

Stat 311 Written Assignment 1

Due uploaded to Canvas by 11:45 PM on Wednesday April 19th

This assignment covers material in Lessons 1 and 2. You will be using the data in Calcium.csv to perform the tasks and answer the questions listed in this handout. The data set includes blood calcium (Calcium) and inorganic phosphate (InorgPhosph) levels (mmol/l) for adults between 65 and 89 years old. Other data included in the data set include Age, Sex, Lab, and AgeGroup. Calcium, InorgPhosph and Age are continuous variables, whereas Sex, Lab and AgeGroup are categorical variables.

Assignment (See Assignment Guidelines on the next page)

1. Summarizing the distribution of blood calcium levels.

- a. Create one table that contains the seven number summary statistics for Calcium for all individuals. Also add rows for the same summaries by Sex. The seven number summaries should include min, Q1, median, Q3, max, mean, and SD. Also report sample size (n).

For consistency in assignments, complete the following table. Label this Table 1 and make sure to add a table caption to the top of the table. The statistical software may report the summary statistics to many decimal places. Where applicable, be sure to round your summary statistic values to two or three decimal places.

Subset	n	Min	Q1	Median	Q3	Max	Mean	SD
All								
Females								
Males								

- b. Create three histograms for blood calcium levels, one for all the data and one for each sex. All three histograms should be in the same figure (use the `par(mfrow=c(2,2))` in R to set the window). Label this Figure 1 and make sure to add a figure caption at the bottom of the figure.
- c. Create two boxplots for blood calcium levels, one for all the data and one comparative boxplot by sex. The two boxplots should be in the same figure (use `par(mfrow=c(1,2))` in R to set the window). Label this Figure 2 and make sure to add a figure caption at the bottom of the figure.
- d. Summarize your overall observations for blood calcium levels using all the summary information in parts a) – c). Comment on overall summaries and summaries by sex. What is the overall distribution of blood calcium for these individuals? Comment on shape and skewness, presence of outliers, etc. Are your observations the same by sex? If not, how are things different?
- e. Would a value of 1.9 mmol/l blood calcium be considered an outlier among all females in the data set? Use the 1.5 IQR criterion to justify your answer.

Stat 311 Written Assignment 1

Due uploaded to Canvas by 11:45 PM on Wednesday April 19th

2. Summarizing the relationship between blood calcium and blood inorganic phosphate levels.
 - a. Create a scatterplot of Calcium (y) on Age (x). Use different colors or symbols to separate the sexes on the plot. Summarize any associations you observe.
 - b. Using software, calculate the sample correlation coefficient between Calcium and Age for all individuals, and separately by sex. In R use the `cor()` function, but add the argument `use="complete.obs"` into the function call. Interpret the three sample correlation coefficients.
 - c. Use the `lm()` function in R to find the simple linear regression equation for blood calcium (y) on age (x) for males only. Include the R output from the summary and write out the regression equation.
 - d. Create a residual plot for your regression from 1c. Label this Figure 3 and make sure to add a figure caption at the bottom of the figure. Interpret the residual plot.
 - e. What is the coefficient of determination for the regression from 1c? Interpret the value.
 - e. Use R to get the mean Age and the standard deviation for Age for the males only. Then combine this with the summary statistics in question 1a and the correlation coefficient from question 2b to show "by-hand" calculations for the regression equation of blood calcium (y) on age (x) for males only. Is the regression equation from your "by-hand" calculations similar to the equation from the call to `lm()`?
 - f. Using the equation from 1c, what is the estimated blood calcium level for an 83-year-old male?
 - g. Do you recommend this equation for estimating blood calcium levels for males, ages 65 – 89? Explain.
3. Create a contingency table using Sex (rows) and AgeGroup (columns). Include the table in your submission and then answer the following. You must show your work for parts a) through e). Report all values as percentages rounded to one decimal point.
 - a. What is the marginal distribution for sex?
 - b. What is the marginal distribution for age group?
 - c. What percentage of subjects are males and in the 65 – 69 age group (joint percentage)?
 - d. What is the conditional distribution of age group given that a subject is female?
 - e. What is the conditional distribution of sex given that a subject belongs to the 70-74 age group?

Stat 311 Written Assignment 1

Due uploaded to Canvas by 11:45 PM on Wednesday April 19th

Assignment Guidelines

We expect the following:

1. Divide and conquer! This is a group assignment, so it is reasonable to divide up the work among members. We expect that you will assign initial primary responsibility of 1 or 2 problems to each member. **As such, please include the name(s) of the primary student assigned to each problem.**
2. Pay attention to formatting! We expect a well formatted/easy to read table paper the includes completely labeled graphs, including x and y axes labels that include units, as well as proper use of Table and Figure captions.
3. Grammar matters! We expect the use of complete sentences and coherent thoughts. None of the interpretations need to lengthy. Rather, they need to be well thought out and succinctly written. We are looking for substance, not “brain dump.”
4. The final submittal sounds like it was written by one person! All members will receive the same grade, so we are expecting that each group member reads through the final product before submission. The final submission should be edited to read as one cohesive assignment.
5. Work well together! There will be a peer review required after the assignments are complete. Peer review results may be used to adjust individual grades upwards or downwards, depending on the results.
6. Pay attention to the final file! We expect the problems to be in order and all pages in the right direction. Remember that your assignment must be submitted as a single pdf, doc, or docx file. You are not required to type your assignment (although we highly recommend that you type the assignment). You can handwrite, insert graphs, and scan—but, be mindful of your final product. Look at your final file before uploading to make sure it is readable (not blurry, not too small), pages are not cut off, and pages are in order and are in the correct orientation.