Assigned Problems for the "Fundamentals" Unit

Please hand in a hard copy of the Excel data problems, each on a separate page, at the start of class on June 13. Electronic submissions will not be accepted. Problems worked entirely by hand are not to be turned in, just discussed in class as needed. Problems marked (*) are not assigned, just extras in case you want more practice. Data sets for all listed homework problems are available on Blackboard or on the CD that (hopefully) came with your textbook.

Ch 2 Graphical and Tabular Descriptive Techniques

2.2 Nominal Data	2.21, (22*) 2.25	**Excel-how-to** p. 20 **Excel-how-to** p. 22
2.3 Interval Data	2.44, (49*),51,(52*)	**Excel-how-to** histogram p 33 stem and leaf p. 43 ogive p.46
2.4 Time Series Data	2.(59*, 60*), 61, 64	**Excel-how-to** p. 51
2.5 Two or more Sets of Nominal Data	2.75, (77*)	**Excel-how-to** p. 59
2.6 Two Interval Variables	2.(87*), 92, 93	**Excel-how-to** p.67

Ch 4 Numerical Descriptive Techniques

4.1,5,21, 23, 24, 45	Hand calculations
p. 150 4.84, (88*)	Use **Excel, how-to** p. 117, 119
Case 4.3 p 151	" " "

Ch 8 Probability Distributions (Emphasis on the Normal Distribution)

8.18, 20, 22	Hand calculations
26, 54, 56, 64	Use Table on B-8, B-9

Ch 9 Sampling Distributions, σ known

9.1 … of the mean	9.5, 6, 7, 8, 9, 18, 24, 28	All hand calculations

Ch 10 Intro to Estimation

10.1 Concepts of Estimation	10.1 – 8	Handwritten
10.3 Selecting Sample Size	10.43, 44, 52	Hand calculations

Ch 11 Intro to Hypothesis testing

11.1 Concepts (Type I and II errors)	11.4, 5	Handwritten
11.3 Calculating Probability of Type II Error	11.63, 64	Handwritten

Ch 12 Inference about a population

12.1 Inference about μ, σ unknown	12.(30*), 32, 38, 40	**Excel-how-to** p. 385
12.2 Inference about σ^2	12.50, 52, (53*)	**Excel-how-to** p.399
12.5 Applications in Accounting Auditing	12.98, 101	

(For Chapter 12 problems only, in addition to the output given by Excel data analysis tool, calculate the sample mean and standard deviation in Excel, and then show how either how the hypothesis test is conducted, or the confidence interval is constructed, as appropriate. That is, type that into the spreadsheet.)

Practice Presentation for the "Fundamentals" Unit

On the web site "Visualizing Economics" you will find the U.S. Census Bureau data for the distribution of household income, up to $250,000, in 2007. Create a five slide presentation illustrating

- this distribution,
- the cumulative relative frequency, that is, an ogive, and
- a table of basic descriptive statistics

using appropriate, well-designed graphics and tables. Your first and last slides should contain a title and conclusion. The remaining three slides should contain one graphic/table each, along with one conclusion about the data based on that graphic.

Please hand in a hard copy of your presentation at the beginning of class on June 13.

Assigned Problems for the Analysis Unit

Please hand in these problems at the beginning of class on June 29 in the format prescribed in class. Problems marked (*) are not assigned, just extras in case you want more practice. Data sets for all listed homework problems are available on Blackboard or on the CD that (hopefully) came with your textbook.

For each problem, describe your data using one (or more) of the techniques utilized in the first unit of this class. Also identify whether the key assumptions maintained by the ANOVA / Regression appear to be met reasonably closely, or not.

Last time the "how to's" were listed for convenience. This time they are not. You are expected to read the pertinent sections of each chapter, and the how to's will be found in these sections.

Ch 14 Analysis of Variance

(14.10*, companion to 14.42), 14.42, 14.60, 14.62, 14.65

Ch 16 Univariate Linear Regression

16.83, 16.88

Ch 17 Multiple Regression (17.1-17.3)

(17.6*), 17.40, 17.42

Ch 18 Model Building (18.1)

(18.43*), 18.48

Ch 20 Time Series and Forecasting

20.25

General Regression Questions

1. Go to the web site of any ISD in the Houston area and find their salary schedule for full time teachers with a bachelor's degree. (Most, but not all, post them online.) Copy and paste this data into a well-organized, labeled spreadsheet titled "teachers.xls". Then:

a) Plot salary (dependent variable) vs. experience (independent variable) for teachers holding bachelor's degrees, using a scatterplot. Make the plot look nice using the techniques described in class.

b) Based on the graph, make a guess at a linear formula that you can use to predict salary given experience. (Don't use an analysis tool,, just literally make your best guess.) Form an additional column in your spreadsheet containing the salary estimates your formula gives you. Plot the predicted values on the same graph that you created above.

c) Make an additional column that contains the residuals–the difference between the value your formula predicts and the value given by the salary schedule. The SUMSQ function can be used to calculate the sum of squared residuals. Do this at the bottom of this column.

d) Use the regression tool in the Excel Analysis toolpack to estimate a regression of salary on experience. Calculate the residuals and the sum of squared residuals as before. The sum of squared errors is smaller using which formula: yours or the regression? Interpret your coefficient estimates.

e) Then reproduce your graph for parts a and b, except that your predicted values are now based on the regression coefficients that you estimated. The regression tool will produce this graph automatically.

2. On Blackboard you will find a spreadsheet, baseball.xls, containing the following data: Major League Baseball teams' attendance (number of people for the full season), the team's winning percentage (expressed in "points"–if you won 56.3% of your games, this variable reads 563), per capita state income (in dollars) in the state where the team is located, stadium age in years, the number of all-stars playing on the team that season, and the population of the metropolitan area where the team is located. These data cover most MLB teams for the period 1990-1997.

We will use these data to examine the economics of baseball. In particular, we observe three phenomena in MLB: 1) teams are obsessed with winning, 2) owners are always fussing about wanting new stadiums, and 3) superstars are paid a whole lot more than average players are. Are these phenomena supported by the economics of running a baseball team? That is, does revenue respond to winning percentage, the number of superstars, and the age of the stadium?

a) Regress attendance–a major source of team revenue–on the other variables. At typical significance levels, does the regression support the claim that economic forces explain the three phenomona described above? For each, answer yes or no, and describe why or why not.

b) Conduct an ANOVA to determine whether winning percentage exerts a statistically significant effect on attendance. Which type of ANOVA is best to conduct here? What do you conclude?

c) In this situation, which is the better analysis technique, regression or ANOVA? Why?