E-mail from Dr. Joan Boyd at YSU:

Hello Dr. Quinn,

John Holcomb gave me your name to possibly assist me in analyzing the results of my patient record search. Wow, is he extremely busy these days! I had my graduate student create the attached Excel spreadsheet. John said that you might have students that could work on this. I am trying to get a manuscript out and any help you can provide is appreciated.

Best,

Joan

How should we reply?

Questions for Dr. Boyd:

* What is the unit of analysis?
* What is your research question(s)? May need more clarification.
* Could you clarify the codes for sex: 1 and 2; age groups: 1 through 5?
* What does each variable mean and what are the values?
* Who did the data import? There looks like data inaccuracies. Is there someone 730 years old? What would we do with that value?
* Should there be missing data? Strange values?
* Are all the subjects independent?
* How soon do you need the analysis?

**OBJECTIVE**: To determine if significant gender differences existed between subjects 65 years of age and older with regard to calcium, phosphorous, and alkaline phosphatase levels.

**DESIGN**: The researchers performed a retrospective chart review of laboratory procedures performed in 6 different physician practices. The data consisted of 178 subjects representing 92 males and 86 females age 65 or older.

**DATA SOURCES**: Patient data were obtained from the charts housed in a cardiac care center. Subjects, with charts preceding them, were referred by a physician to the cardiac center. The laboratory procedures had been performed previously in the laboratories of the referring physicians.

**DATA DICTIONARY**:

|  |  |  |
| --- | --- | --- |
| Column | Variable | Notes |
| OBSNO | Chart number | Can be used as a reference to check data entry. |
| AGE | Age in years |  |
| SEX | Gender of subject | 1=Male; 2= Female |
| LAB | Laboratory used | 1=Metpath; 2=Deyor; 3=St. Elizabeth; 4=CB Rouche; 5=YOH; 6=Horizon |
| ALKPHOS | alkaline phosphatase (IU/L) |  |
| CAMMOL | calcium (mmol/L) |  |
| PHOSMMOL | inorganic phosphorus (mmol/L) |  |
| AGEGROUP | Age group | 1=65-69; 2=70-74; 3=75-79; 4=80-84; 5=85-89 years |

The **first task** of the assignment is to check the validity of the data. Determine if this is a "messy" data set with variable values that appear incorrect. Attempt to recover the correct values by looking up the true values from the actual data records. Copies of these can be found on <http://academic.csuohio.edu/holcombj/clean/bigtable.htm>. Create a “clean” data set.

The objective of the study for which you will analyze the data was to determine if significant gender differences existed between subjects 65 years of age and older with regard to calcium, phosphorous, and alkaline phosphatase levels (Boyd et al., 1998). The researchers performed a retrospective chart review of laboratory procedures performed in 6 different physician practices. The data consisted of 178 subjects representing 92 males and 86 females age 65 or older. In the data set, there are three discrete variables, sex, lab, and agegroup. The other variables of age, alkphos - alkaline phosphatase (IU/L), cammol - calcium (mmol/L), and phosmmol – inorganic phosphorus (mmol/L), are continuous.

1. The first task of the assignment is to check the validity of the data. Determine if this is a "messy" data set with variable values that appear incorrect. Attempt to recover the correct values by looking up the true values from the actual data records. – DISCUSSION DONE
2. Once the data is "clean", perform a summary analysis of the three discrete variables (sex, lab, and agegroup). For the variables alkphos, cammol and phosmmol, report the mean, median, standard deviation, min and max broken down by sex. Also summarize the variables alkphos, cammol and phosmmol in a similar way with the factor variable as lab.
3. Then construct side by side box plots of the variables alkphos, cammol, and phosmmol with the factor variable as sex. Then construct side by side box plots of the alkphos, cammol, and phosmmol continuous variables with the factor variable as lab.
4. Do you believe a significant difference exists in alkphos, cammol, or phosmmol levels with respect to sex? Why or why not? Do you believe a significant difference exists in alkphos, cammol, or phosmmol levels with respect to lab? Why or why not?
5. Conduct a blocked ANOVA analysis using laboratory as the blocking variable. Explain the results.
6. Suppose Mr. and Mrs. Contrarian are married and Mrs. Contrarian has lower calcium than Mr. Contrarian. She refuses to believe the results of the study that men tend to have lower calcium than women because she has lower calcium than her husband. Using your results to question #3, explain to Mrs. Contrarian the flaw in her thinking.
7. One of the objectives of this research was to propose a reference range of values that are to be considered “normal” for calcium, inorganic phosphorus, and alkaline phosphatase. Looking at the results for cammol alone for each of the labs, explain why a single reference range is so difficult to establish.

Write a short, professional report that addresses each of questions 2 through 7. Computer output should be included in an Appendix. In the body of the report, please use professional looking tables and not simply computer output. You may assume a reader understands statistical analysis.