EWMBA 296 (Fall 2015)

Section 8

GSI: Fenella Carpena

December 10, 2015

Agenda for Today

- Practice Problem # 1: Gender Gap in Earnings
- Practice Problem # 2: Diet Burgers
- Practice Problem # 3: Advertising Media
- Practice Problem # 4: More Advertising
- Practice Problem # 5: Wine Prices
- Practice Problem # 6: Adapted from Stock and Watson, Chap 25, Q37

Practice Problem # 1: Gender Gap in Earnings

This problem is inspired by a study of the "gender gap" in earnings in top corporate jobs (Bertrand and Hallock (2001)). The study compares total compensation among top executives in a large set of US public corporations in the 1990s.

(a) Let *Female* be a dummy variable equal to 1 for females and 0 for males. An SRM of log earnings on *Female* yields

	Coefficients	Standard Error
Intercept	6.48	0.01
Female	-0.44	0.05

(i) Interpret the coefficient on *Female*.

(ii) Suppose that the R^2 of the above regression is 0.10. Interpret the R^2.

Practice Problem # 1: Gender Gap in Earnings

This problem is inspired by a study of the "gender gap" in earnings in top corporate jobs (Bertrand and Hallock (2001)). The study compares total compensation among top executives in a large set of US public corporations in the 1990s.

(a) Let *Female* be a dummy variable equal to 1 for females and 0 for males. An SRM of log earnings on *Female* yields

	Coefficients	Standard Error
Intercept	6.48	0.01
Female	-0.44	0.05

(iii) Does this regression suggest that female top executives earn less than top male executives? Explain.

(iv) Does this regression suggest that there is gender discrimination? Explain.

Practice Problem # 1: Gender Gap in Earnings

(b) Two new variables, the market value of the firm (a measure of firm size, in millions of dollars) and stock return (a measure of firm performance, in percentage points), are added to the regression in natural log:

	Coefficients	Standard Error
Intercept	3.86	0.03
Female	-0.28	0.04
ln(MarketValue)	0.37	0.004
Return	0.004	0.003

(i) Interpret the coefficient on ln(MarketValue).

(ii) The coefficient on *Female* is now -0.28. Explain why it has changed from the regression in (a).

Practice Problem # 1: Gender Gap in Earnings

(c) Are larger firms more likely to have female top executives than small firms?

Practice Problem # 2: Diet Burgers

A fast-food chain tests each day that the number of calories in their "Diet-Burger" is no more than 400. Due to imperfections in the cooking processes, the number of calories in their Diet-Burger is normally distributed with standard deviation 30 calories. Let μ denote the population mean calories of the burger. The decision rule adopted by the fast-food chain is to reject the null hypothesis $H_0 : \mu \leq 400$ in favor of $H_a : \mu > 400$ if the sample mean number of calories is more than 410.

(a) If a random sample of size 40 burgers is selected, what is the probability of a Type I error, using this decision rule?

Practice Problem # 2: Diet Burgers

(b) If a random sample of size 10 burgers is selected and the same decision rule is applied, do you think the probability of a Type I error will be (select one):

(1) Lower than in part (a)

(2) The same as in part (a)

(3) Higher than in part (a)

(4) There is not enough information

Practice Problem # 2: Diet Burgers

(c) Suppose that the true mean number of calories is 422 (and the standard deviation is 30). If a random sample of 40 burgers is selected, what is the probability of a Type II error, using the same decision rule?

Practice Problem # 3: Advertising Media

A large consumer products company wants to measure the effect of different local advertising media on the sales of its products. Specifically, they considered TV and newspaper advertising, and also considered providing cents-off coupons in newspapers. Over a period of three months, these variables were measured in 22 cities of roughly equal population and demographics, and the results were analyzed using multiple regression. The variables were:

- SALES = sales in 1,000,000 dollars
- TVAD = TV ad budget, in 10,000 dollars
- NEWSAD = Newspaper ad budget, in 1,000 dollars
- COUPON = 1 if coupons were given out in local newspapers, 0 otherwise.

The regression output is shown below.

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

Practice Problem # 3: Advertising Media

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

(a) Fill in the blank spaces in the regression output labelled (i) and (ii).

Practice Problem # 3: Advertising Media

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

(b) Interpret the coefficient for COUPON in words. Develop a 95% confidence interval for the population coefficient of COUPON and interpret this confidence interval in words.

Practice Problem # 3: Advertising Media

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

(c) For Pittsburgh, a city typical of those studied, the proposed local advertising budgets were \$47,000 for TV ads and \$25,000 for newspaper ads. No coupons were distributed in this area. What is the predicted level of sales in Pittsburgh in dollars?

Practice Problem # 3: Advertising Media

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

(d) If the cost of distributing coupons is \$20,000, which of the three advertising media considered here has the largest predicted effect on sales?

Practice Problem # 3: Advertising Media

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

(e) Based on this regression, which of the three advertising media (if any) does not have a statistically significant effect on sales?

Practice Problem # 3: Advertising Media

	Coefficients	Standard Error	t-Stat	p-value
Intercept	0.376	0.130		
TVAD	0.127	0.017	(i)	(ii)
NEWSAD	0.016	0.003		
COUPON	0.100	0.075		

(f) Let β_2 denote the population coefficient for NEWSAD in this regression. Test the hypothesis $H_0 : \beta_2 \leq 0.01$, $H_a : \beta_2 > 0.01$ at the 5% level. Interpret your results in words suitable for a person who has little background in statistics.

Practice Problem # 4: More Advertising

Suppose a retailer has asked you to help optimize their advertising efforts. You are given quarterly sales data and want to estimate the relationship between sales in each quarter (Y) and the level of advertising each quarter (X).

(a) You suspect that, irrespective of advertising, sales in the fourth quarter of each year are generally higher because of holiday shopping.

(i) Write a regression model that would allow for this holiday season effect. Precisely define all the variables in the model.

(ii) How would you use this regression model to test for a holiday season effect?

Practice Problem # 4: More Advertising

(b) You suspect that in addition to sales in the fourth quarter being higher, that the effect of advertising on sales might differ in the fourth quarter. How would you modify your model in part (a) to accommodate this? Precisely define all the variables in the model. Explain the meaning of each of the parameters in the modified model, including the intercept(s).

Practice Problem # 5: Wine Prices

You are someone who loves wine seeking to use the knowledge you acquired in Data and Decisions to understand how wine prices are determined. You hypothesize that the current market price of fancy wine is affected by the weather during the growing season just before the grapes are harvested. Accordingly, you obtain data on bottles of wine harvested between 1970 and 2000 and sold in auction this year, and hope to estimate the following equation:

$$\widehat{Price} = b_0 + b_1 Vintage + b_2 Temp + b_3 Rainfall$$

where *Vintage* is the year that the grapes where harvested, *Temp* is the ave. temperature during the growing season in the area where the grapes were grown, and *Rainfall* is the amount of rainfall when the grapes were harvested.

Practice Problem # 5: Wine Prices

(a) Unfortunately, the rainfall data you want were not collected in the 1970s for some of the vineyards. You suspect that vineyards producing low-end wines were the last to start collecting weather data. You must decide whether to: (i) run the full regression on a subset of the wines for which rainfall data is available even for the 1970s, or (ii) run a regression without the rainfall variable, but using all of the wines in your sample. Briefly discuss the trade-offs to adopting strategy (i) versus strategy (ii).

Hint: You may want to list the cost(s) and benefit(s) of each strategy.

Practice Problem # 5: Wine Prices

(b) You've decided to instead run the following, shorter regression (i.e., fewer X variables) on your entire sample:

$$\widehat{Price} = b_0 + b_1 Vintage + b_2 Temp$$

but you would like to know by what *percent* the price changes when the vintage changes by a year, all else equal. Select the correct statement:

(i) You should regress $\widehat{Price} = b_0 + b_1 Vintage + b_2 Temp$.

(ii) You should regress $\widehat{log(Price)} = b_0 + b_1 Vintage + b_2 Temp$.

(iii) You should regress $\widehat{Price} = b_0 + b_1 log(Vintage) + b_2 Temp$.

(iv) You should regress $\widehat{log(Price)} = b_0 + b_1 log(Vintage) + b_2 Temp$.

Practice Problem # 5: Wine Prices

(c) Discuss the validity of your regression analysis using these data with respect to making predictions about the price of California wines.

Practice Problem # 6: Adapted from Stock and Watson Chap 25, Q37

(a) The following output summarizes the fit of a response variable Y on a dummy variable D, an explanatory variable X, and their interaction. Does the fit of the model suggest parallel equations for the two groups?

	Coefficients	Standard Error	t-Stat	p-value
Intercept	6.0486	0.2090	28.94	< 0.0001
X	2.0340	0.1800	11.30	< 0.0001
D	-5.9992	0.5063	-11.85	< 0.0001
D $*$ X	-0.1865	0.2645	-0.71	0.4843

Practice Problem # 6: Adapted from Stock and Watson Chap 25, Q37

(b) How would the output change if the coding of the dummy variable D were reversed (so that 0s became 1s and vice versa)?

	Coefficients	Standard Error	t-Stat	p-value
Intercept	6.0486	0.2090	28.94	< 0.0001
X	2.0340	0.1800	11.30	< 0.0001
D	-5.9992	0.5063	-11.85	< 0.0001
D $*$ X	-0.1865	0.2645	-0.71	0.4843

Practice Problem # 6: Adapted from Stock and Watson Chap 25, Q37

(c) Would you drop any of the variables included in the model above? Explain.

	Coefficients	Standard Error	t-Stat	p-value
Intercept	6.0486	0.2090	28.94	< 0.0001
X	2.0340	0.1800	11.30	< 0.0001
D	-5.9992	0.5063	-11.85	< 0.0001
D $*$ X	-0.1865	0.2645	-0.71	0.4843

Thank you for a wonderful semester! :-)