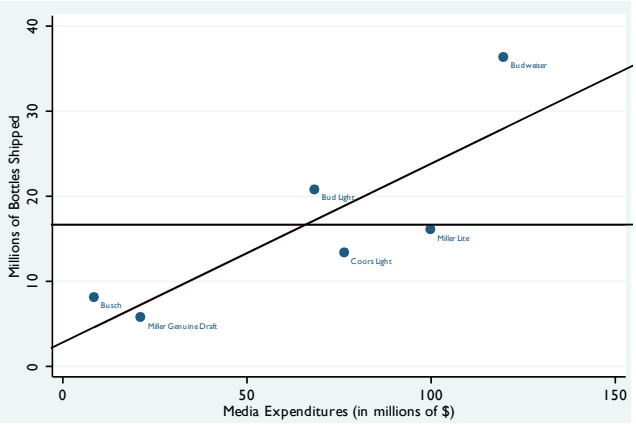
Activity 14: Simple Linear Regression - models, assumptions, tests for significance

1) Let's return to our example of predicting Y = (the number of bottles shipped) for six major beer brands. In this example, we found the best-fitting (ordinary least squares) regression line to be: $\hat{Y} = 2.76 + 0.21x$

Brand	X = media expenditures (millions of $)	Y = bottles shipped (millions of bottles)
Busch	8.7	8.1
MGD	21.5	5.6
Bud Light	68.7	20.7
Coors Light	76.6	13.2
Miller Lite	100.1	15.9
Budweiser	120.0	36.3
Mean:	**65.933**	**16.633**
Std. Dev:	**43.5017**	**11.0471**
	Correlation = 0.829	

Source: *Superbrands 1998*; 10/20/97



2) When we conduct a linear regression analysis, we're often interested in finding the most parsimonious model that can predict or explain the variance in the dependent variable. To find this model, we may try several models, each increasing in complexity. For example:

a) We may start with the most simple model in which we predict Y by a single constant: $\hat{Y} = b_0$.
Since we know our observations won't have identical Y-values, we can add an error term: $Y_i = \beta_0 + \varepsilon_i$.

b) We could then add a predictor to the model so that Y is predicted by: $Y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$
We would then compare this model to the previous model to see if it provided a better prediction.

c) We could then add yet another predictor: $Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$ and compare it to the previous model. If this new model provided a significantly better prediction (explained a significant amount of previously unexplained variance), then we would keep this new model. If the model didn't really improve our prediction, we could decide to keep the old, simpler model.

At each stage in building our regression model, we can assess the value of adding predictors (complexity) through formal hypothesis testing methods. These methods allow us to determine which independent variables to keep in our model.

We could also work through this process backwards. We could start with the most complex model -- one that predicts Y with many independent variables -- and eliminate the predictors that explain the least amount of variation in Y. We'll use this method later; for now, let's try building a model by adding a predictor.

3) When comparing regression models, I recommend writing out the **full model** (the more complex) and the **reduced model** (the less complex). Outside of this class, you would choose which models to compare. Because we're just beginning to learn this, I will choose our models in this activity. In this example, we want to see if X (media expenditures) provides a better prediction than a model with no predictors. Write the full & reduced models:

Full model: _____

Reduced model: _____

4) We know that the best regression line minimizes the sum of squared errors (SSE). Let's see how much error we would have with both models:

Observed Data			Reduced Model			Full Model		
Brand	**X = media**	**Y = bottles**	**Predicted**	**Error**	**Error^2**	**Predicted**	**Error**	**Error^2**
Busch	8.7	8.1	16.633	-8.533	72.812	4.587	3.513	12.341
MGD	21.5	5.6	16.633	-11.033	121.727	7.275	-1.675	2.806
Bud Light	68.7	20.7	16.633	4.067	16.540	17.187	3.519	12.383
Coors Light	76.6	13.2	16.633	-3.433	11.785	18.846	-5.646	31.877
Miller Lite	100.1	15.9	16.633	-0.733	0.537	23.781	-7.881	62.110
Budweiser	120.0	36.3	16.633	19.667	386.791	27.96	8.340	69.556
Mean:	**65.933**	**16.633**		**SUM:**	**610.192**		**SUM:**	**191.073**

Source: *Superbrands 1998*; 10/20/97

What do those sums in the bottom row represent?

5) The table shows that when we added X to our model, the sum of squares reduced by 610.192 - 191.073 = 417.277. What does this number represent?

6) Write these sums of squares into the following ANOVA summary table. Explain what SSY, SSR, and SSE represent. How many degrees of freedom will we have?

Source	SS	df	MS	MSR
Regression (b_1 | b_0)				
Error				
Total				

1) Compare your MSR to the appropriate F value in the table. What conclusion can you make?

2) Another way to make a similar conclusion would be to run a test to determine if β_1 is significant (significantly different from zero). What would we conclude if this coefficient was not significant?

Let's conduct a simple t-test:

Note:

$$t_{n-2} = \frac{\hat{\beta}_1 - 0}{\frac{S_{y|x}}{S_x\sqrt{n-1}}} = \frac{\frac{S_y}{S_x} r_{xy}}{\frac{S_y\sqrt{1-r^2}\sqrt{\frac{n-1}{n-2}}}{S_x\sqrt{n-1}}} = \frac{\frac{S_y}{S_x} r_{xy}}{\frac{S_y\sqrt{1-r^2}\sqrt{\frac{n-1}{n-2}}}{S_x\sqrt{n-1}}} = \frac{\frac{S_y}{S_x} r_{xy}\left(S_x\sqrt{n-1}\right)}{S_y\sqrt{1-r^2}\sqrt{\frac{n-1}{n-2}}} =$$

$$= \frac{\frac{S_y}{S_x} r_{xy}\left(S_x\sqrt{n-1}\right)}{S_y\sqrt{1-r^2}\sqrt{\frac{n-1}{n-2}}} = \frac{r_{xy}\sqrt{n-1}}{\sqrt{1-r^2}\sqrt{\frac{n-1}{n-2}}} =$$

$$= t^2_{n-2} = \frac{r^2(n-1)}{\frac{\left(1-r^2\right)(n-1)}{(n-2)}} = \frac{r^2(n-2)}{\left(1-r^2\right)} =$$

$$= t_{n-2} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r^2}} =$$

This is the test statistic for testing the significance of a correlation coefficient.

1) What did we just learn? If you wish to test the significance of a correlation coefficient, you conduct a simple t-test with n-2 degrees of freedom. This t-test also tests the significance of the regression coefficient. Let's test the significance of our correlation of 0.829:

2) The value of that test statistic should look familiar. It's the same value we got when we tested the beta coefficient. This is important. Since the correlation and regression coefficient are so closely related (look at their formulas), we can use the same hypothesis test for both. But that still doesn't explain why we used an F-test (in our ANOVA summary table) to determine if our full model was better than our reduced model.

   Look again at the F-test we conducted. Let's rearrange some terms:

$$F = \frac{SS_{reg} / df_{reg}}{SSE / df_E} = \frac{r^2(SSY)/df}{(1-r^2)(SSY)/df} = \frac{r^2/(n-2)}{(1-r^2)/(1)} = t^2_{n-2}$$

   From this, we can see that an F-test is equivalent to the square of a t-test. We calculated a t-statistic of 2.96. Therefore, we should have calculated an F-statistic of 2.96^2 = 8.76. You can verify this by looking at the table in question #6.

3) We will learn another way to calculate the F-statistic (using what we will call the omnibus F-test). Let's try it now:

$$F^{\,k_{full}-k_{reduced}}_{n-k_{full}-1} = \frac{\left(R^2_{full} - R^2_{reduced}\right)/\left(k_{full} - k_{reduced}\right)}{\left(1-R^2\right)/\left(N - k_{full} - 1\right)}$$

$\overline{X} = \frac{1}{n}\sum x_i = 65.933$	$\overline{Y} = \frac{1}{n}\sum y_i = 16.633$
$S_x^2 = \frac{\sum (X - \overline{X})}{n-1} = 1892.397$	$S_y^2 = \frac{\sum (Y - \overline{Y})}{n-1} = 122.038$
$S_x = \sqrt{\frac{\sum (X - \overline{X})}{n-1}} = 43.5017$	$S_y = \sqrt{\frac{\sum (Y - \overline{Y})}{n-1}} = 11.0471$
$r_{xy} = \frac{1}{n-1}\sum \frac{(x_i - \overline{X})(y_i - \overline{Y})}{s_x s_y} = 0.829$	$R_{xy}^2 = \frac{SS_{reg}}{SSY} = \frac{SSY - SSE}{SSY} = 0.687$
$S_{Y\|X}^2 = \frac{\sum (Y - \hat{Y})^2}{n-2} = \sigma_Y^2 (1 - R^2)\left(\frac{n-1}{n-2}\right) = 47.756$	$S_{Y\|X} = \sqrt{S_{Y\|X}^2} = S_Y\sqrt{1 - R^2}\sqrt{\frac{n-1}{n-2}} = 6.9106$
$1 - R_{xy}^2 = \frac{SSE}{SSY} = 0.313$	N = 6
$\hat{\beta}_1 = \frac{S_y}{S_x} r_{xy} = 0.21$	$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 2.67$
$SSY = \sum (Y - \overline{Y})^2 = (n-1)S_Y^2 = 610$	$SSE = \sum (Y - \hat{Y})^2 = (n-2)S_{Y\|X}^2 = (1 - r^2)SSY = 191$
$SS_{reg} = \frac{SSY - SSE}{SSY} = \sum (\hat{Y} - \overline{Y})^2 = (R^2)SSY = 417$	You can see most of these statistics are related (just rearrange terms)