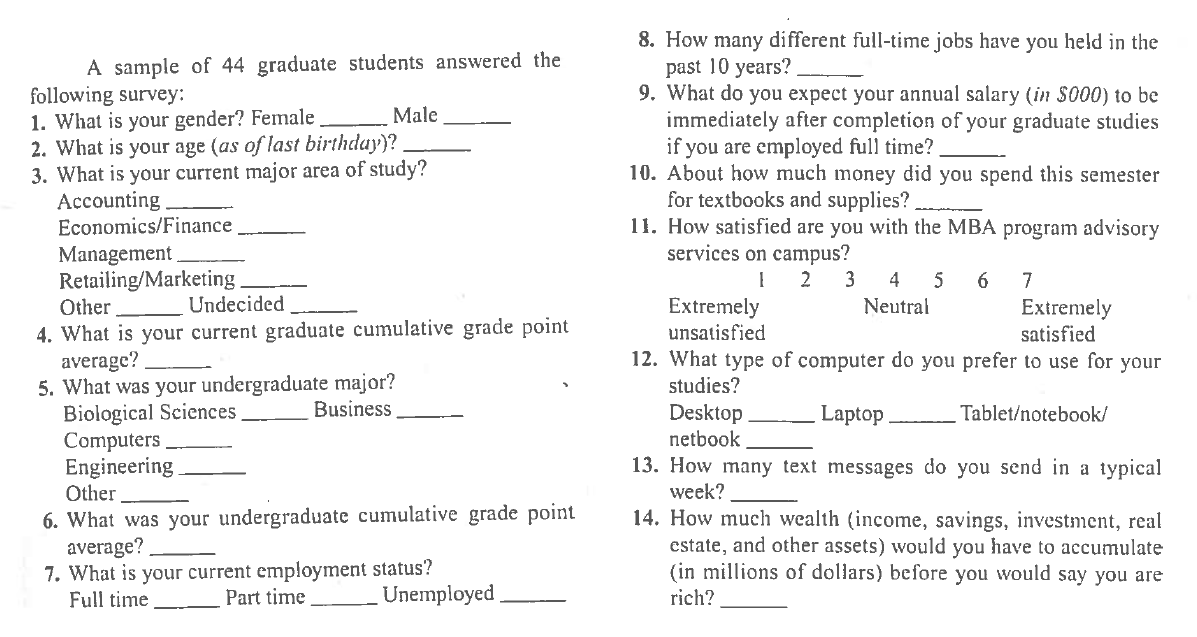
Name:

Test 1

Acts 170

Use the data file ‘GradSurvey’. A list of questions that were asked to graduate students is below.



The first few lines of data look like this:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Gender | Age | GMajor | GGPA | UMajor | UGPA | Employ | FTJobs | Salary | Spending | Advisor | Computer | Texts | Wealth |
| 1 | Male | 26 | Economics/Finance | 3.2 | Other | 3 | Full-Time | 3 | 50 | 70 | 6 | Tablet | 100 | 1 |
| 2 | Male | 25 | Other | 3.2 | Biological Sciences | 3.1 | Full-Time | 2 | 140 | 150 | 4 | Laptop | 200 | 10 |
| 3 | Male | 26 | Other | 3 | Other | 2.9 | Unemployed | 1 | 100 | 200 | 4 | Laptop | 500 | 1 |
| 4 | Male | 23 | Management | 3 | Business | 3.8 | Part-Time | 0 | 50 | 300 | 4 | Laptop | 0 | 0.9 |
| 5 | Male | 27 | Retailing/Marketing | 3 | Business | 3.9 | Full-Time | 1 | 80 | 150 | 4 | Desktop | 100 | 10 |

1.)

a. Compute descriptive statistics for UGPA and Texts. Looking at their mean and median, what type of skewness do you expect and why?

b. Create boxplots, one on top of the other for UGPA and Texts. Does the shape of the distributions correspond with what you expected above? Comment on anything else noticeable in the plots.

c. Create a side by side histogram and a (normal) qq plot for Salary. Comment on each plot.

d. Take a log transformation of Salary and make a side by side histogram of the original variable and the log one. Comment on the plots.

e. Make boxplots of Undergraduate GPA by their Undergraduate major. Comment on the plots.

f. Make a scatterplot of undergraduate GPA and Salary, with Salary as the y variable. One student, Sebastian, has a very unusual expectation of his future Salary. Put his name close to his point as a label on the plot.

g. Produce a table of correlations for all of your quantitative variables. What two variables are most highly correlated with Wealth?

h. Fit a regression model using Wealth as the outcome variable and all of the quantitative variables, as well as Gender, Employ, and Computer as the explanatory variables. Comment on R² and Ra². Interpret Ra².

i. Do a hypothesis test to determine whether Advisor rating is a significant predictor of Wealth. Provide hypothesis statements, t test statistic, t critical value, and a decision. Use alpha = 0.05.

j. Build a new model removing the three variables UGPA, Spending, and Advisor since they were especially terrible. Compare this new model’s s to the s from the first model. Did you expect this change given the p-value of the variable that you dropped? Explain.

k. Using the model from part (j) and an alpha = 0.10, which variables are significant? How do you know?

l. Using the model from part (j), interpret the slope for the full time jobs variable.

m. Using the model from part (j), interpret the slope for the gender variable.

n. Suppose you have a 30 year old female, who has a graduate GPA of 3.5, has had 2 full time jobs, sends 100 texts a week, an expected salary of 90 thousand dollars, is currently employed part time, and prefers to use a laptop. Use the model from part (j) and find a 95% prediction interval for her wealth needed to say she is rich. Interpret this interval. (Hint: You may need to use “ “ for your categorical variables to work. Like “Female”.)

o. What is the difference between a confidence interval and a prediction interval? You do not need to find a confidence interval, just explain to me the difference between what the 2 find.

p. Produce a scatterplot matrix of all quantitative variables you used from your model in part (j) and Wealth. Is there anything noticeable in your plots? If you could square only 1 term, which one would you choose? (You do not need to actually do this.)

q. Take GGPA from the model in part (j) and compute its partial correlation coefficient. What is the interpretation of this coefficient?

r. Produce an added variable plot for the partial correlation coefficient above. Comment on the difference of the original correlation and scatterplot and the partial correlation and added variable plot now.

s. Take a subset of your original dataset by using only students who have greater than a 3.0 graduate GPA. Redo the model from part (j). Comment on any notable changes.