EWMBA 296 (Fall 2015)

Section 6

GSI: Fenella Carpena

December 3, 2015

Agenda for Today

- Practice Problem # 1: Cheese, Please
- Why do we care about F-test?
- Practice Problem # 2: Advertising: TV or Print?

Practice Problem # 1: Cheese, Please

As cheddar cheese matures, a variety of chemical processes take place. The taste of matured cheese is related to the concentration of several chemicals in the final product. In a study of cheddar cheese from the LaTrobe Valley of Victoria, Australia, a sample of 30 cheeses were analyzed for their chemical composition and were subjected to taste tests. The variables measured during the study include:

- *Taste*: Subjective taste test score, ranging from 0-100, obtained by combining the scores of several tasters. The higher the score, the better.
- *LogH2S*: Log of concentration of hydrogen sulfide in parts per thousands
- *LogAcetic*: Log of concentration of acetic acid in parts per thousand
- *Lactic*: Concentration of lactic acid in parts per thousand

Practice Problem # 1: Cheese, Please

(1a) You begin with a regression of *Taste* on *H2S*, and you obtain the following regression output. Interpret the coefficient of *H2S* in words.

Regression Statistics	
Multiple R	0.75575227
R Square	0.5711615
Adjusted R Square	0.55584584
Standard Error	10.8333823
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4376.74585	4376.74585	37.2926453	1.37378E-06
Residual	28	3286.140816	117.362172		
Total	29	7662.886667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-9.7868374	5.957910275	-1.6426628	0.11163772	-21.9910634	2.41738853
LogH2S	5.77608861	0.945849964	6.10677045	1.3738E-06	3.838602787	7.71357443

Practice Problem # 1: Cheese, Please

(1b) Now you regress *Taste* on all 3 chemical compounds (*LogAcetic*, *LogH2S*, *Lactic*). The table below presents the regression results.

(i) What is the F-statistic of this model?

(ii) What is the p-value associated with this F-statistic?

(iii) Calculate the F-statistic by hand, and verify that you obtain the same result as in the regression table.

Regression Statistics	
Multiple R	0.80732564
R Square	0.65177469
Adjusted R Square	0.61159485
Standard Error	10.1307056
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	4994.47558	1664.82519	16.2214343	3.81018E-06
Residual	26	2668.411086	102.631196		
Total	29	7662.886667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-28.87677	19.73541835	-1.4631952	0.15539915	-69.443503	11.6899638
LogAcetic	0.32774129	4.45975656	0.0734886	0.94197977	-8.83941961	9.49490219
LogH2S	3.91184107	1.248430294	3.13340768	0.00424708	1.345655852	6.4780263
Lactic	19.6705434	8.629054829	2.27957102	0.03107948	1.933267124	37.4078196

Practice Problem # 1: Cheese, Please

(1c) Consider the same regression as in part (b).

(i) Suppose that you wish to test whether *all* of the explanatory variables collectively have no effect on the taste of the cheese. State the null and alternative hypothesis.

(ii) Carry out the hypothesis test from part (i) at the 5% level. Do you reject or fail to reject H_0?

(iii) Interpret the result of your hypothesis test from part (ii).

Practice Problem # 1: Cheese, Please

(1d) Point to two specific pieces of evidence in the regression from part (b) that suggest the presence of collinearity.

Regression Statistics	
Multiple R	0.80732564
R Square	0.65177469
Adjusted R Square	0.61159485
Standard Error	10.1307056
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	4994.47558	1664.82519	16.2214343	3.81018E-06
Residual	26	2668.411086	102.631196		
Total	29	7662.886667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-28.87677	19.73541835	-1.4631952	0.15539915	-69.443503	11.6899638
LogAcetic	0.32774129	4.45975656	0.0734886	0.94197977	-8.83941961	9.49490219
LogH2S	3.91184107	1.248430294	3.13340768	0.00424708	1.345655852	6.4780263
Lactic	19.6705434	8.629054829	2.27957102	0.03107948	1.933267124	37.4078196

Practice Problem # 1: Cheese, Please

Regression of *Taste* on *LogH2S*

Regression Statistics	
Multiple R	0.75575227
R Square	0.5711615
Adjusted R Square	0.55584584
Standard Error	10.8333823
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4376.74585	4376.74585	37.2926453	1.37378E-06
Residual	28	3286.140816	117.362172		
Total	29	7662.886667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-9.7868374	5.957910275	-1.6426628	0.11163772	-21.9910634	2.41738853
LogH2S	5.77608861	0.945849964	6.10677045	1.3738E-06	3.838602787	7.71357443

Practice Problem # 1: Cheese, Please

(1e) Suppose that you run an additional regression where you keep all 3 explanatory variables (*LogAcetic*, *LogH2S*, *Lactic*), and include an additional variable called *Time*, which is the number of days the cheese has been aged. In comparing this new model that uses 4 explanatory variable to the previous model which uses 3 explanatory variables, what change would you expect to see in the value of the R-squared?

Select one from the following options.

(1) The new model's R-squared value is lower.

(2) The new model's R-squared value is higher.

(3) Cannot tell from the information given.

Practice Problem # 1: Cheese, Please

(1f) Continued from part (e): What change would you expect to see in the value of the F-statistic?

Select one from the following options.

(1) The new model's F-statistic is lower.

(2) The new model's F-statistic is higher.

(3) Cannot tell from the information given.

Why do we care about F-test?

I can think of the following three reasons.

Reason # 1: It helps us to understand the overall explanatory power of our X variables. Consider the regression from Lecture # 12.

Reason # 2: It helps us to diagnose collinearity problems, as seen in Question 1d of the previous exercise. We can compare the F-stat with the individual t-stats to determine whether collinearity is present.

Reason # 3: In some cases the F-stat for the hypothesis that all explanatory variables are zero is the focus of the study. For example:

- Suppose we wanted to know how economic policies affects economic growth.
- For our explanatory variables, we may include several policy instruments (balanced budgets, inflation, trade openness, etc.).
- We could conduct a hypothesis test to determine if all of these policies variables are jointly significantly different from zero.
- This hypothesis is potentially the one of interest: theory rarely tells us which particular policy variable is important, but rather a broad category of variables.

Aside: Why do we care about F-test?

Regression from Lecture # 12

Source	SS	df	MS		
				Number of obs =	81
				F(20, 60) =	1.11
Model	284849698	20	14242484.9	Prob > F =	0.3621
Residual	768332997	60	12805549.9	R-squared =	0.2705
				Adj R-squared =	0.0273
Total	1.0532e+09	80	13164783.7	Root MSE =	3578.5

dow	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
random1	35.80516	481.0383	0.07	0.941	-926.4148	998.0251
random2	247.2956	392.14	0.63	0.531	-537.1012	1031.692
random3	912.3516	481.8561	1.89	0.063	-51.50411	1876.207
random4	80.44477	534.8426	0.15	0.881	-989.3997	1150.289
random5	266.1893	470.0973	0.57	0.573	-674.1454	1206.524
random6	-67.22133	490.011	-0.14	0.891	-1047.389	912.9465
random7	512.3638	417.5273	1.23	0.225	-322.8152	1347.543
random8	-305.2283	539.4625	-0.57	0.574	-1384.314	773.8573
random9	75.60148	492.9816	0.15	0.879	-910.5085	1061.711
random10	-764.1952	483.6255	-1.58	0.119	-1731.59	203.1998
random11	106.9755	434.1365	0.25	0.806	-761.4267	975.3778
random12	591.6559	498.2542	1.19	0.240	-405.0009	1588.313
random13	318.5282	393.0607	0.81	0.421	-467.7102	1104.767
random14	372.8452	522.5232	0.71	0.478	-672.3568	1418.047
random15	23.61284	426.5991	0.06	0.956	-829.7123	876.938
random16	-735.4908	449.7786	-1.64	0.107	-1635.182	164.2002
random17	1237.224	446.4344	2.77	0.007	344.2219	2130.225
random18	-731.3945	467.363	-1.56	0.123	-1666.26	203.4707
random19	699.8461	463.0684	1.51	0.136	-226.4285	1626.121
random20	-118.2315	454.6812	-0.26	0.796	-1027.729	791.2663
_cons	2685.739	484.8051	5.54	0.000	1715.984	3655.493

The t-statistic on the coefficient of *random*17 is really high, but the overall explanatory power of the regression is low. This provides evidence that the high t-stat for the coefficient on *random*17 is likely due to chance.

Practice Problem # 2: Advertising: TV or Print?

You are a manager charged with allocating advertising dollars between TV ads and print ads. You have a dataset describing spending and sales at the company during the previous 200 months. You start by regressing company sales on TV ad spending:

R Square	0.07
Adjusted R Square	0.07
F-Statistic	15.61
p-value for F-statistic	0.0001

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2556.1	233.9	10.93	0.00
TV Ads (millions $)	9.01	2.28	3.95	0.00

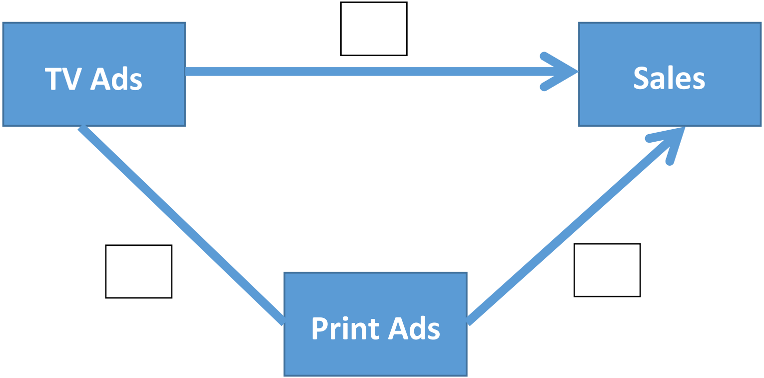
Next, you add print advertising to the regression:

R Square	0.07
Adjusted R Square	0.07
F-Statistic	7.94
p-value for F-Statistic	0.0005

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2486.5	264.4	9.41	0.00
TV Ads (millions $)	5.75	6.18	0.93	0.35
Print Ads (millions $)	3.96	6.97	0.57	0.56

(2a) Construct a path diagram to describe the relationship among sales, TV ad spending, and print ad spending. Is the correlation between TV ads and print ads positive, negative or zero?

Practice Problem # 2: Advertising: TV or Print?



Practice Problem # 2: Advertising: TV or Print?

(2b) Mark each of the following statements TRUE or FALSE.

(i) In the MRM of sales on TV ad spending and print ad spending, the R^2 is 0.07. In the SRM of sales on TV advertising alone, the R^2 is also 0.07. Therefore, the R^2 of a regression of sales on print advertising alone would be 0.00.

(ii) The F-statistic in the MRM of sales on TV ad spending and print ad spending is used to test the null hypothesis that the intercept, the slope of TV ad spending, and the slope of print ad spending are all zero.

Practice Problem # 2: Advertising: TV or Print?

(2c) Well trained at the Haas School of Business, you immediately recognize in the regression results several signs of collinearity. List two. For each of the two, refer to specific features of the regression results above, and explain briefly why it is a sign of collinearity.

Practice Problem # 2: Advertising: TV or Print?

Regression of sales on TV ad spending:

R Square	0.07
Adjusted R Square	0.07
F-Statistic	15.61
p-value for F-statistic	0.0001

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2556.1	233.9	10.93	0.00
TV Ads (millions $)	9.01	2.28	3.95	0.00

Regression of sales on TV ad spending and Print ad spending:

R Square	0.07
Adjusted R Square	0.07
F-Statistic	7.94
p-value for F-Statistic	0.0005

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2486.5	264.4	9.41	0.00
TV Ads (millions $)	5.75	6.18	0.93	0.35
Print Ads (millions $)	3.96	6.97	0.57	0.56

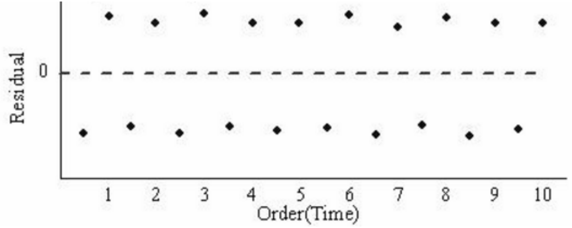
Practice Problem # 2: Advertising: TV or Print?

(2d) Based on the regression evidence, one of your colleagues argues that the company should be spending more on TV advertising, and less on print advertising. Do you agree? Briefly explain.

Practice Problem # 2: Advertising: TV or Print?

(2e) Recall that the data contain TV/Print ad spending and sales during the previous 200 months. Suppose that after you estimated the regression of sales on TV Ads and Print Ads, you plotted the residuals vs month as shown below. Do you detect any violation of the MRM assumptions using this graph?

Note: The graph shows only the first 10 months as an example, but assume that the rest of the months look the same.




even staring at a wall
becomes interesting
while studying