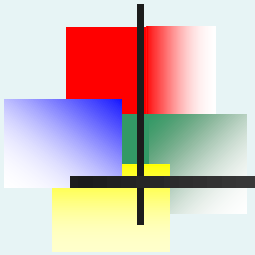


Trendline and Regression Analysis (Part 2):

Multiple Regression

BM 4419

Business Analytics

The Multiple Regression Model

Idea: Examine the linear relationship between
1 dependent (Y) & 2 or more independent variables (X_i)

Multiple Regression Model with k Independent Variables:

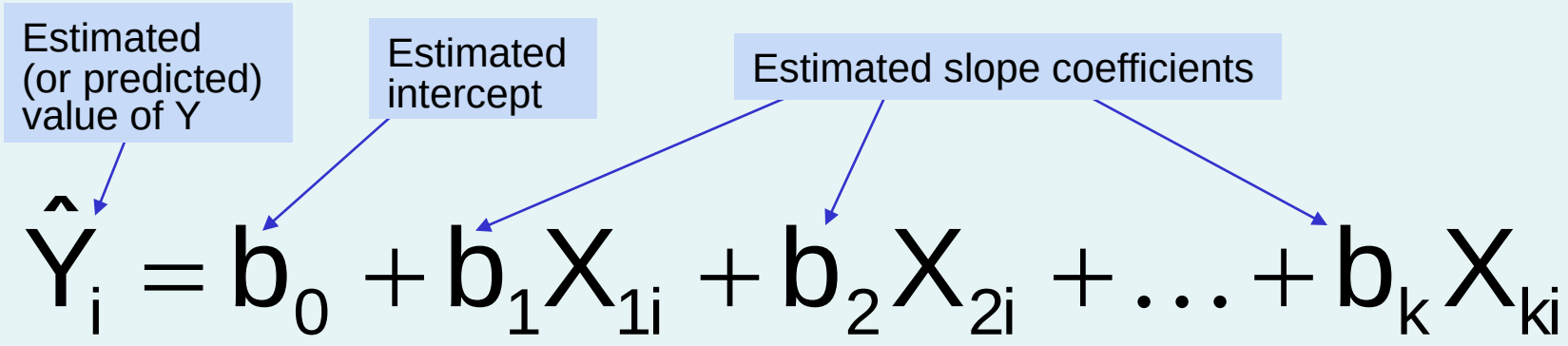
Y-intercept | Population slopes | Random Error

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon$$

Multiple Regression Equation

The coefficients of the multiple regression model are estimated using sample data

Multiple regression equation with k independent variables:



In this chapter we will always use Excel to obtain the regression slope coefficients and other regression summary measures.

Example: 2 Independent Variables

- A distributor of frozen desert pies wants to evaluate factors thought to influence demand
  - Dependent variable: Pie sales (units per week)
  - Independent variables: Price (in $), Advertising ($100's)
- Data are collected for 15 weeks

Pie Sales Example

Week	Pie Sales	Price ($)	Advertising ($100s)
1	350	5.50	3.3
2	460	7.50	3.3
3	350	8.00	3.0
4	430	8.00	4.5
5	350	6.80	3.0
6	380	7.50	4.0
7	430	4.50	3.0
8	470	6.40	3.7
9	450	7.00	3.5
10	490	5.00	4.0
11	340	7.20	3.5
12	300	7.90	3.2
13	440	5.90	4.0
14	450	5.00	3.5
15	300	7.00	2.7

Multiple regression equation:

$$\widehat{Sales} = b_0 + b_1 (Price) + b_2 (Advertising)$$

Estimating a Multiple Linear Regression Equation

- Excel will be used to generate the coefficients and measures of goodness of fit for multiple regression

- Excel:
  - Tools / Data Analysis... / Regression
- PHStat:
  - PHStat / Regression / Multiple Regression…

Multiple Regression Output

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

$$\widehat{Sales} = 306.526 - 24.975(Price) + 74.131(Advertising)$$

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

The Multiple Regression Equation

$$\widehat{Sales} = 306.526 - 24.975(Price) + 74.131(Advertising)$$

where

Sales is in number of pies per week

Price is in $

Advertising is in $100's.

b_1 = -24.975: sales will decrease, on average, by 24.975 pies per week for each $1 increase in selling price, holding all other variables constant.

b_2 = 74.131: sales will increase, on average, by 74.131 pies per week for each $100 increase in advertising, holding all other variables constant.

Using The Equation to Make Predictions

Predict sales for a week in which the selling price is $5.50 and advertising is $350:

$$\widehat{\text{Sales}} = 306.526 - 24.975(\text{Price}) + 74.131(\text{Advertising})$$

$$= 306.526 - 24.975\,(5.50) + 74.131\,(3.5)$$

$$= 428.62$$

Predicted sales is 428.62 pies

Note that Advertising is in $100's, so $350 means that $X_2 = 3.5$

Coefficient of Multiple Determination

- Reports the proportion of total variation in Y explained by all X variables taken together

$$r^2_{Y.12..k} = \frac{SSR}{SST} = \frac{\text{regression sum of squares}}{\text{total sum of squares}}$$

Multiple Coefficient of Determination

(continued)

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

$$r^2_{Y.12} = \frac{SSR}{SST} = \frac{29460.0}{56493.3} = .52148$$

52.1% of the variation in pie sales is explained by the variation in price and advertising

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

Adjusted r^2

- r^2 never decreases when a new X variable is added to the model
  - This can be a disadvantage when comparing models
- What is the net effect of adding a new variable?
  - We lose a degree of freedom when a new X variable is added
  - Did the new X variable add enough explanatory power to offset the loss of one degree of freedom?

Adjusted r^2

(continued)

- Shows the proportion of variation in Y explained by all X variables adjusted for the number of X variables used

$$r^2_{adj} = 1 - \left[(1 - r^2_{Y.12..k}) \left(\frac{n-1}{n-k-1} \right) \right]$$

(where n = sample size, k = number of independent variables)

- Penalize excessive use of unimportant independent variables
- Smaller than r^2
- Useful in comparing among models

Adjusted r^2

(continued)

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

$r^2_{adj} = .44172$

44.2% of the variation in pie sales is explained by the variation in price and advertising, taking into account the sample size and number of independent variables

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888