

Problem set 8

Regression

1. Access **bschools2002.xls** which contains data regarding the top 30 business schools based on the 2002 *Business Week* ratings.

Students from better schools might command a higher salary. Comparing a school's mean base salary to its rank might help us understand this relationship.

Use R to solve this exercise.

- (a) Develop a scatterplot for these variables with mean base salary as the dependent variable.
- (b) Does this relationship appear linear?
- (c) Use R to perform a regression of mean base salary vs. rank. Write the estimated regression equation.
- (d) Use your regression equation to estimate the mean base salary for a school ranked eighth.
- (e) Use your regression equation to estimate the mean base salary for a school ranked 25th.
- (f) Use the coefficient on the rank variable to estimate the expected increase in mean base salary from a one-unit improvement in a school's rank. Provide a 95% confidence interval for your estimate.
- (g) How confident are you that the true slope, β_1 , is significantly different from zero?

2. Estimating betas

Access the **stock.xls** dataset and use it to estimate betas for the following stocks: Apple, IBM and HP. Suppose the excess returns on the stock market (as measured by the S&P500, stored under **ESP** in the dataset) were to be negative 20% next month.

- (a) What would you expect to be the excess return on Apple shares next month? How about IBM and HP shares? Base your estimate on the estimated beta and the theoretical CAPM equation; that is, discard the estimated constant (alpha) as we did in the chapter.
- (b) How much money would you expect to lose next month if you had \$ 10,000 invested in Apple shares at the beginning of the month? For the purposes of answering this part of the question only, assume that the risk-free rate next month is 0.25%.

In the example in class, we used the variable **Market** to measure the market excess return. In this problem, we ask you to use an alternative method of measuring the market excess return using the variable ESP. So, for this exercise, use ESP. One problem with the CAPM is that it is not obvious how to measure the market return. Market is a combination of bonds and the S&P 500, and ESP includes only the S&P 500. Finally, all the variables in the dataset are **excess market returns** (i.e. market minus the risk-free interest rate).

3. Access the **retailsales.xls** data file. This data file reports the percentage change in total domestic retail sales and the percentage change in the U.S. GDP over a recent ten-year period (from A.C. Nielsen *Facts, Figures and the Future*, February 2003).
 - (a) Use R to perform a regression of **% chg in Retail Sales** using **% chg in GDP** as the independent variable.
 - i. Write the estimated regression equation.
 - ii. Use the regression to estimate how much a one percentage point increase in GDP will affect retail sales.
 - iii. Provide a 95 % confidence interval for your estimate in part ii.
 - iv. Provide a 90 % confidence interval for your estimate in part ii.
 - v. Using $\alpha = 0.05$, can you reject the null hypothesis that the true coefficient multiplying % chg in GDP is zero?
 - (b) Use the regression from the previous questions.
 - i. Predict the **% chg in Retail Sales** in a year where the GDP increases by 3.0%.
 - ii. Provide a 95% prediction interval for your estimate.
 - iii. Provide a 98% prediction interval for your estimate.
 - iv. Using the same prediction, estimate the probability that the **% chg in Retail Sales** will be greater than 8.5.
 - (c) Overall how would you rate the quality of this regression? Justify your answer.

4. Valuing and MBA for yourself

The purpose of this example is to compare the “value-added” of two different business schools by looking at the incomes of the student body prior to beginning the MBA program, and comparing it to the incomes

after completing the program. The data consist of information on 400 students, half from school A and the other half from school B.

'pre-MBA' = income in year beginning the program, in thousands of dollars

'post-MBA' = income in year after completing the program, in thousands of dollars

'school' = a dummy variable equal to 0 for students attending school A, and 1 for students attending school B

The following regression output was obtained:

Regression: postMBA			
	constant	preMBA	school
coefficient	24.659	1.83628	1.732
std error of coef	1.868	0.04178	1.136
t-ratio	13.2000	43.9600	1.5200
p-value	0.0000%	0.0000%	12.8000%
standard error of regression		11.26	
R-squared		83.10%	
adjusted R-squared		83.00%	

- (a) Explain clearly, and as concisely as possible, the interpretation of the coefficient of the school variable.

Suppose we define a new variable as follows:

'school'*preMBA' = 'school' multiplied by 'preMBA'.

We redo the regression with this extra variable added as another predictor and obtain the following regression output:

Regression: postMBA				
	constant	preMBA	school	school*preMBA
coefficient	30	1.70426	-7.314	0.23227
std error of coef	2.67	0.06306	3.447	0.08364
t-ratio	11.2300	27.0300	-2.1200	2.7800
p-value	0.0000%	0.0000%	3.4000%	0.6000%
standard error of regression		11.17		
R-squared		83.40%		
adjusted R-squared		83.30%		

Answer the remaining questions, basing your answers in this second regression:

- (b) Suppose your income this year is \$15,000 and you are choosing between the two schools' programs. Assume the two schools have the same fees, similar locations, etc. Which one should you choose? What if your current income is \$65,000?

We ask R to predict the post-MBA income of someone entering school A with a pre-MBA income of \$40,000 and to give a 90% confidence and prediction intervals for post-MBA income. This gives the following additional output:

for post-MBA income. This gives the following additional output.

predicted value of		98.171
standard error of estimated mean		0.79
confidence level	90.00%	
confidence limits for prediction	lower	79.71
	upper	116.632
confidence limits for estimated mean	lower	96.868
	upper	99.474

- (c) What is the predicted post-MBA income of graduates of school A having pre-MBA income of \$40,000? If 60 students entering school A this year have pre-MBA incomes of \$ 40,000, about how many of those students do you estimate will make less than \$ 80,000 the year they leave?
 - (d) Explain briefly the meaning of the R-squared statistic in this context (i.e., do not simply say what it means in the abstract, but say what it means for this regression and application).
 - (e) In a few, non-technical words, summarize what the difference seems to be between the two schools.
5. Access the **eurodata2b.xls** dataset. This file contains information about unemployment and wage growth in Germany and Greece. The dummy variable Germany is set to 1 in Germany and 0 in Greece.
- Use R to perform a regression of Wage Growth vs. Unemployment, Germany and DE*Unemployment.
- (a) Write out the full estimated regression equation
 - (b) Write out the estimated regression equation for Germany
 - (c) Write out the estimated regression equation for Greece
 - (d) How does a one percentage point increase in unemployment relate to the growth rate of wages in Germany?
 - (e) How does a one percentage point increase in unemployment relate to the growth rate of wages in Greece?
 - (f) Estimate the difference in how unemployment relates to wage growth between the two countries.

- (g) Provide a 95% confidence level for the difference in how unemployment relates to wage growth between the two countries.
- (h) Predict the growth rate in wages for each country in a year that has 3% unemployment.
- (i) Provide a 90% confidence interval for each prediction from the previous question.

6. Access the **pizzasales.xls** file.

The Waialua Pizza Company is a medium-sized chain of pizzerias located at beaches all over the South Pacific. The chain is known for its delicious pizzas served at all the nice beaches, and it is known for its use of statistical techniques to improve operations.

The company has obtained data reflecting its sales in its 50 beachfront stores. The Waialua Pizza Company feels the income levels of the nearby community and the presence or absence of competition might be major factors in determining sales.

The following variables were tallied:

Sales = \$ per day

Income = Average per-capita income in \$ per week in the surrounding neighborhood

Competitor = 1 when one or more competing pizzerias are located within 1/2 mile;

0 when no other pizzerias are located nearby

- (a) Conduct a regression of Sales vs. Competitor (only use this one independent variable for now) and use the results to answer the following questions:
 - i. Estimate the daily sales for a store that has no competition.
 - ii. Estimate the daily sales for a store that faces competition.
 - iii. Calculate the difference between your two estimates and comment on the practical and statistical significance of this gap.
 - iv. Provide a 95% confidence interval for the effect of competition on sales.
 - v. What percentage of the variance in sales can be explained using only the Competitor variable?

- (b) Conduct a regression of Sales vs. Income (only use this one independent variable for now) and use the results to answer the following questions:
- Estimate the daily sales for a store whose neighborhood income is \$200 per week.
 - Estimate the daily sales for a store whose neighborhood income is \$300 per week.
 - Estimate the impact of a \$ 100 increase in neighborhood income per week on sales.
 - Provide a 95 % confidence interval for your estimate in c.
 - What percentage of the variance in sales can be explained using only the Income variable?
- (c) Create a scatterplot of Sales vs. Income and have R plot the regression line as well. Does the picture reveal any likely opportunities to improve your model?
- (d) Construct a new variable, Comp*Inc , by multiplying the Competitor and Income variable together. Run a regression to predict sales using all three variables: Competitor, Income and Comp*Inc .
- Is the Competitor variable in this model statistically significant?
 - Estimate the daily sales for a store without competition whose neighborhood income is \$300 per week.
 - Estimate the daily sales for store with a competitor whose neighborhood income is \$ 300 per week.
 - Compare the answers to part ii. and iii. Reconcile the results of this comparison with your answer to part i.