Topic 4

Multiple Linear Regression

Learning objectives

- Distinguish between simple linear regression and multiple linear regression
- Understand the applications of multiple linear regression.
- Use the F-test
- Calculate the Adjusted R-Squared

The Multiple Regression Model

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

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k-1,i} + \varepsilon_i$$

Multiple Regression Equation

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

Multiple regression equation with k independent variables:

Estimated (or predicted) value of y

Estimated intercept

Estimated slope coefficients

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{k,i}$$

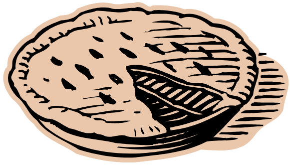
We will always use a computer to obtain the regression slope coefficients and other regression summary measures.

Example 1: Sales

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

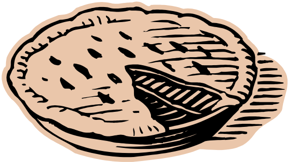
Multiple regression equation:

$$\widehat{\text{Sales}_t} = b_0 + b_1\,(\text{Price})_t + b_2\,(\text{Advertising})_t + e_t$$



Multiple Regression Output

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



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

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 formal F-test for slope parameter β_i

Null hypothesis H$_0$: $\beta_1 = \beta_2 = 0 \parallel R^2 = 0$
Alternative hypothesis H$_A$: $\beta_1 \neq \beta_2 \neq 0 \parallel R^2 \neq 0$

Test statistic

$$F^* = \frac{R^2/k}{(1-R^2)/n-k-1}$$

F-critical (from F-tables) $Fcritical = F_{k,n-k-1}$

Column Row

Formal F-test

- Decision rule

- When F*>F-critical, Reject H_0 and conclude that

R^2 is statistically significant

  - ✓ β_1 and β_2 are jointly significant

- When F*<F-critical, Fail to reject H_0 and conclude that R^2 is not statistically significant

  - ✓ β_1 and β_2 are jointly significant

Equivalence of F-test to t-test

- For a given α level, the F-test of $\beta_1 = 0$ versus $\beta_1 \neq 0$ is algebraically equivalent to the two-tailed t-test.
- Will get exactly same P-values, so…
  - If one test rejects H_0, then so will the other.
  - If one test does not reject H_0, then so will the other.

Should I use the F-test or the t-test?

- The F-test is only appropriate for testing that the slope differs from 0 ($\beta_1 \neq 0$).
- Use the t-test to test that the slope is positive ($\beta_1 > 0$) or negative ($\beta_1 < 0$).
- F-test is more useful for multiple regression model when we want to test that more than one slope parameter is 0. Test if β_1 and β_2 are jointly significant

Adjusted-R^2

- The adjusted R-squared compares the explanatory power of regression models that contain different numbers of predictors
- It is adjusted based on the df (i.e. the number of predictors in the model)
- Relevant in multiple regression
- Adjusted R^2 can actually get smaller as additional variables are added to the model.
- As N gets bigger, the difference between R^2 and Adjusted R^2 gets smaller and smaller.
- $$R^2_{adj} = 1 - (1 - R^2)\,\frac{n - 1}{n - k - 1}$$

Adjusted-R^2

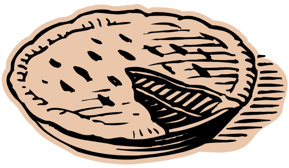
- One main difference between R^2 and the adjusted R^2
- R^2 assumes that every single variable explains the variation in the dependent variable.
- The adjusted R^2 tells you the percentage of variation explained by only the independent variables that actually affect the dependent variable.
- The adjusted R^2 is always lower than the R^2

Class exercise

- Given that $R^2 = 0.52$, n=15, and k=2
- Calculate the Adjusted R^2
- Test the statistical significance of R^2
- Use the t-test to test the statistical significance of β_1 and β_2
- Describe the equivalence of the F-test and the t-test above.

Solution: 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

Class test

- You are given the following regression equation:

$$Y = \underset{(0.006)}{98} - \underset{(0.002)}{9.3X_1} - \underset{(0.003)}{0.029X_2}$$

$$R^2 = 0.83$$

$$n = 12$$

1. Interpret the coefficient of determination (2 marks)
2. Test the significance of the coefficient of determination (6 marks)
3. Compute the adjusted R^2 and interpret your result (5 marks)
4. Why is the adjusted R^2 different from the R^2 value? (2 Marks)