MGT 2250 Management Statistics

Sample Test 3

1. R&D Expenses This data file contains a variety of accounting and financial values that describe 493 companies operating in several technology industries in 2004: software, systems design, and semiconductor manufacturing. One column gives the expenses on research and development (R&D), and another gives the total assets of the companies. Both columns are reported in millions of dollars.

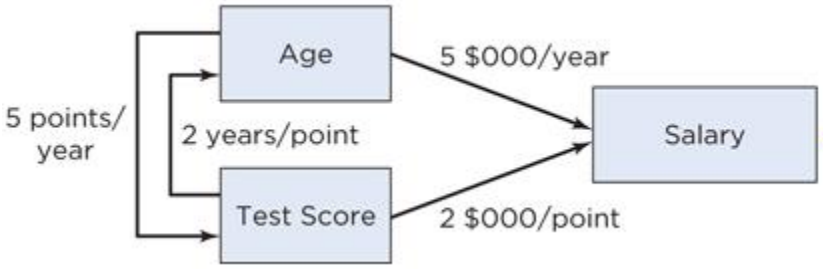
- (a) Scatterplot R&D Expense on Assets. Does a line seem to you to be a good summary of the relationship between these two variables? Describe the outlying companies.
- (b) Estimate the least squares linear equation for R&D Expense on Assets. Interpret the fitted intercept and slope. Be sure to include their units. Note if either estimate represents a large extrapolation and is consequently not reliable.
- (c) Interpret the summary values r^2 and s_e associated with the fitted equation. Attach units to these summary statistics as appropriate. Does the value of r^2 seem fair to you as a characterization of how well the equation summarizes the association?
- (d) Inspect the histograms of the *x*- and *y*-variables in this regression. Do the shapes of these histograms anticipate some aspects of the scatterplot and the linear relationship between these variables?
- (e) Plot the residuals from this regression. Does this plot reveal patterns in the residuals? Does s_e provide an adequate summary of the residual variation?

2. Seattle Homes This dataset contains the listed prices (in thousands of dollars) and the number of square feet for 28 homes in the Seattle area. The data come from the Web site of a Seattle realtor offering homes in the area for sale. For the SRM to work, we need to formulate the model in terms of cost per square foot. Use the selling price per square foot as *Y* and the reciprocal of the number of square feet as *X*. (*Note:* If you keep track of the dimensions for the slope and intercept, you'll see that one represents fixed costs and the other, marginal costs.)

- (a) Give a 95% confidence interval for the fixed cost (the portion of the cost that does not change with the size of the home) associated with these home prices, along with a brief interpretation.
- (b) Give a 95% confidence interval for the marginal costs (the cost that is related to number of square feet), along with a brief interpretation.
- (c) How much might a buyer pay, per square foot, for a specific home with 3,000 square feet in the Seattle area? Give a range, rounded appropriately, to show the buyer.
- (d) How much in *total* might a buyer pay for a 3,000-square-foot home in the Seattle area? Give a range, rounded appropriately, to show the buyer.

3. The human resources department at a firm developed a multiple regression to predict the success of candidates for available positions. Drawing records of new hires from five years ago, analysts regressed current annual salary on age at the time of hire and score on a personality test given to new hires. The

path diagram below summarizes the fitted model. Age is coded in years, the test is scored from 1 to 20, and annual salary is in thousands of dollars ($000).



- (a) Write down the equation for the multiple regression model.
- (b) Which is larger: the direct or indirect effect of test score?
- (c) Find the marginal slope of salary on test score.
- (d) If you were a new applicant and could take a special course that in a week's time could raise your test score by 5 points, would the course be worth the $25,000 being charged? Which slope is relevant: marginal or partial?

4. Download Before purchasing videoconferencing equipment, a company tested its current internal computer network. The tests measured how rapidly data moved through its network given the current demand on the network. Eighty files ranging in size from 20 to 100 megabytes (MB) were transmitted over the network at various times of day, and the time to send the files (in seconds) recorded. The time is given as the number of hours past 8 A.M. on the day of the test. Use the transfer time as the response, with the file size and time of day as explanatory variables.

- (a) Examine scatterplots of the response versus the two explanatory variables as well as the scatterplot between the explanatory variables. Do you notice any unusual features in the data? Do the relevant plots appear straight enough for multiple regression?
- (b) Do you think, before fitting the multiple regression, that the partial slope for the file size will be the same as its marginal slope? Explain.
- (c) Fit the multiple regression of the transfer time on the file size and the time of day. Summarize the estimates obtained for the fitted model.
- (d) Does the fit of this model meet the conditions of the MRM?
- (e) Compare the sizes of the *t*-statistics of the fitted model to the overall *F*-statistic. Do these tests agree with each other?
- (f) Compare the confidence interval for the marginal slope for file size to the confidence interval for the partial slope for file size. How are these different?
- (g) Does the path diagram for the multiple regression offer a suggestion for the differences noticed in the previous questions?