

Raising Expectations for ST516 Homework

At this point, you should all be familiar with R and the mechanics of data analysis.

Consequently, I am going to raise my expectations for the two remaining homework assignments so that it more accurately reflects what a working analyst/data scientist would produce.

Any deviations from the expectations will result in points being deducted.

Here is what I will be looking for in all remaining submissions:

- Your name, the class, the date, and the number of the homework assignment in the header of **both written and R submissions**.
 - Why? Because I want to know immediately who produced this document, when, and what it was for.
- Correct numbering of answers in **both written and R submissions**.
 - Why? Because when I start reading your answer to question 4, it should match number 4 in the homework. If you skip a conceptual question, make sure your numbering scheme stays true. Better yet, write ‘Question 3: Skipped’ if it is one of the questions you choose not to answer.
- No raw code in your written portion without a description of its purpose/function.
 - Why? Because code is slow to interpret, especially for someone without extensive knowledge of the coding language you are using. Get used to explaining what every line does. Statements such as:

```
# Simulate 100 standard normal random variables  
x <- rnorm(100)
```

make your code far more readable and easy to follow.
- No raw output in your written submission without interpretation.
 - If you do a t test for example and just copy/paste the raw output from R into your written submission without any sort of explanation of what the output means, I’m not going to read it.
 - Why? Because it is your responsibility to interpret the output for me. Your future audiences may have no knowledge of R nor how to interpret output for themselves. I am far more concerned about your ability to read the output and interpret it in a concise and erudite fashion.

- Do not require me to run R code to grade your assignment.
 - Why? Because output is meaningless without context and interpretation. And because your future audiences are not going to have R up and running when you hand them your work.
 - Your .pdf submission should be a complete, self-contained work that answers all questions on the assignment without the need to execute code. The only time I'm going to run your code is if I think you did something wrong and I'm trying to find the problem.
 - Note that I will still check your code for formatting.
- Proper formatting of R files which includes:
 - Numbering each section of code so it corresponds to its respective homework problem.
 - Not extending a line of code past 80 characters.
 - Using comments to briefly explain what the code is doing.
 - Why? Code is slow for humans to read and interpret. Comments, spacing, and good formatting makes diagnosing problems in the code far easier. Huge blocks of code with no spacing or comments means I have to parse through each line trying to figure out what you are trying to accomplish. I don't like doing that.
- Proper written answers to questions regarding hypothesis tests.
 - Any question about a hypothesis test should include the following
 - * A properly formatted statement of the null and alternative hypotheses which includes a description of the parameter being tested, proper units, and a sentence explaining each hypothesis.
 - * A conclusion that describes the results of the test, the parameter in question, the testing method, the p -value, the sample size, the point estimate, and a confidence interval.
 - * No double-negatives.
 The sentence "There is no evidence that the mean is not 15 units" is confusing. "There is insufficient evidence to suggest the mean differs from 15 units" is much more clear and conveys the same information.