CHAPTER 5

ELEMENTS OF MULTIPLE REGRESSION ANALYSIS: TWO INDEPENDENT VARIABLES

Today's Class

- Multiple Regression
  - Moving from one IV to multiple Ivs

- Similar stuff you saw with a single IV…

Moving from 1 Independent Variable to Multiple IV's

- The simple linear regression equation with one IV is as follows:

$$Y = a + bX + e$$

- Inferring this to multiple IV's is simple, just add more *b*'s!

$$Y = a + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + e$$

Multiple Regression

- Each independent variables has its own regression coefficient.

- Instead of the slope, we can think of each regression coefficient as follows:
  - The amount of change in the predicted value of Y as we increase 1 unit of X_i.

Calculation of Multiple Regression

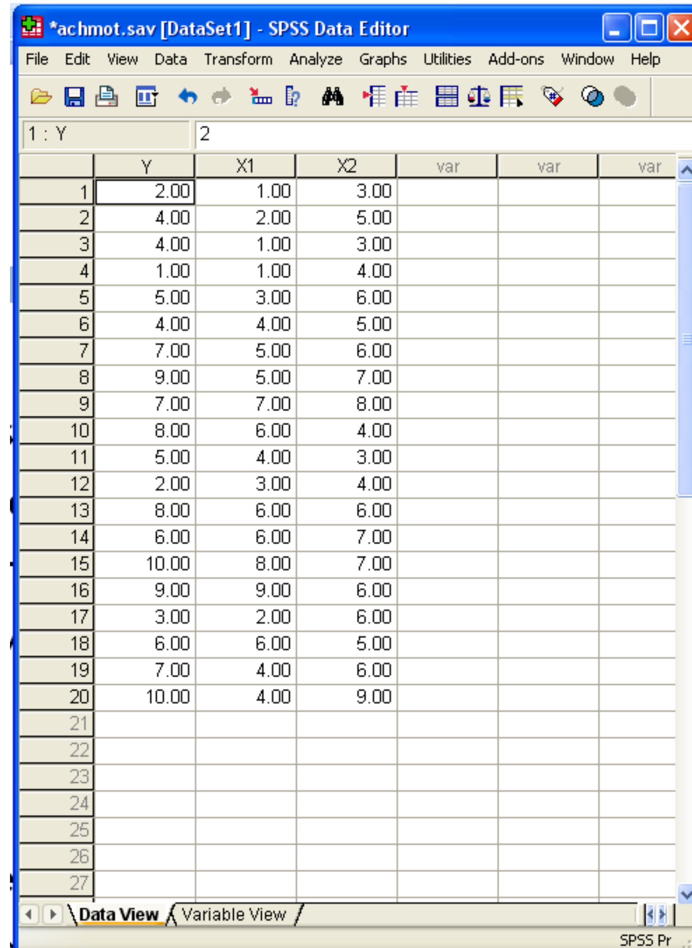
- If we have only two independent variables, calculation is tedious by not too difficult.

- Once we have more than 2 IV's we will have to rely on Matrix operations to perform calculations. We will learn more about that in the lectures to come. (YAY!)

Data Set

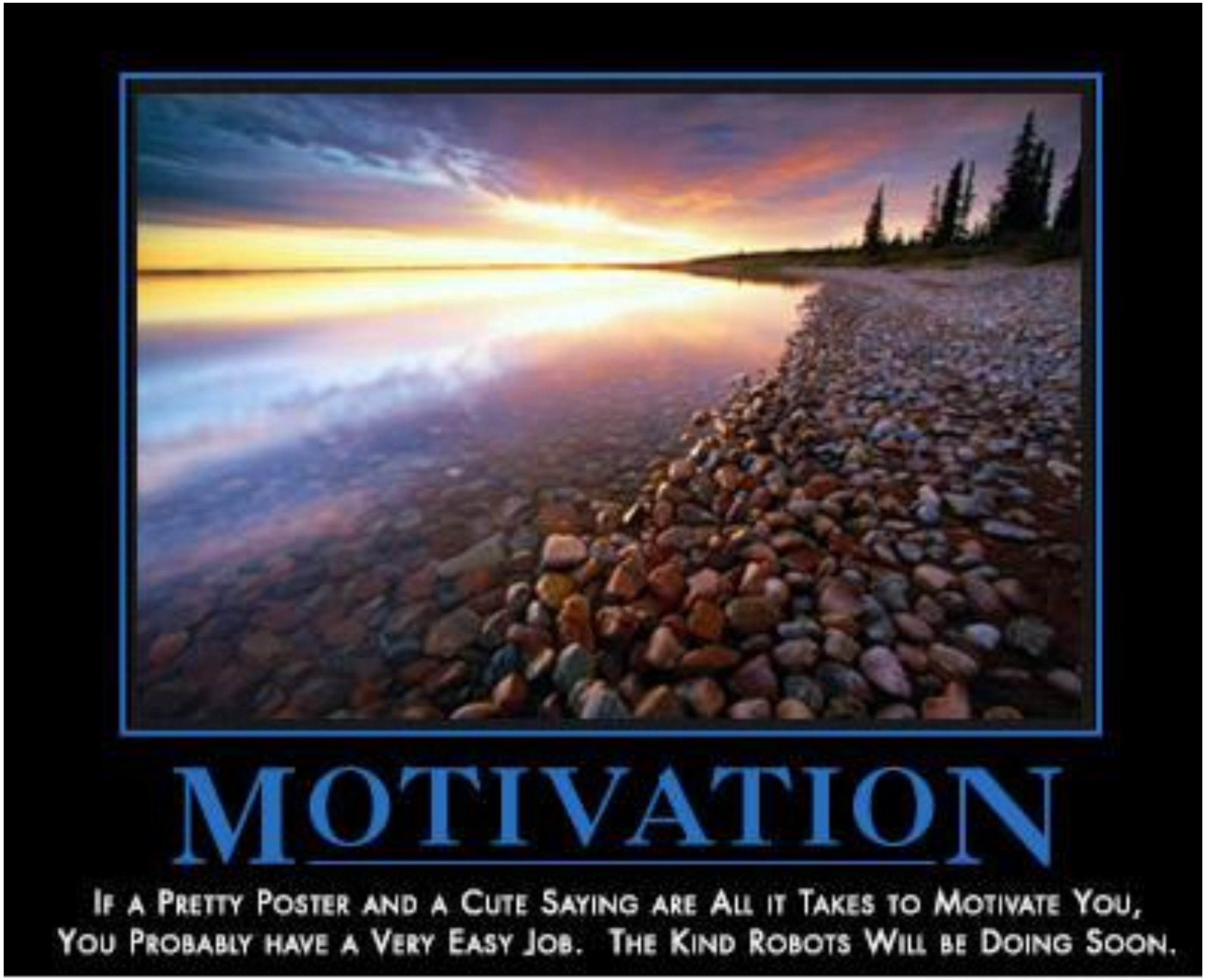
- 20 students were given a test that measured Reading Achievement (Y), Verbal Aptitude (X_1), and Achievement Motivation (X_2).

- The data is given on Pg. 98 in the text.



	Y	X1	X2	var	var	var
1	2.00	1.00	3.00			
2	4.00	2.00	5.00			
3	4.00	1.00	3.00			
4	1.00	1.00	4.00			
5	5.00	3.00	6.00			
6	4.00	4.00	5.00			
7	7.00	5.00	6.00			
8	9.00	5.00	7.00			
9	7.00	7.00	8.00			
10	8.00	6.00	4.00			
11	5.00	4.00	3.00			
12	2.00	3.00	4.00			
13	8.00	6.00	6.00			
14	6.00	6.00	7.00			
15	10.00	8.00	7.00			
16	9.00	9.00	6.00			
17	3.00	2.00	6.00			
18	6.00	6.00	5.00			
19	7.00	4.00	6.00			
20	10.00	4.00	9.00			
21						
22						
23						
24						
25						
26						
27						

Knowing the Data



Preliminary Calculations

$$\sum Y = 117$$

$$\sum X_1 = 87$$

$$\sum X_2 = 110$$

$$\sum Y^2 = 825$$

$$\sum X_1^2 = 481$$

$$\sum X_2^2 = 658$$

$$\sum X_1Y = 604$$

$$\sum X_2Y = 702$$

$$\sum X_1X_2 = 517$$

$$N = 20$$

$$\sum y^2 = \sum Y^2 - \frac{{\sum Y}^2}{N} = 825 - \frac{{117}^2}{20} = 825 - 684.45 = 140.55$$

$$\sum x_1^2 = \sum X_1^2 - \frac{{\sum X_1}^2}{N} = 481 - \frac{{87}^2}{20} = 481 - 378.45 = 102.55$$

$$\sum x_2^2 = \sum X_2^2 - \frac{{\sum X_2}^2}{N} = 658 - \frac{{110}^2}{20} = 658 - 605 = 53$$

$$\sum x_1 y = \sum X_1 Y - \frac{\sum X_1 \quad \sum Y}{N} = 604 - \frac{87 \quad 117}{20} = 604 - 508.95 = 95.05$$

$$\sum x_1 y = \sum X_2 Y - \frac{\sum X_2 \quad \sum Y}{N} = 702 - \frac{110 \quad 117}{20} = 702 - 643.5 = 58.5$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{\left(\sum X_1\right)\left(\sum X_2\right)}{N} = 517 - \frac{(87)(110)}{20} = 517 - 478.5 = 38.5$$

Calculation of b_1

$$b_1 = \frac{\left(\sum x_2^2\right)\left(\sum x_1 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_2 y\right)}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$b_1 = \frac{(53)(95.05) - (38.5)(58.5)}{(102.55)(53) - (38.5)^2} = \frac{5037.65 - 2252.25}{5435.15 - 1482.25}$$

$$b_1 = \frac{2785.4}{3952.9} = .7046$$

Calculation of b_2

$$b_2 = \frac{\left(\sum x_1^2\right)\left(\sum x_2 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_1 y\right)}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$b_2 = \frac{(102.55)(58.5) - (38.5)(95.05)}{(102.55)(53) - (38.5)^2} = \frac{5999.175 - 3659.425}{5435.15 - 1482.25}$$

$$b_2 = \frac{2339.75}{3952.9} = .5919$$

Calculation of a

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

$$a = 5.85 - \;.7046 \quad 4.35 \; - \; .5919 \quad 5.5$$

$$a = -.4705$$

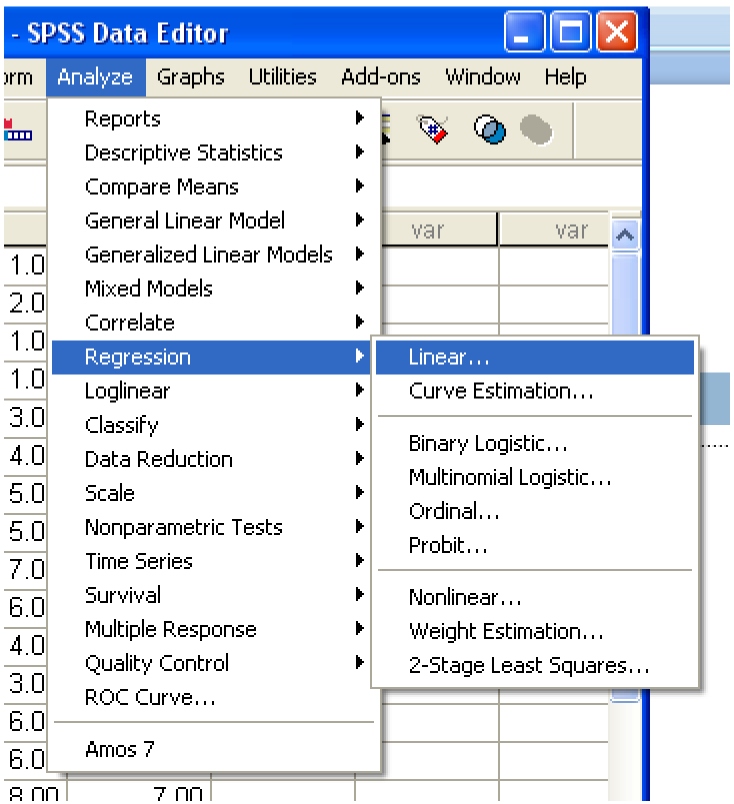
The final regression equation

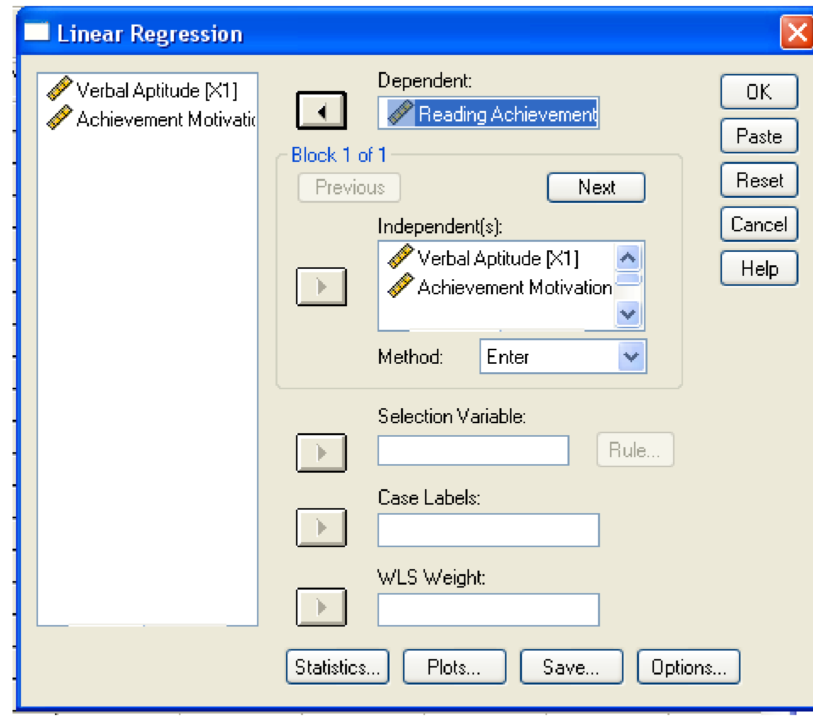
- After we finish all the calculations, we can put it all together

$$Y' = a + b_1 X_1 + b_2 X_2$$

$$Y' = -.4705 + .7046 \; X_1 + .5919 \; X_2$$

Multiple Regression in SPSS





SPSS Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.850[a]	.723	.690	1.51360

a. Predictors: (Constant), Achievement Motivation, Verbal Aptitude

Proportion of Variance in Y accounted for by X1 and X2

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	101.603	2	50.802	22.175	.000[a]
	Residual	38.947	17	2.291		
	Total	140.550	19			

a. Predictors: (Constant), Achievement Motivation, Verbal Aptitude

b. Dependent Variable: Reading Achievement

Test of hypothesis that either X1 or X2 (or both) have some relationship with Y (a non-zero regression coefficient)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.471	1.194		-.394	.698
	Verbal Aptitude	.705	.175	.602	4.021	.001
	Achievement Motivation	.592	.244	.363	2.428	.027

a. Dependent Variable: Reading Achievement

Test of hypothesis the coefficient for X1 is zero.

Test of hypothesis the coefficient for X2 is zero.

Predicted Values

- Once we have the final equation, we can use it for prediction.

- Let's try to predict 2 subjects:
  - Subject 1: $X_1=1$ and $X_2=3$
  - Subject 20: $X_1=4$ and $X_2=9$

$$Y'(\text{Subject 1}) = -.4705 + .7046\ \ 1\ \ + .5919\ \ 3\ \ = 2.0098$$

$$Y'(\text{Subject 20}) = -.4705 + .7046\ \ 4\ \ + .5919\ \ 9\ \ = 7.6750$$

Residuals

- Once we have the predicted values, we can calculate residuals for these two subjects.

$$\text{Residual Subject 1} \;\; e_1 = Y - Y' = 2 - 2.0098 = -.0098$$

$$\text{Residual Subject 20} \;\; e_{20} = Y - Y' = 10 - 7.765 = 2.325$$

Sum of Squares

$$ss_{reg} = b_1 \sum x_1 y + b_2 \sum x_2 y$$

$$ss_{reg} = .7046 \quad 95.05 \; + \; .5919 \quad 58.5 \; = 101.6$$

$$ss_{res} = \sum y^2 - ss_{reg}$$

$$ss_{res} = 140.55 - 101.6 = 38.95$$

Squared Multiple Regression Coefficient (R^2)

- Remember from CH. 2, R^2 was the amount of variance accounted for by the independent variable

$$R^2 = \frac{SS_{reg}}{\sum y^2}$$

From our previous example this would be:

$$R^2 = \frac{101.60}{140.55} = .723$$

- 72% of the total variance of the dependent variable (Y) was accounted for by the two independent variables (X_1&X_2)

Alternative methods of Calculations

- We can calculate this by hand in terms of the correlation coefficients…(example shown in book).

- We can also perform the calculations with the click of the mouse in SPSS.

Tests of Significance

- Once we have calculated the parameter values it is important to determine if they are significant.
- The following are tests used for all the parameters

Test of R^2

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

df = k,N-k-1
Df = 2,17

$$F = \frac{.723 / 2}{(1 - .723) / (20 - 2 - 1)} = \frac{.3615}{.0163} = 22.18$$

Or alternatively,

$$\text{F} = \frac{\text{ss}_{\text{reg}} / \text{df}_{\text{reg}}}{\text{ss}_{\text{res}} / \text{df}_{\text{res}}} = \frac{101.6 / 2}{38.95 / 17} = \frac{50.8}{2.29} = 22.18$$

Test of b's

- Want to test if b is significantly different from 0.
- This is done in the same way as it was done with one variable....take the b value and divide it by its standard error.

In our example with 2 IV's, the standard error for b_1 is:

$$s_{b_{y1.2}} = \sqrt{\frac{s^2_{y.12}}{\sum x_1^2 \left(1 - r_{12}^2\right)}}$$

the standard error for b_2 is:

$$s_{b_{y2.1}} = \sqrt{\frac{s^2_{y.12}}{\sum x_2^2 \left(1 - r_{12}^2\right)}}$$

Test of b's

- Using our same example from before

$$s^2_{y.12} = \frac{ss_{res}}{N-k-1}$$

$$ss_{res} = 38.95$$

$$s^2_{y.12} = \frac{38.95}{20-2-1} = 2.29$$

$$\sum x_1^2 = 102.55$$

$$\sum x_2^2 = 53$$

$$b_1 = .7046$$

$$b_2 = .5919$$

$$r_{12} = .522$$

$$s_{b_1} = \sqrt{\frac{2.29}{102.55\ 1-.522^2}} = .1752$$

$$t_{b_1} = \frac{b_1}{s_{b_1}} = \frac{.7046}{.1752} = 4.02$$

$$s_{b_2} = \sqrt{\frac{2.29}{53\ 1-.522^2}} = .2437$$

$$t_{b_2} = \frac{b_2}{s_{b_2}} = \frac{.5919}{.2437} = 2.43$$

Test of R^2 vs. Test of b

- The test of R^2 is the same as testing all the b's simultaneously.

- When we test each b individually, we are testing the given b while controlling for all other independent variables.

Confidence Intervals

- We can calculate confidence intervals in multiple regression similar to the way we did in simple regression:

$$b \pm t_{\alpha/2,df}\, s_b$$

Using our same example:

CI for b_1

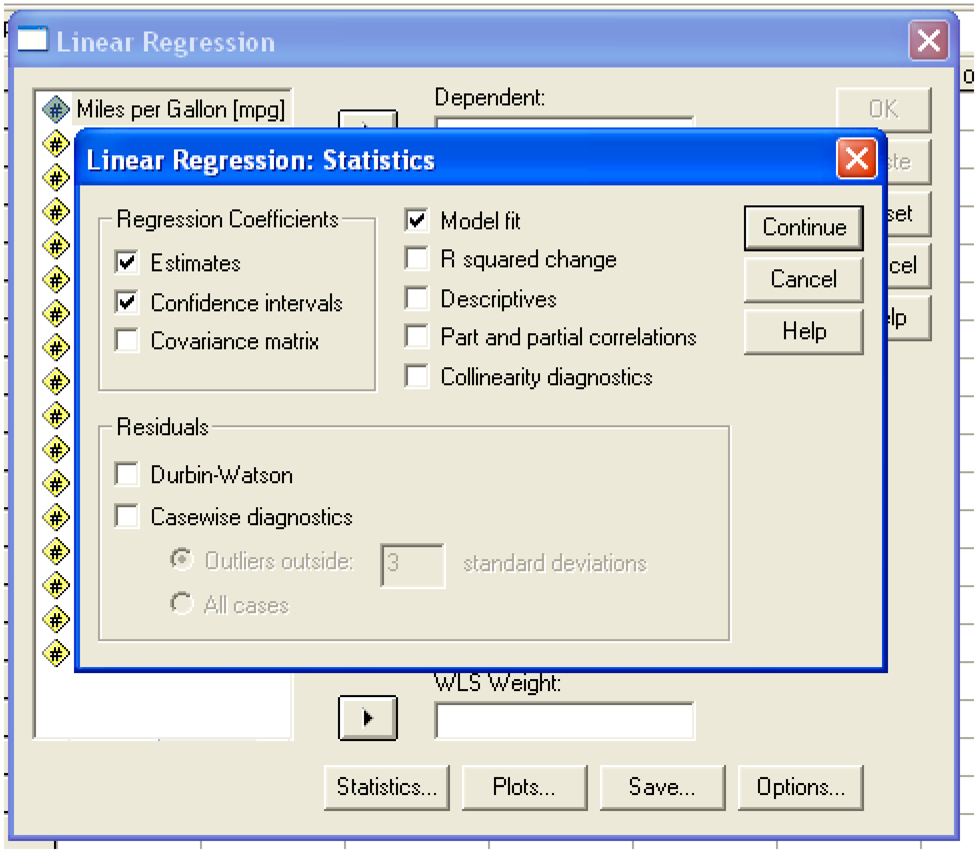
$.7046 \pm 2.11 \quad .1752 = .3349, 1.0743$

CI for b_2

$.5919 \pm 2.11 \quad .2437 = .0777, 1.1061$

- Our CI's do not include 0, again confirming that the regression coefficients significantly differ from 0.

Confidence Intervals in SPSS



SPSS Output

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-.471	1.194		-.394	.698	-2.990	2.049
	Verbal Aptitude	.705	.175	.602	4.021	.001	.335	1.074
	Achievement Motivation	.592	.244	.363	2.428	.027	.078	1.106

a. Dependent Variable: Reading Achievement

Test for increment in proportion of variance accounted for

- This tests the amount of variance accounted for as increased due to the adding of another independent variable.

- In our example, we can test the increment due to adding X_2 on top of what information we already have from X_1….along with testing the increment due to adding X_1 on top of X_2.

- This test is actually equivalent to testing the individual b coefficient.

Test for increment in proportion of variance accounted for

The test for X_2

$$F = \frac{R^2_{y.12} - R^2_{y.1} \Big/ (2-1)}{1 - R^2_{y.12} \Big/ N - 2 - 1}$$

$$F = \frac{.723 - .6273 \Big/ (2-1)}{1 - .723 \Big/ 20 - 2 - 1} = \frac{.0957}{.0163} = 5.87$$

df = 1,N-2-1
df =1,17

The test for X_1

$$F = \frac{R^2_{y.12} - R^2_{y.2} \Big/ (2-1)}{1 - R^2_{y.12} \Big/ N - 2 - 1}$$

$$F = \frac{.723 - .4597 \Big/ (2-1)}{1 - .723 \Big/ 20 - 2 - 1} = \frac{.2633}{.0163} = 16.15$$

Relative Importance of Variables

- The magnitude of b is in part affected by the scale of measurement.
- For example, if you measure objects in inches instead of feet, the nature of the regression and the tests of significance will not change.
- What will change is the magnitude of the b.
- Therefore remember, it isn't the size of the b that is important, it is its significance.

Relative Importance of Variables

- Thinking back to our example:
  - $b_1 = .602$ and $b_2 = .364$
- This does not mean that b_1 is twice as important as $b_{2.}$
- They simply represent different variables measured on different scales.

Next Time

- Lab session: Regression in practice.
  - Meet in room 228.