Chapter 4
Multiple Regression Analysis
(Part 1)

Terry Dielman

Applied Regression Analysis:
A Second Course in Business and Economic Statistics, fourth edition

4.1 Using Multiple Regression

- In Chapter 3, the method of least squares was used to describe the relationship between a dependent variable *y* and an explanatory variable *x*.

- Here we extend that to two or more predictor variables, using an equation of the form:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

Basic Exploration

- In Chapter 3 our main graphic tool was the X-Y scatter plot.
- Exploratory graphics are a bit harder to produce here because they need to be multidimensional.
- Even if there were just two *x* variables a 3-D display is needed.

Estimation of Coefficients

We want an equation of the form:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

As before we use least squares. The coefficients b_0 b_1 b_2 ... b_k are determined by minimizing the sum of squared residuals.

Formulae are Very Complex

- Can show exact formula when k=1 (simple regression). Refer to Section 3.1.
- Few texts show the formulae for k=2 (the simplest of multiple regressions)
- Appendix D shows formula in matrix notation
- This is totally a computer problem

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i, \qquad i = 1, 2, \cdots, n$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

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

Find β to Minimize $Q = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = (\boldsymbol{y} - X\boldsymbol{\beta})^T (\boldsymbol{y} - X\boldsymbol{\beta})$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \mathbf{0} \longrightarrow \frac{\partial Q}{\partial \boldsymbol{\beta}} = -2X^T(\boldsymbol{y} - X\boldsymbol{\beta}) = \mathbf{0} \longrightarrow X^T X\boldsymbol{\beta} = X^T \boldsymbol{y}$$

Normal Equation

If $X^T X$ is a nonsingular matrix,

$$\hat{\boldsymbol{\beta}}_{LSE} = (X^T X)^T X^T \boldsymbol{y}$$

Sum of Squares (SS) in Matrix Form

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \boldsymbol{y}^T \left(I_n - \frac{J_n}{n} \right) \boldsymbol{y},$$

where $I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{n \times n}$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \boldsymbol{y}^T \left(P - \frac{J_n}{n} \right) \boldsymbol{y}$$

$J_n = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}_{n \times n}$

where $P = X(X^T X)^{-1} X^T$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \boldsymbol{y}^T (I_n - P) \boldsymbol{y}$$

ANOVA Table

Source	d.f.	S.S.	M.S.	Fo
Reg.	k	$\boldsymbol{y}^T \left(P - \frac{J_n}{n} \right) \boldsymbol{y}$	SSR/k	$\frac{MSR}{MSE}$
Error	n−k−1	$\boldsymbol{y}^T \left(I_n - P \right) \boldsymbol{y}$	$SSE/(n-k-1)$	
Total	n−1	$\boldsymbol{y}^T \left(I_n - \frac{J_n}{n} \right) \boldsymbol{y}$		

If $F_0 > F_{0.05,\, k,\, n-k-1}$, then Reject $H_0 : \beta_1 = \beta_2 = \cdots \beta_k = 0$

Example 4.1 Meddicorp Sales

n = 25 sales territories

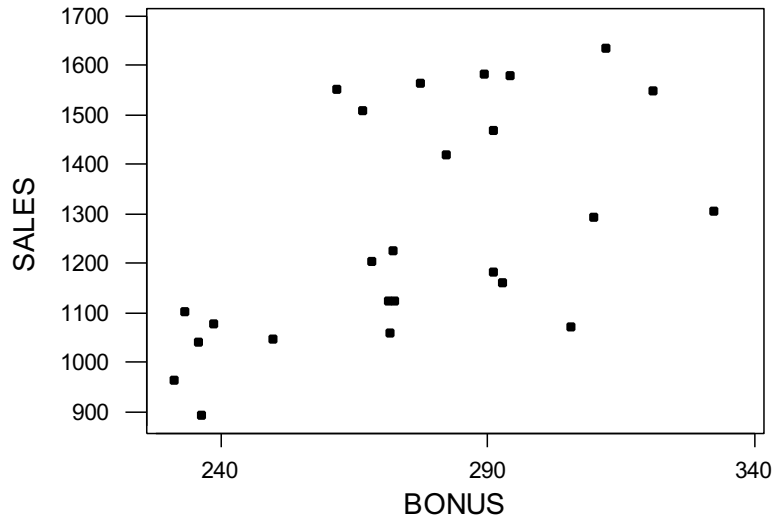
Y = Sales (1000$) in each territory

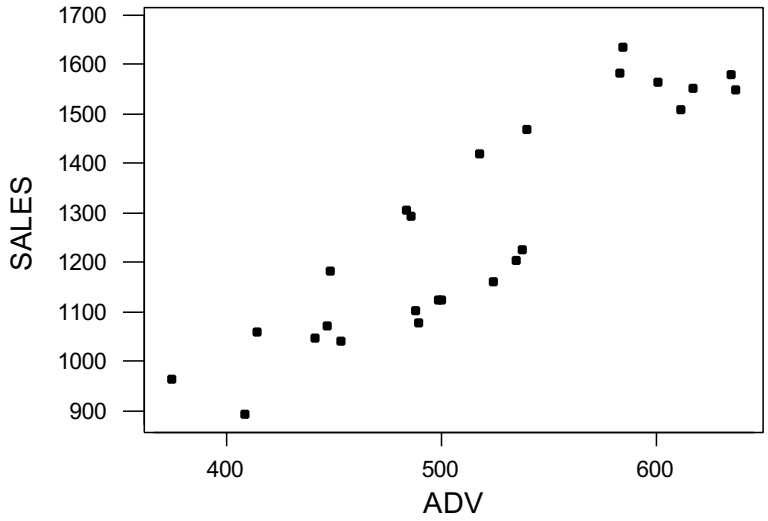
X_1 = Advertising (100$) in territory

X_2 = Amount of bonuses (100$) paid to salespersons in the territory

Data set MEDDICORP4

Plots and Correlation

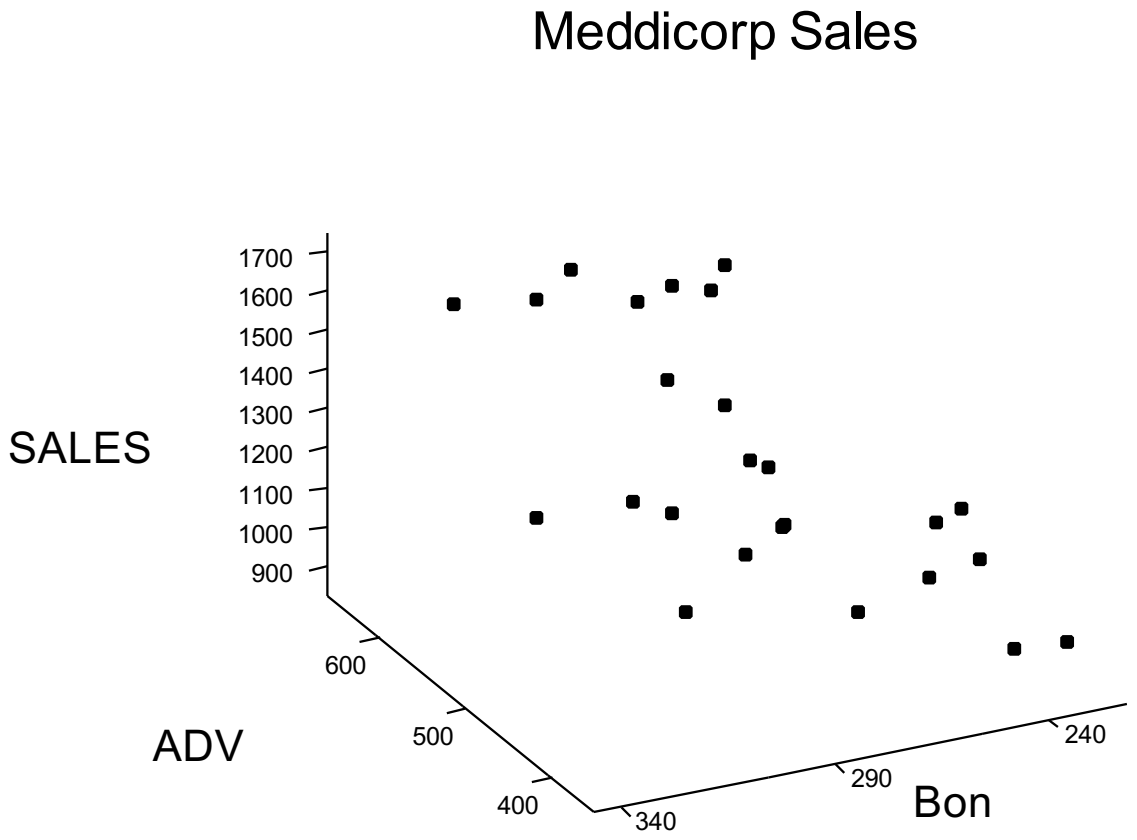




Correlations

	SALES	ADV
ADV	0.900	
BONUS	0.568	0.419

3D Graphics



Minitab Regression Output

The regression equation is
SALES = - 516 + 2.47 ADV + 1.86 BONUS

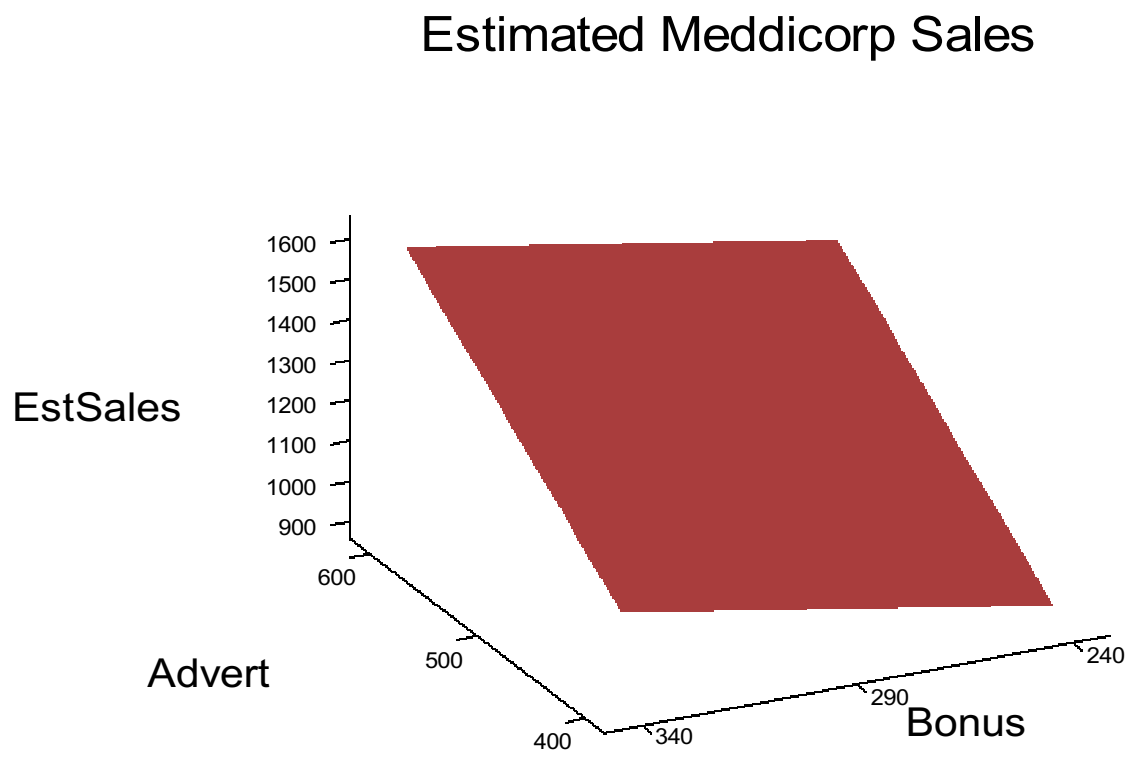
Predictor	Coef	SE Coef	T	P
Constant	-516.4	189.9	-2.72	0.013
ADV	2.4732	0.2753	8.98	0.000
BONUS	1.8562	0.7157	2.59	0.017

S = 90.75 R-Sq = 85.5% R-Sq(adj) = 84.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1067797	533899	64.83	0.000
Residual Error	22	181176	8235		
Total	24	1248974			

3D Surface Graph



The regression equation is
SALES = - 516 + 2.47 ADV + 1.86 BONUS

Interpretation of Coefficients

The regression equation is
SALES = - 516 + 2.47 ADV + 1.86 BONUS

- Recall that sales is in $1000s and advertising and bonus in $100s.
- If advertising is held fixed, sales increase $1860 for each $100 of bonus paid.
- If bonus were fixed, sales increase $2470 for each $100 spent on ads.

4.2 Inferences From a Multiple Regression Analysis

In general, the population regression equation involving K predictors is:

$$\mu_{y|x_1,x_2,\ldots,x_K} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K$$

This says the mean value of *y* at a given set of *x* values is a point on the surface described by the terms on the right-hand side of the equation.

4.2.1 Assumptions Concerning the Population Regression Line

An alternative way of writing the relationship is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i$$

where *i* denotes the ith observation and e_i denotes a random error or disturbance (deviation from the mean).

We make certain assumptions about the e_i.

Assumptions

1. We expect the average disturbance e_i to be zero so the regression line passes through the average value of y.
2. The e_i have constant variance σ_e^2.
3. The e_i are normally distributed.
4. The e_i are independent.

Inferences

- The assumptions allow inferences about the population relationship to be made from a sample equation.
- The first inferences considered are those about the individual population coefficients β_1 β_2 … β_K.
- Chapter 6 examines what happens when the assumptions are violated.

4.2.2 Inferences about the Population Regression Coefficients

If we wish to make an estimate of the effect of a change in one of the *x* variables on *y*, use the interval:

$$b_j \pm t_{n-K-1} s_{b_j}$$

this refers to the j^{th} of the *K* regression coefficients. The multiplier *t* is selected from the *t*-distribution with *n-K-1* degrees of freedom.

Tests About the Coefficients

A test about the marginal effect of x_j on y may be obtained from:

$$H_0: \ \beta_j = \beta_j^*$$
$$H_a: \ \beta_j \neq \beta_j^*$$

where β_j^* is some specific value that is relevant for the j^{th} coefficient.

Test Statistic

The test would be performed by using the standardized test statistic:

$$t = \frac{b_j - \beta^*_j}{s_{b_j}}$$

The most common form of this test is for the parameter to be 0. In this case the test statistic is just the estimate divided by its standard error.

Example 4.2 Meddicorp (Continued)

Refer again to the portion of the regression output about the individual regression coefficients:

Predictor	Coef	SE Coef	T	P
Constant	-516.4	189.9	-2.72	0.013
ADV	2.4732	0.2753	8.98	0.000
BONUS	1.8562	0.7157	2.59	0.017

This lists the estimates, their standard errors and the ratio of the estimates to their standard errors.

Tests For Effect of Advertising

To see if an increase in advertising expenditure affects sales, we can test:

H_0: $\beta_{ADV} = 0$ (An increase in advertising has no effect on sales)

H_a: $\beta_{ADV} \neq 0$ (Sales do change when advertising increases)

The df are $n\text{-}K\text{-}1 = 25\text{–}2\text{-}1 = 22$. At a 5% significance level, the critical point from the t-table is 2.074

Test Result

From the output we get:

$t = (2.4732 – 0)/.2753 = 8.98$

This is above the critical value of 2.074, so we reject H_0.

Note that we could also make use of the p-value (.000) for the test.

One-Sided Test on Bonus

We can modify the test to make it one sided

H_0: $\beta_{BONUS} = 0$ (Increased bonuses do not affect sales)

H_a: $\beta_{BONUS} > 0$ (Sales increase when bonuses are higher)

At a 5% significance level, the (one-sided) critical point is 1.717.

One-Sided Test Result

From the output we get:

t = 1.8562/.7157 = 2.59 which is > 1.717

We reject H_0 but this time make a more specific conclusion.

The listed p-value (.017) is for a two-sided test. For our one-sided test, cut it in half.

Interval Effect of advertising

Recall that sales are measured in 1000$ and ADV in 100$

$b_{adv} = 2.4732$ and has standard error = .2753

$2.4732 \pm 2.074(.2753) = 2.4732 \pm .5709$

$= 1.902$ to 3.044

Each $100 spent on advertising returns $1902 to $3044 in sales.

4.3 Assessing the Fit of the Model

Recall how we partitioned the variation in the previous chapter:

SST = Total variation in the sample of Y values

Split up into two components SSE, SSR

SSE = Error or unexplained variation

SSR = Explained by the Yhat function

4.3.1 The ANOVA Table and R^2

- These are the same statistics we briefly examined in simple regression.
- They are perhaps more important here because they measure how well all the variables in the equation work together.

```
S = 90.75       R-Sq = 85.5%     R-Sq(adj) = 84.2%

Analysis of Variance

Source            DF          SS          MS          F          P
Regression         2     1067797      533899      64.83      0.000
Residual Error    22      181176        8235
Total             24     1248974
```

R^2 – a Universal Measure of Fit

R^2 = SSR / SST = proportion of variation explained by the regression equation.

If multiplied by 100, interpret as %

If only one *x*, R^2 is square of correlation

For multiple, R^2 is square of correlation between the Y values and Y-hat values

For our example

```
S = 90.75       R-Sq = 85.5%     R-Sq(adj) = 84.2%

Analysis of Variance

Source            DF          SS          MS          F          P
Regression         2     1067797      533899      64.83      0.000
Residual Error    22      181176        8235
Total             24     1248974
```

R^2 = 1067797 / 1248974 = .85494

85.5% of the variation in sales in the 25 territories is explained by the different levels of advertising and bonus

Adjusted R^2

- If there are many predictor variables to choose from, the best R^2 is always obtained by throwing them all in the model.
- Some of these predictors could be insignificant, suggesting they contribute little to the model's R^2.
- Adjusted R^2 is a way to balance the desire for high R^2 against the desire to include only important variables.

Computation

The "adjustment" is for the number of variables in the model.

$$R^2_{adj} = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)}$$

Although regular R^2 may decrease when you remove a variable, the adjusted version may actually increase if that variable did not have much significance.

4.3.2 The *F* Statistic

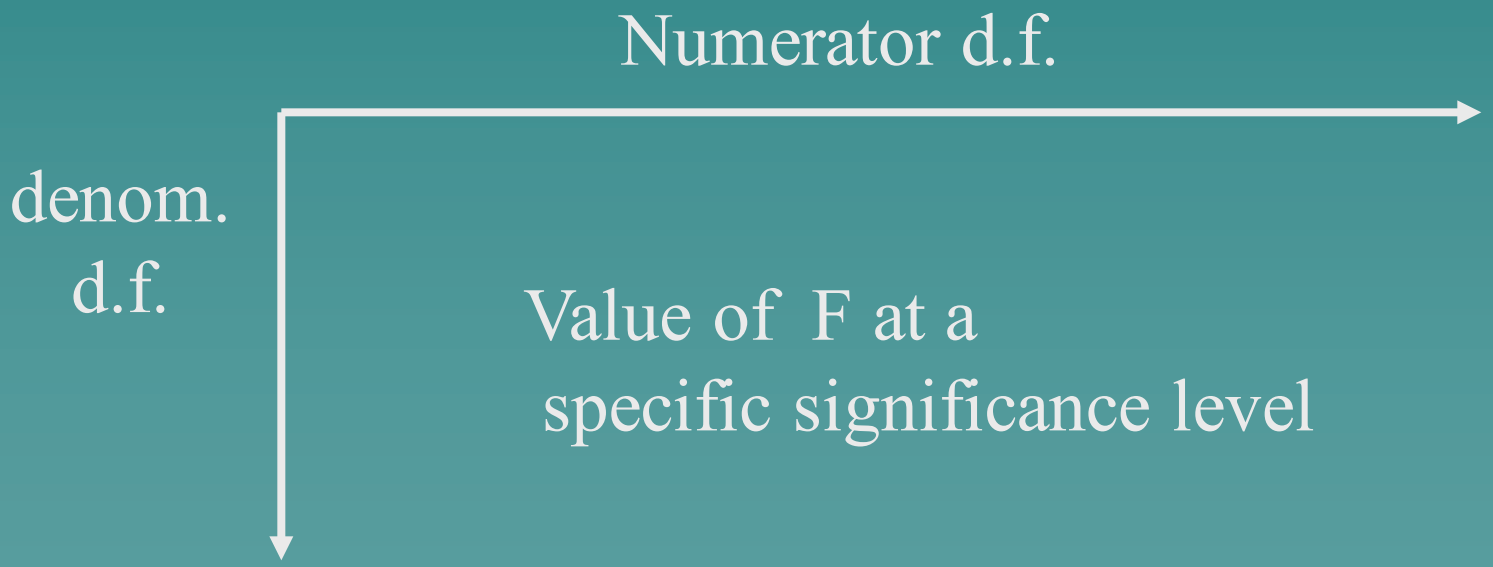
- Since R^2 is so high, you would certainly think that the model contains significant predictive power.
- In other problems it is perhaps not so obvious. For example, would an R^2 of 20% show any prediction ability at all?
- We can test for the predictive power of the entire model using the *F* statistic.

F Tests

- Generally these compare two sources of variation
- F = V1/V2 and has two df parameters
- Here V1 = SSR/*K* has *K* df
- And V2 = SSE/(*n-K-1)* has n-k-1 df

F Tables

Usually will see several pages of these; one or two pages at each specific level of significance (.10, .05, .01).



F Test Hypotheses

H_0: $\beta_1 = \beta_2 = \ldots = \beta_K = 0$ (None of the Xs help explain Y)

H_a: Not all βs are 0 (At least one X is useful)

H_0: $R^2 = 0$ is an equivalent hypothesis

F test for our example

S = 90.75 R-Sq = 85.5% R-Sq(adj) = 84.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1067797	533899	64.83	0.000
Residual Error	22	181176	8235		
Total	24	1248974			

F = 533899 / 8235 = 64.83 has p-value = 0.000

From tables, $F_{2,22,.05} = 3.44$ and $F_{2,22,.01} = 5.72$

Confirms that $R^2 = 85.5\%$ is not near zero