analysis is to exploit the relationship between two (or more) variables so that we can gain information about one of them through knowing values of the other(s).

1 A motivating example

Verizon Wireless is an American telecommunications company. In the first project, the company used its customer databases in order to develop a model to predict which of its customers were most likely to defect to another provider at the expiration of their current contracts based on the following:

- the current plan a customer had,
- a customer's historical calling patterns,
- the number and the type of services requests made by a customer.

The second project

- used the model developed in the first project to create samples of customers likely to leave at the end of the current contract.
- offered each of these samples a different experimental new plan offer.
- tracked which customers in each segment who accepted the offer.
- the data generated from these experimental samples were combined with the customer calling pattern and service request data to create a second set of models.
- determine the best new plan offer to make to a customer who is likely to leave Verizon before the current contract expires.

The outcomes of the project are as follows. This database marketing project helped:

- decrease its attrition rate from 2 percent per month to 1.5 percent per month (a reduction of 25 percent from the original attrition rate),
- The value of the reduction in churn is roughly $700M per year.
- the company's direct mail budget form "churning mailings" fell 60 percent since the promotional mailings are now highly targeted.

2 A case in Medicine

EXAMPLE 1 The following data set records the plasma levels of total cholesterol (in mg/ml) of 24 patients with hypercholesterolemia admitted to a hospital:

3.5	1.9	4.0	2.6	4.5	3.0	2.9	3.8	2.1	3.8	4.1	3.0
2.5	4.6	3.2	4.2	2.3	4.0	4.3	3.9	3.3	3.2	2.5	3.3

■

Question: Predict the cholesterol level of the next patient to be admitted to the hospital with hypercholesterolemia.
Intuitive answer: Use the average of the 24 observations: 3.354 (horizontal reference line in Fig. 1(a)). Observations scatter around the average but are subject to considerable

[The above is justifiable if the observations are i.i.d. for example. In the absence of further information, this seems to be the best we can do.]

Suppose the hospital has also collected data on the **ages** of the 24 patients:

46	20	52	30	57	25	28	36	22	43	57	33
22	63	40	48	28	49	52	58	29	34	24	50

Each observation (corresponding to each patient) consists of values of two *variables*:

(X, Y) = (age, cholesterol level).

We can see a strong linear relationship between the two variables. As far as prediction is concerned, it seems more reliable to assume a linear function relating age and cholesterol level of hypercholesterolemia patients, and predict the next patient's cholesterol level by his/her age.

- ▷ Fig. 1(b) *fits* a sloped straight line to the scatterplot by the *least squares method* (to be discussed later). This straight line summarizes the relationship between cholesterol level and age and can be used for predicting future

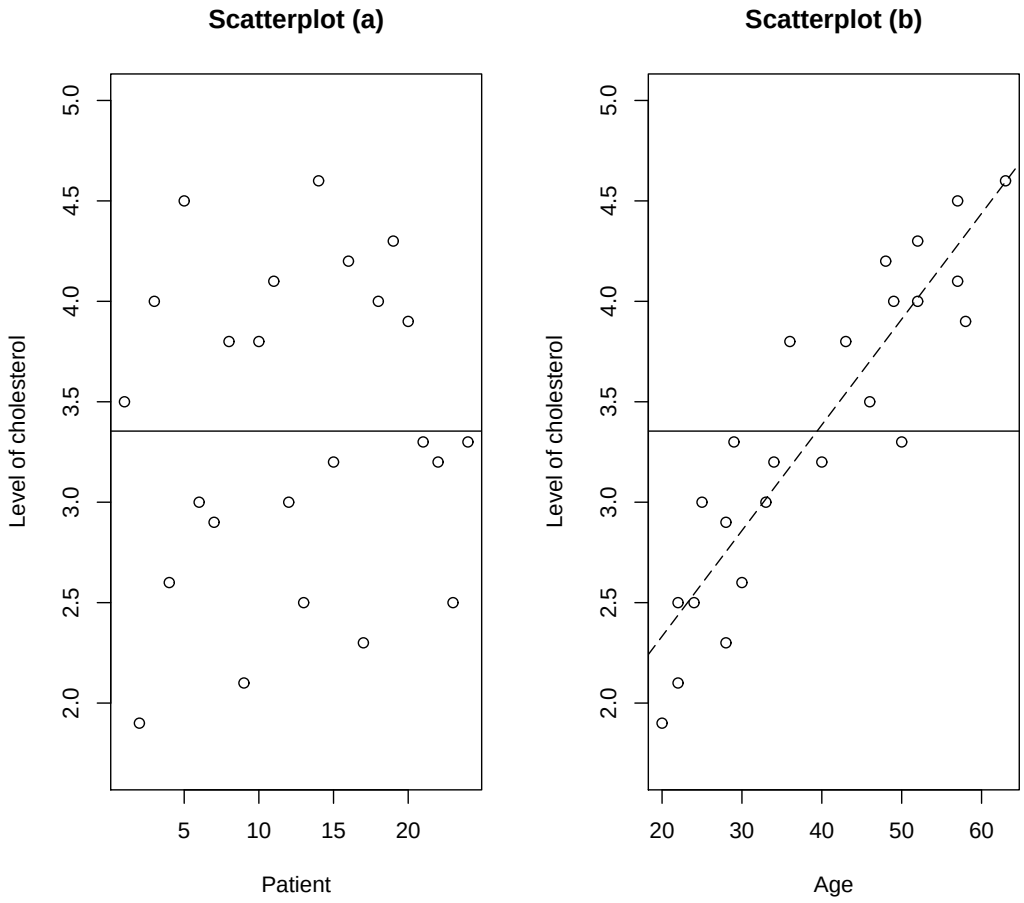

Figure 1: Plasma levels of total cholesterol (in mg/ml).

patients' cholesterol levels. Compared to Fig. 1(a), fluctuations of the 24 observations around the sloped straight line are much smaller. A function linear in age (the sloped straight line) can better account for the observed variation in cholesterol level than a simple constant function (the horizontal line).

▷ The above example highlights the importance in data analysis of collecting data on some other variables (e.g. age) relevant to the main variable of interest (e.g. cholesterol level) in order to obtain a model which can better explain the observed variation in the main variable.

3 Elements of Regression

We first introduce some terminology in regression. Generally speaking, there are two types of variables: quantitative and qualitative.

DEFINITION 1 **Quantitative variables** can be measured in a numerical form: e.g. age, GPA, income, time, temperature etc. **Qualitative variables** are not numerical in nature: e.g. gender, categorized age, education level, type of crime committed, style of cuisine served in a restaurant etc.

DEFINITION 2 The variable to be predicted, y, is called **the response variable**. Or, the variable which is of our primary interest is called the response variable (output variable, outputs, Y-variables or dependent variable). The remaining variables are called **predictor variables** (input variable, inputs, X-variables, regressors or independent variable).

DEFINITION 3 (a) Functional Relationships- The value of the dependent variable Y can be computed exactly if we know the value of the independent variable X. (e.g., Y=2X).

(b) Statistical Relationships-Not a perfect or exact relationship. The expected value of the response variable Y is a function of the explanatory or predictor variable X.

Usually

$$\text{Response variable} = \text{Model function}(Ey) + \text{Random error}. \tag{3.1}$$

When systematic component (Model function) is a linear function of parameters, it is also so-called Linear Regression.

4 Multiple Linear Regression

Multiple Linear Regression is used to explore the possible relationship between one response variable and (one or more than one) predictor variable(s). Obviously, *Simple Linear Regression* is a special case of *Multiple Linear Regression*.

Consider

Y: response variable, $X_1, \ldots, X_p$: predictor variables.

A sample of n observations is observed in the form of $(y_i, x_{i1}, \ldots, x_{ip})$ for the ith observation:

Y	X_1	X_2	$\cdots$	X_p
y_1	x_{11}	x_{12}	$\cdots$	x_{1p}
y_2	x_{21}	x_{22}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_n	x_{n1}	x_{n2}	$\cdots$	x_{np}

DEFINITION 4 A *multiple linear regression model* (MLR) assumes:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \tag{4.1}$$

where

- $X_1, \ldots, X_p$ are non-random variables,
- $\epsilon_1, \ldots, \epsilon_n$ are *i.i.d.* with $\epsilon_i \sim N(0, \sigma^2)$,
- Note that, hence, $y_1, ..., y_n$ are independent, and normally distributed..

Equivalently, in matrix form,

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad \epsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n),$$

where

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \text{ and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Notations:

- $\beta_0, \beta_1, \ldots, \beta_p$: unknown regression coefficients;
- β_j is associated with effect of X_j on Y, and also called the coefficient of X_j;
- $\mathbf{X}$: $n \times (p+1)$ design matrix with known entries;
- Assume the columns of $\mathbf{X}$ are linearly independent, so that $(\mathbf{X}'\mathbf{X})^{-1}$ exists;

▷ $\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$.

Note: The word 'linear' in a linear model refers to the property that $E[\mathbf{Y}]$ is 'linear' in the regression coefficients $\beta_0, \ldots, \beta_p$, but not necessarily in each explanatory variable.

EXAMPLE 2 Linear models:

(i) SLR $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$;

(ii) Polynomial regression

$$y_i = \beta_0 + \beta_1 z_i + \beta_2 z_i^2 + \epsilon_i,$$

predictor variables: $x_{i1} = z_i$, $x_{i2} = z_i^2$; Then the above model becomes

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i.$$

(iii) Interaction effects

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i.$$

predictor variables: $x_{i3} = x_{i1} x_{i2}$.

(iv) $y_i = \beta_0 + \beta_1 \sin 2\pi z_{i1} - \beta_2 \log z_{i2} + \epsilon_i$,
predictor variables: $x_{i1} = \sin 2\pi z_{i1}$, $x_{i2} = -\log z_{i2}$. ■

We can use linear regression models to deal with almost any "function" of a predictor variable.

EXAMPLE 3 NOT linear models:

(i) $y_i = \exp\{\beta_0 + \beta_1 x_i + \epsilon_i\}$;

(ii) $y_i = 1/(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i)$. ■

But one should note that after taking transformation, the above two models can still be converted to multiple regression models.

5 Estimation of Regression Coefficients

In the multiple regression model, the least squares (LS) estimates $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$ for $\beta_0, \beta_1, \cdots, \beta_p$ are chosen to minimize

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2. \tag{5.1}$$

Using calculus the $p+1$ first order conditions are obtained as follows.

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) = 0\\
&\sum_{i=1}^{n} x_{i1}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) = 0\\
&\qquad\qquad \cdots\\
&\sum_{i=1}^{n} x_{ip}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) = 0. \qquad (5.2)
\end{aligned}$$

Since the above expression is cumbersome we would like to have a neat solution.

To this end, by (4.1) rewrite (5.1) as

$$\sum_{i=1}^{n}\epsilon_i^2 = (\epsilon_1\epsilon_2\cdots\epsilon_p)\begin{pmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_p\end{pmatrix} = \epsilon'\epsilon = (\mathbf{Y}-\mathbf{X}\beta)'(\mathbf{Y}-\mathbf{X}\beta). \qquad (5.3)$$

To minimize (5.3) we need to take derivative with respect to the vector β. It is like a quadratic function. Using the chain rule we find

$$\frac{d}{d\beta}\Big[(\mathbf{Y}-\mathbf{X}\beta)'(\mathbf{Y}-\mathbf{X}\beta)\Big] = -2\mathbf{X}'(\mathbf{Y}-\mathbf{X}\beta).$$

Setting this equal to zero and solving for β we obtain **the normal equation**

$$2\mathbf{X}'\mathbf{X}\beta - 2\mathbf{X}'\mathbf{Y} = \mathbf{0}.$$

It follows that

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \qquad (5.4)$$

6 Estimation of σ^2 in MLR

It seems reasonable to assume that the greater the variability of the random error ε (which is measured by its variance σ^2), the greater will be the errors in the estimation of the model parameters $\beta_0, \beta_1, \cdots, \beta_p$, and in the error of prediction when $\hat{y}$ is used to predict y for some values of x. Consequently, you should not be surprised, as we proceed through this chapter, to find σ^2 appears in the formulas for all confidence intervals and test statistics that we use.

In most practical situations, σ^2 will be unknown and we must use the data to estimate its value.

Recall the way of estimating σ^2 from i.i.d sample $Y_1, \cdots, Y_n$ with $EY_i = 0$ and $var(Y_i) = \sigma^2$.

▷ find

$$\sum_{i=1}^{n}(Y_i - \hat{EY_i})^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Square the difference between each observation and the estimate of its mean.

▷ divide by degrees of freedom

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

SLR model with $E(y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ and $var(y_i) = \sigma^2$, independent but not identically distributed. Let's do the same two steps.

▷ find

$$\sum_{i=1}^{n}(y_i - \hat{Ey_i})^2 = \sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}))^2 = SSE.$$

Square the difference between each observation and the estimate of its mean. Here SSE denotes the error sum of squares.

▷ divide by degrees of freedom

$$s^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i + \cdots + \hat{\beta}_p x_{ip}))^2 = \frac{SSE}{n-p-1}.$$

s^2 is an unbiased estimate of σ^2, that is

$$E(s^2) = \sigma^2.$$

7 Fitted values and residuals

Denote the fitted value of y_i by $\hat{y}_i$ and the residual by $e_i = y_i - \hat{y}_i$. Let

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_p \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \cdots + \hat{\beta}_p x_{1p} \\ \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \cdots + \hat{\beta}_p x_{2p} \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \cdots + \hat{\beta}_p x_{np} \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix}.$$

The fitted values are then represented by

$$\hat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} \tag{7.1}$$

and the residual value by

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\widehat{\beta}. \tag{7.2}$$

8 An example

EXAMPLE 4 Hotel chain wants to decide where to open next branch. The hotel collects the data of 48 existing hotel branches for 3 years across the following dimensions:

- Profit (Response Variable)
- B. Competetitive Properties (Predictor/Explanatory Variable)
  (1) Hotel rooms in vicinity and average room rate
  (2) Demand generators (hospital/ offices/ colleges in 4 mile radius)
  (3) Demographics (location population, unemployment rate and median family income)
  (4) Market awareness (years in business and state population per hotel)
  (5) Physical consideration (accessibility, distance to downtown and sign visibility)

The hotel used regression as a tool to analyze the positive and negative relation between these variables and defined an equation that best explains the relationships among various dimensions. To obtain the scatterplot matrix for all these variables, just give the R command:

```
pairs(hotel, pch=19)
```

The scatterplot matrix will appear on the graphics device in R.

This matrix enables you to tell whether the response variable appears to have any association with any of the predictor variables, and if any two of the predictor variables appear to be correlated. For the categorical variable Holiday the Scatterplot matrix is not very helpful.

It turns out that

$$\text{Predicted Profitability}$$
$$= \beta_0 - \beta_1 \cdot StatePop + \beta_2 \cdot Price - \beta_3 \cdot \sqrt{MedIncome} + \beta_4 \cdot ColStudent$$

This equation predicts that profitability will increase

- when room rate and the number of college students increase
- when state population and median income decreases. ■

This model is currently being used by the hotel for site screening process.

Table 1: Environmental data

ozone	radiation	temperature	wind
3.45	190.00	67.00	7.40
3.30	118.00	72.00	8.00
2.29	149.00	74.00	12.60
2.62	313.00	62.00	11.50
2.84	299.00	65.00	8.60
2.67	99.00	59.00	13.80
2.00	19.00	61.00	20.10
2.52	256.00	69.00	9.70
2.22	290.00	66.00	9.20
2.41	274.00	68.00	10.90
2.62	65.00	58.00	13.20
2.41	334.00	64.00	11.50
3.24	307.00	66.00	12.00
1.82	78.00	57.00	18.40
3.11	322.00	68.00	11.50
2.22	44.00	62.00	9.70
1.00	8.00	59.00	9.70
2.22	320.00	73.00	16.60
1.59	25.00	61.00	9.70
3.17	92.00	61.00	12.00
2.84	13.00	67.00	12.00
3.56	252.00	81.00	14.90
4.86	223.00	79.00	5.70
3.33	279.00	76.00	7.40
3.07	127.00	82.00	9.70
4.14	291.00	90.00	13.80
3.39	323.00	87.00	11.50
2.84	148.00	82.00	8.00
2.76	191.00	77.00	14.90
3.33	284.00	72.00	20.70

9 An example

A data set is taken from an environmental study that measured four variables for 30 consecutive days, which contains 30 observations (rows) and 4 variables (columns), is shown in Table 1:

ozone (Y) — ozone surface concentration of ozone in New York, in parts per million;

radiation (X_1) — solar radiation;

temperature (X_2) — observed temperature, in degrees Fahrenheit;

wind (X_3) — wind speed, in miles per hour.

- ▷ Figure 2 displays a scatter plot matrix of the data, providing visual information about pairwise relationships among the 4 variables.

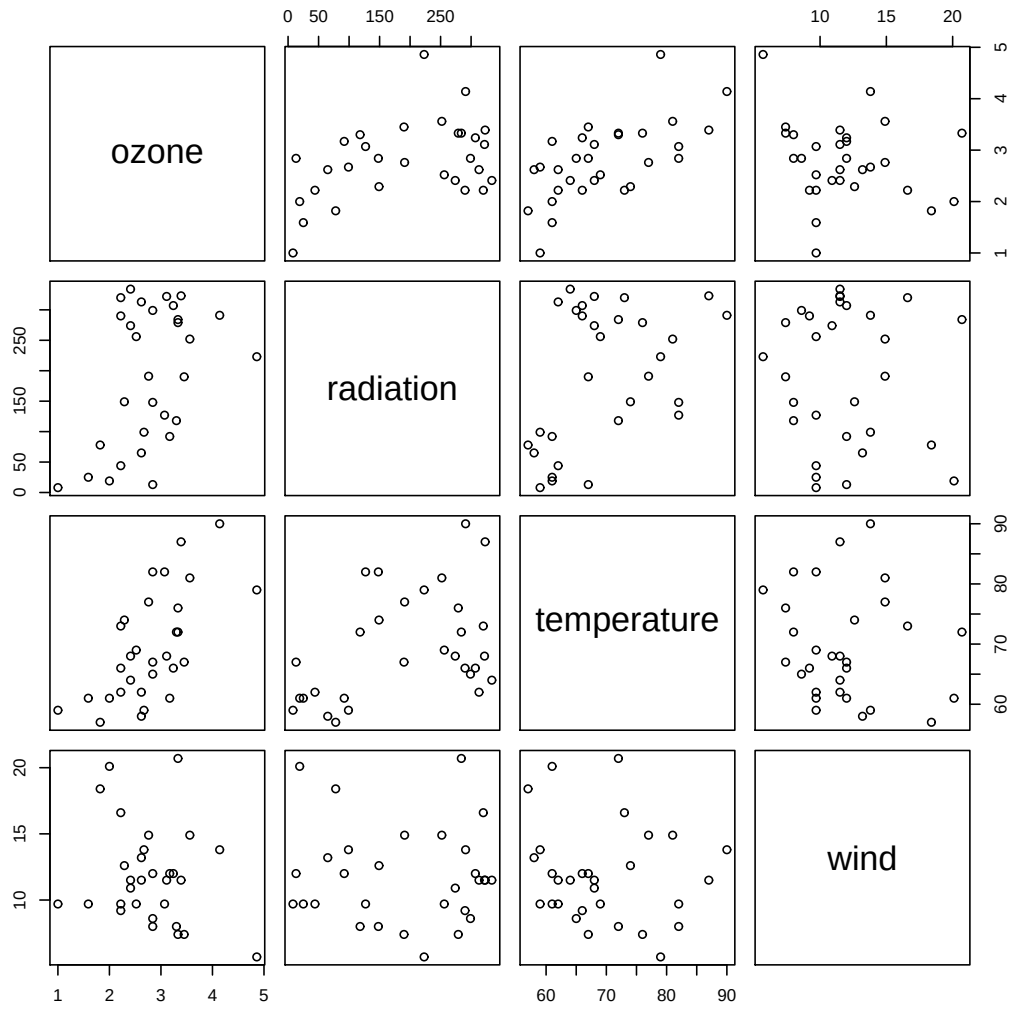

Figure 2: Scatter plot matrix for the environmental data

- To study the variation of ozone concentration (Y), we fit an MLR model to the data using 3 explanatory variables: radiation (X_1), temperature (X_2) and wind (X_3).

- The 30×4 design matrix $\mathbf{X}$ equals

$$\mathbf{X} = \begin{bmatrix} 1 & 190.00 & 67.00 & 7.40 \\ 1 & 118.00 & 72.00 & 8.00 \\ 1 & 149.00 & 74.00 & 12.60 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

▷ Then the LSE is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} -0.295271027 \\ 0.001305816 \\ 0.045605744 \\ -0.027843496 \end{bmatrix}.$$

▷ According to the signs of estimates, it seems that ozone concentration increases as radiation and temperature increase, but is reduced by wind. Statistical tests have to be conducted if we wish to have a more rigorous study.

10 Model checks for adequacy

10.1 ANOVA table

Hence, formulas for sums of squares are given as follows.

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$$

$$SSE = \sum (y_i - \hat{y}_i)^2$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2 = \sum \hat{y}_i^2 - n\bar{y}^2.$$

From the above matrix expression for sums of squares one may verify the following partition

$$S_{yy} = SSR + SSE.$$

Table below shows these analysis of variance results.

Source	df	SS	MS	F	p-value
Regression	p	$SSR = \sum(\hat{y}_i - \bar{y})^2$	$\mathrm{MS}_{\mathrm{Reg}}$	$\mathrm{MS}_{\mathrm{Reg}}/s^2$	
Residual	$n-p-1$	$SSE = \sum(y_i - \hat{y}_i)^2$	s^2		
Total	$n-1$	$S_{yy} = \sum(y_i - \bar{y})^2$			

where

$$\mathrm{MS}_{\mathrm{Reg}} = \frac{SSR}{df_R}.$$

The difference SSR measures how effective the variables $X_1, ..., X_p$ are to explain the variation in the response Y collectively.

10.2 F test

It can be proved that $E(s^2) = \sigma^2$, as for SLR. The expectation of MS_{Reg} is σ^2 plus a quantity that is nonnegative. For instance, when $p = 2$, we have

$$E(MS_{Reg}) = \sigma^2 + \frac{1}{2}\Big[\beta_1^2 \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2 + \beta_2^2 \sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2 + 2\beta_1\beta_2 \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)\Big].$$

Note that if both β_1 and β_2 are equal to zero, then $E(MS_{Reg}) = \sigma^2$. Otherwise $E(MS_{Reg}) > \sigma^2$.

This suggests that a comparison of MS_{Reg} and s^2 is useful for testing whether there is a regression relation between the response variable y and the predictor variables $x_1, \cdots, x_p$.

- ▷ The formal F test for the significance of the MLR is equivalent to testing

$$H_0 : \beta_1 = \cdots = \beta_p = 0 \quad \text{vs} \quad H_1 : \text{at least one} \quad \beta_j \neq 0$$

- ▷ $F = \frac{\text{MS}_{\text{Reg}}}{s^2} \sim F(p, n-p-1)$.

  Rejection rule: reject H_0 iff the p-value $P(F_{p,n-p-1} > F)$ is smaller than α,
  or equivalently, iff $F > F^{(\alpha)}_{p,n-p-1}$.

Then, by Cochran's Theorem, we have that SSE and SSR are independent, and

$$SSR \sim \chi^2_p \quad \text{and} \quad SSE \sim \chi^2_{n-p-1}.$$

10.3 R^2 statistic

R^2 statistic:

$$R^2 = \frac{SSR}{S_{yy}}.$$

Adding more variables will make R^2 go up because SSE will never become larger with more predicator variables and SS_y is always the same for a given set of responses ; so cannot really use R^2 to determine whether a variable should be added.

Adjusted R^2 statistic:

$$R^2_a = 1 - \frac{SSE/(n-p-1)}{S_{yy}/(n-1)}.$$

Sum of squares are adjusted by degree of freedom. **Note that:**

- R_a^2 is mainly used to measure the performance of the fitted model on different data sets, especially with **different sample size**.
- Generally, for a good model, the R^2 should not be small ($< 60\%$) or not be large ($> 95\%$).

EXAMPLE 5 **Example 1** (cont'd)
ANOVA table:

Source	df	SS	MS	F	p-value
Regression	3	7.6685	2.5562	7.3244	
Residual	26	9.0738	0.3490		
Total	29	16.7423			

- Calculating using computer,

  $$\text{p-value} = P(F_{3,26} > 7.3244) = 0.001026,$$

  which is very small $\Longrightarrow$ MLR highly significant.
- For a size α test, we read the critical values from standard statistical tables or computer software. For example,

α	Critical value $F_{3,26}^{(\alpha)}$
5%	2.9752
1%	4.6365

  In either case, the F-ratio 7.3244 is bigger than the critical value and we should reject H_0 at the stated significance level.
- The coefficient of determination is $R^2 = 7.6685/16.7423 = 0.4580$, which shows that the MLR does not give a good fit (even though it is highly significant!) ■

11 Inference for Regression Coefficients

Tests model significance give no indication of which variable(s) in particular are important. So we need to consider tests for individual regression coefficients.

Since $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$ we have $\widehat{\beta}_i \sim N_{p+1}(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. It follows that

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2(\mathbf{X}'\mathbf{X})_{ii}^{-1})$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{ii}$ denotes the i diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. Note that σ^2 is unknown in practice. As in SLR one may verify that $s^2 = SSE/(n-p-1) = MSE$ is an unbiased estimator for σ^2. Moreover we have

$$\frac{\hat{\beta}_i - \beta_i}{s.e.(\beta_i)} \sim t_{n-p-1},$$

where $s.e.(\beta_i) = s\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{i+1,i+1}}$.

To test

$$H_0 : \beta_i = 0 \quad H_\alpha : \beta_i \neq 0.$$

The test statistic is

$$t^* = \frac{\hat{\beta}_i}{s.e.(\beta_i)}$$

and the decision rule is

$$\text{if} \quad |t^*| > t_{n-p-1,\alpha/2}, \quad \text{reject} \quad H_0.$$

The confidence interval is

$$(\hat{\beta} - t_{n-p-1,\alpha/2}s.e.(\beta_i, \hat{\beta} + t_{n-p-1,\alpha/2}s.e.(\beta_i),$$

EXAMPLE 6 **Example 1** (cont'd)
Suppose we want to see if temperature can be removed from the full model, i.e. to test $H_0 : \beta_2 = 0$ against $H_0 : \beta_2$ unrestricted.

Note that

$$\hat{\beta}_2 = 0.04561 \quad \text{and} \quad s = \sqrt{0.3490} = 0.5908.$$

Calculate

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 2.8559225951 & 0.0006173356 & -0.0353506868 & -0.0409097295 \\ 0.0006173356 & 0.0000033414 & -0.0000180726 & 0.0000000141 \\ -0.0353506868 & -0.0000180726 & 0.0005338541 & 0.0001439104 \\ -0.0409097295 & 0.0000000141 & 0.0001439104 & 0.0026139210 \end{bmatrix},$$

■

so that

$$\text{s.e.}(\hat{\beta}_2 - \beta_2) = s\sqrt{0.0005338541} = 0.01365.$$

The 95% confidence interval for β_2 is

$$[\, 0.04561 - 0.01365(2.0555),\ 0.04561 + 0.01365(2.0555)\,] = [\, 0.01755,\ 0.07366\,],$$

where $t_{26}^{(0.025)} = 2.0555$.

The point zero falls outside the above interval and so we reject H_0 at the 5% significance level. It seems that including temperature in the model significantly improves our fit.

12 Logistic regression

Often, we are interested in predicting the probability that a customer will respond in a favorable way (e.g. that his or her response will be "yes" to our offer).

- ▷ Logistic regression is the preferred method for developing a regression-like predictive model when the target variable is binary.
- ▷ The predicted probability of a favorable response should fall between zero and one.
- ▷ However, using linear regression, we will often predict probabilities that fall outside the zero to one range.

Denote by p_i the probability that a customer will respond in a favorable way (Y takes the value 1). Then

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i.$$

The estimator of p_i is

$$\hat{p}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}}}{1 + \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}}$$

after obtaining the maximum likelihood estimates $\hat{\beta}_j$ of β_j.

R code for fitting Multiple Regression:

```
> mlr<-lm(y~x1+x2+x3,data=ozone)
> summary(mlr)
```

Output for multiple regression model:

```
Call:
lm(formula = y ~ x1 + x2 + x3, data = ozone)

Residuals:
     Min       1Q   Median       3Q      Max
-1.13583 -0.39280  0.00007  0.39270  1.41993


Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) -0.295271   0.998348  -0.296  0.76976
x1           0.001306   0.001080   1.209  0.23746
x2           0.045606   0.013650   3.341  0.00253 **
x3          -0.027843   0.030203  -0.922  0.36507

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5908 on 26 degrees of freedom
Multiple R-squared:  0.458,     Adjusted R-squared:  0.3955
F-statistic: 7.324 on 3 and 26 DF,  p-value: 0.001026
```

R codes for ANOVA table:

```
> anova(mlr)


Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
x1         1 3.1469  3.1469  9.0172 0.005845 **
x2         1 4.2250  4.2250 12.1062 0.001787 **
x3         1 0.2966  0.2966  0.8498 0.365073
Residuals 26 9.0738  0.3490
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R codes for logistic model:

```
model <- glm(Win ~ Distance, data = train, family="binomial")
```