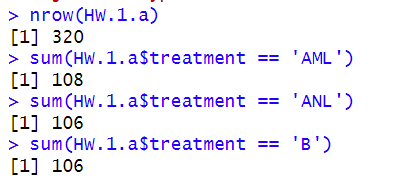
**Question 1**

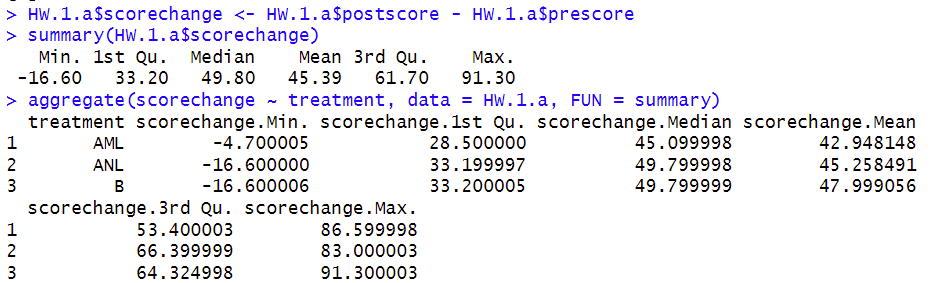
1. Counterfactual is an unobserved scenario that helps understand what would have happened (potential outcome) if an individual/group was or wasn’t treated. For example, for a treated individual (i) we know the outcome (y) given they were treated (x=1) but can’t observe their outcome if they were not treated (x=0). This unobserved outcome is a counterfactual.
2. Selection refers to systematic differences between the treated and control group even before the treatment has occurred. It is a component of the difference in outcomes between the treated and untreated that is not related to the treatment itself. For treatment to have a causal effect, selection should be eliminated, that is, the assignment of the treatment should be random.
3. The causal effect of variable a (explanatory) on variable b (outcome) is the change in variable b associated with a change in variable a, given all other variables are controlled for (no selection).
4. The "back door" is a term used to describe unobserved variables that may be associated with the treatment and outcome, biasing the true relationship between them. Identifying and controlling for unobserved variables helps compute the true causal effect.
5. Randomized Controlled Trial (RCT) is a research method where the assignment into treatment and control groups is random. Random assignment eliminates (or minimizes) selection bias, which allows a more accurate assessment of the treatment’s causal effect on the outcome.
6. Pre-experimental balance refers to the process of making sure the treatment and control groups are as similar as possible in their characteristics before the experiment. It ensures that the effect of other confounding variables is eliminated. One example, would be the proportion of sick people in a clinical trial should be similar amongst the treatment and control groups

**Question 2**

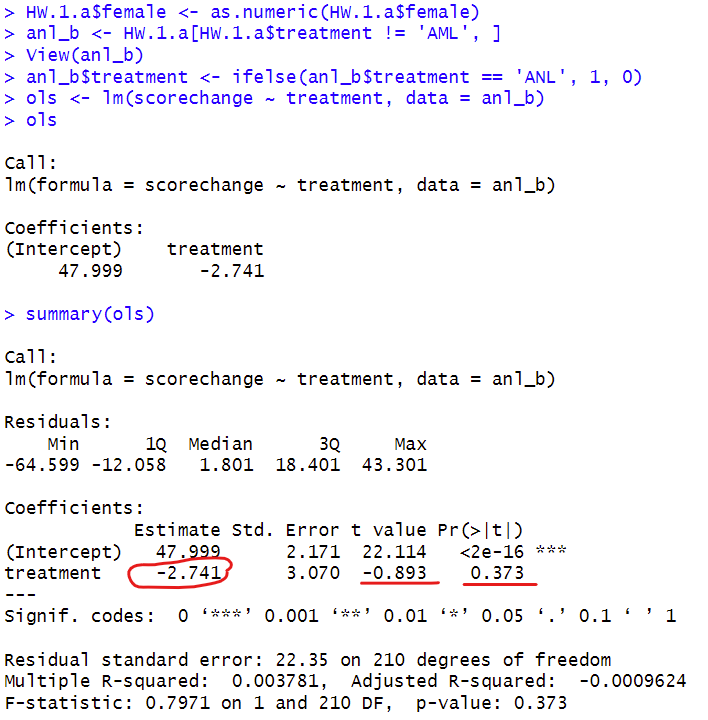
1. There are 320 students included in this dataset: 108 students with “AML” treatment, 106 students with “ANL” treatment, and 106 students with “B” treatment.



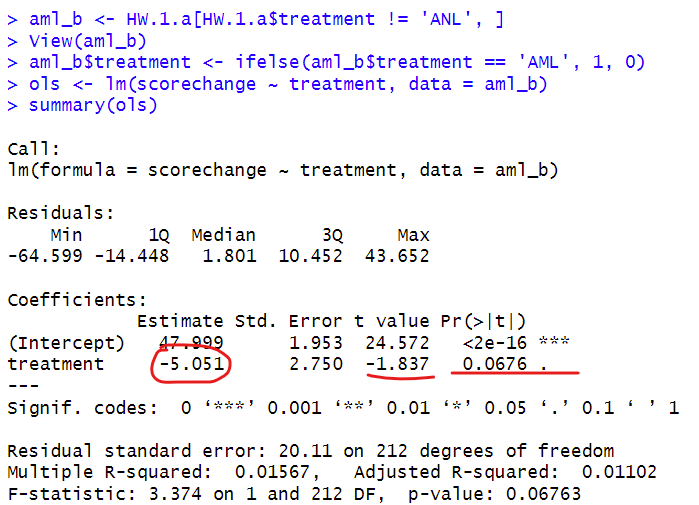
1. Descriptive statistics for the change in scores for all students and student groups based on treatment:



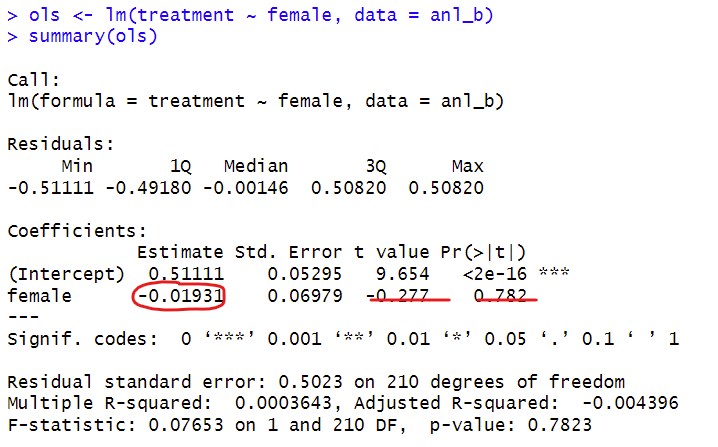
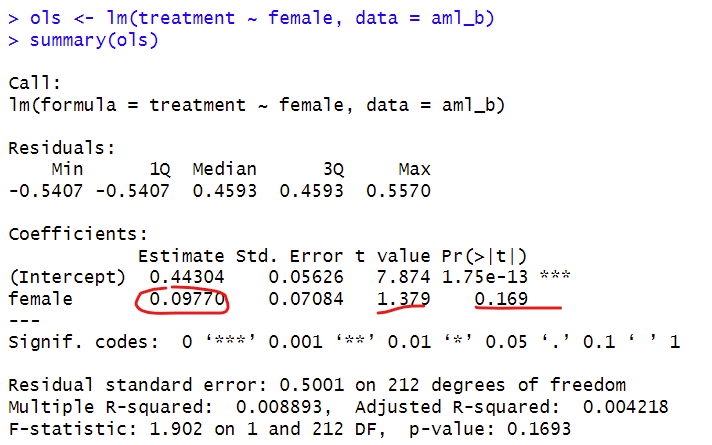
c) Yes, the average change in scores is smaller for students who were subject to ANL condition (improved by 45.3 points) compared to students with B condition (improved by 48 points). However, the relationship is not statistically significant, as shown below.

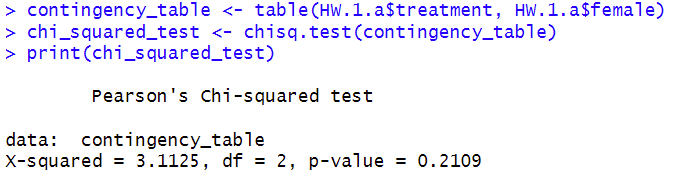
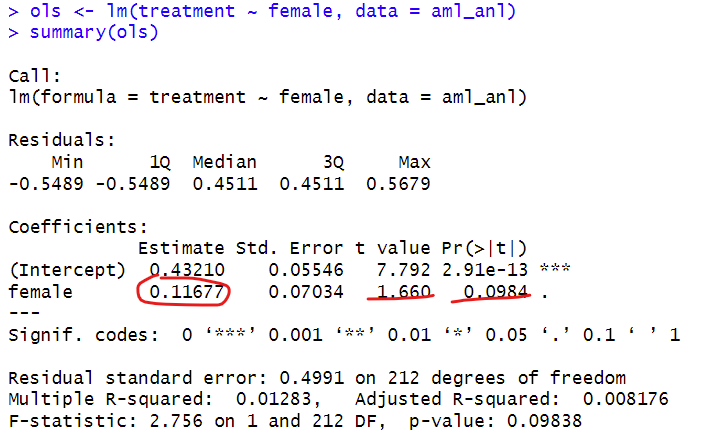


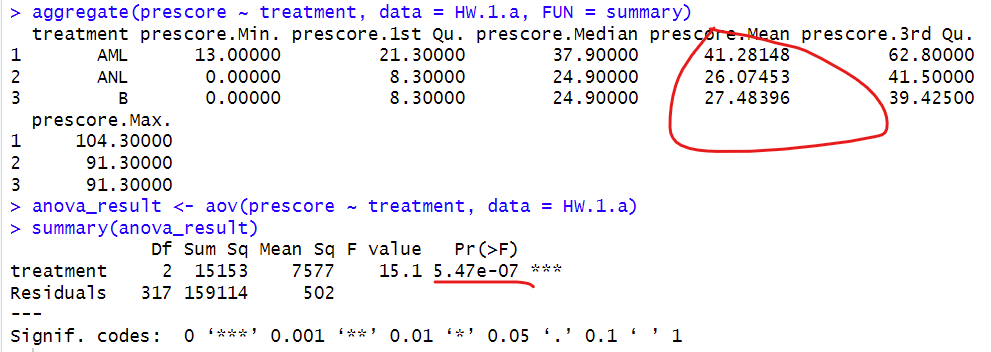
d) Yes, the average change in scores is smaller for students who were subject to AML condition (improved by 42.9 points) compared to students with B condition (improved by 48 points). The average improvement with condition AML is even smaller than for students with condition ANL. However, the difference is not statistically significant, as shown below.



e) No, we cannot consider the relationships found in c) and d) as causal. There is no information on the randomness in the assignment of conditions, so selection bias might have affected the results. Chi-Squared Test for Independence illustrates that there is insufficient evidence to conclude a significant association between treatment and gender in the observed data. However, there were statistically significant differences in prescores between the treatment groups.

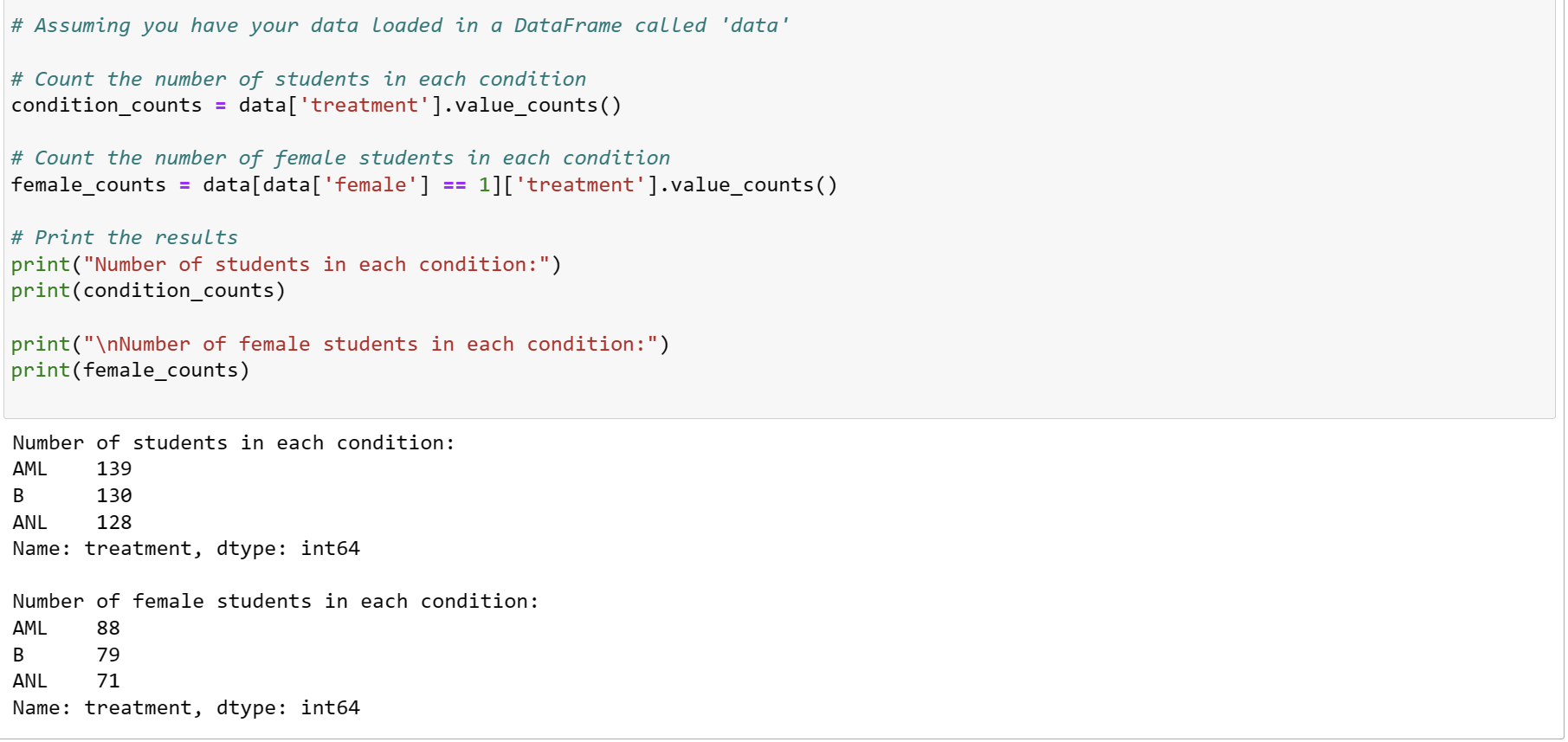






**Question 3**

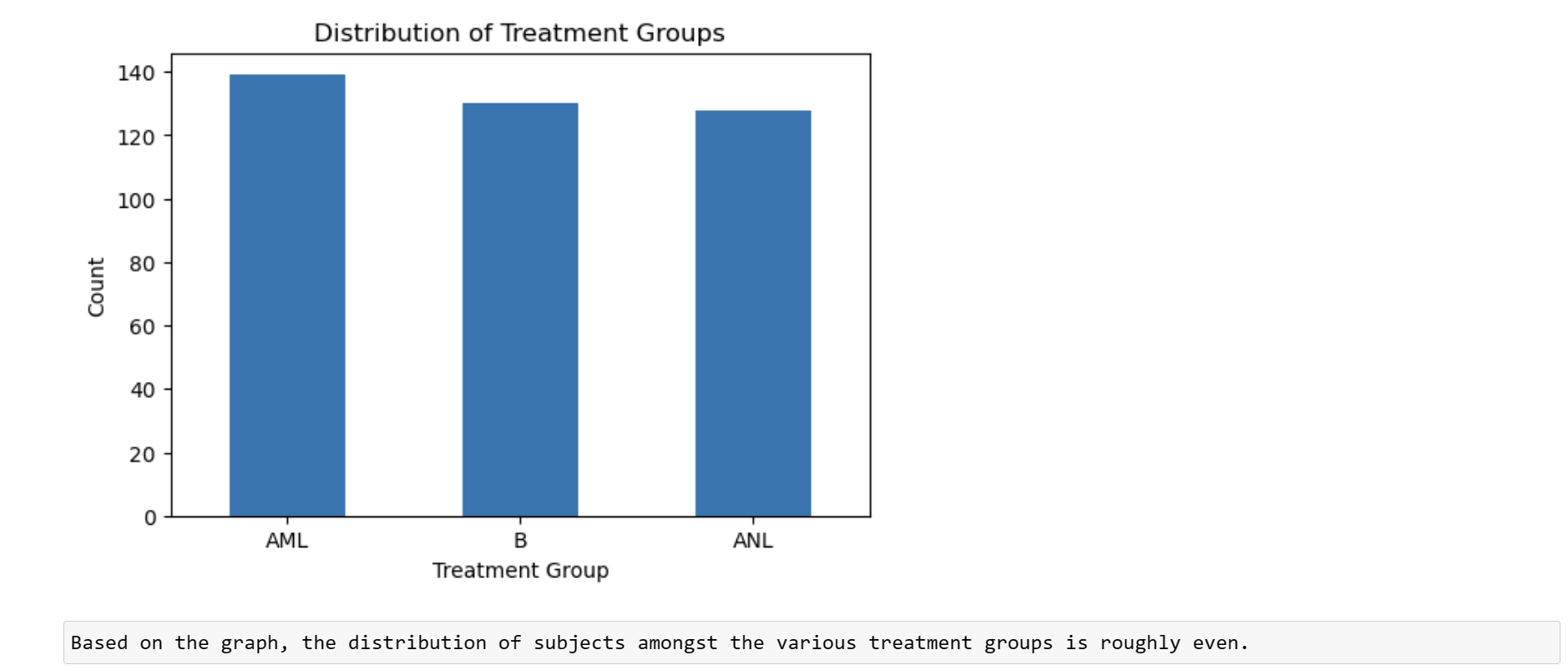
1. How many students participated in the experiment in each condition? How many of them are female?

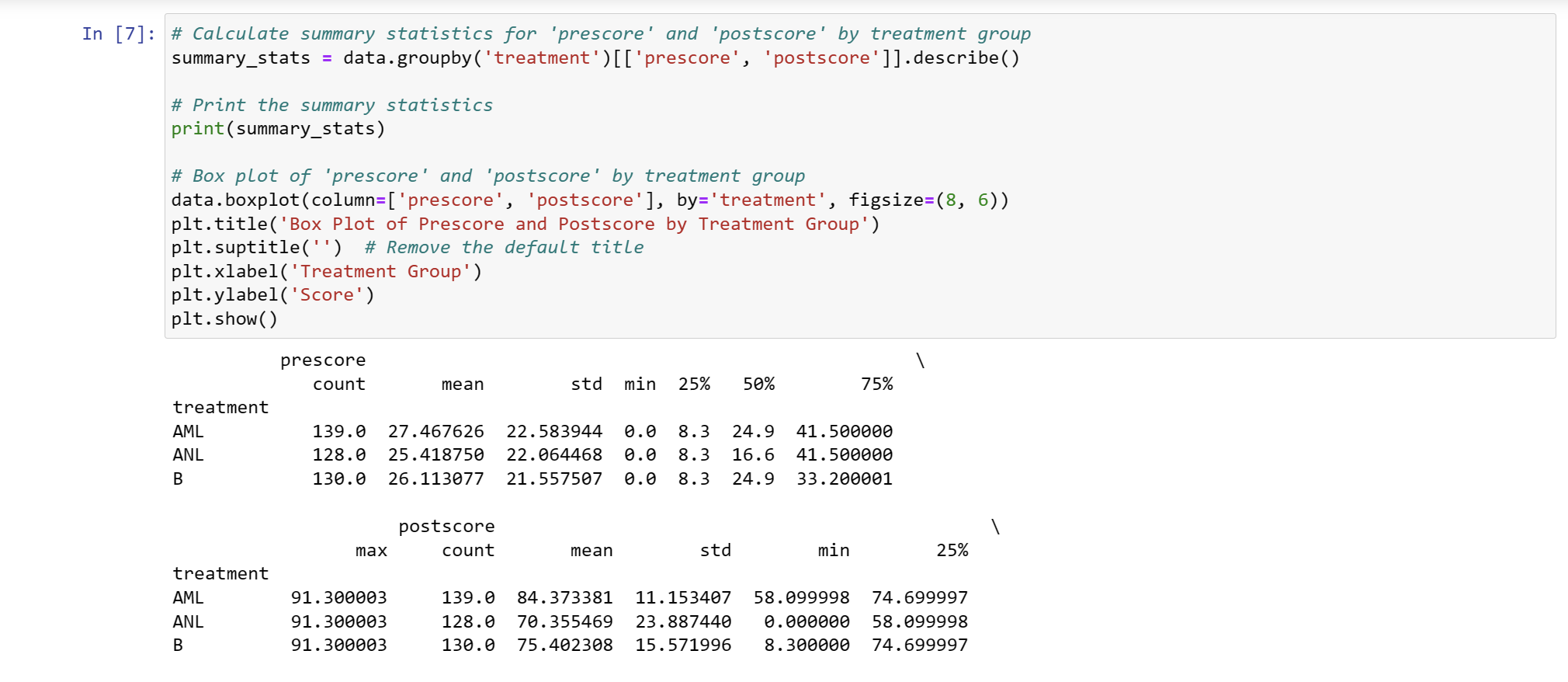


1. Show suggestive evidence that this data came from a well executed RCT

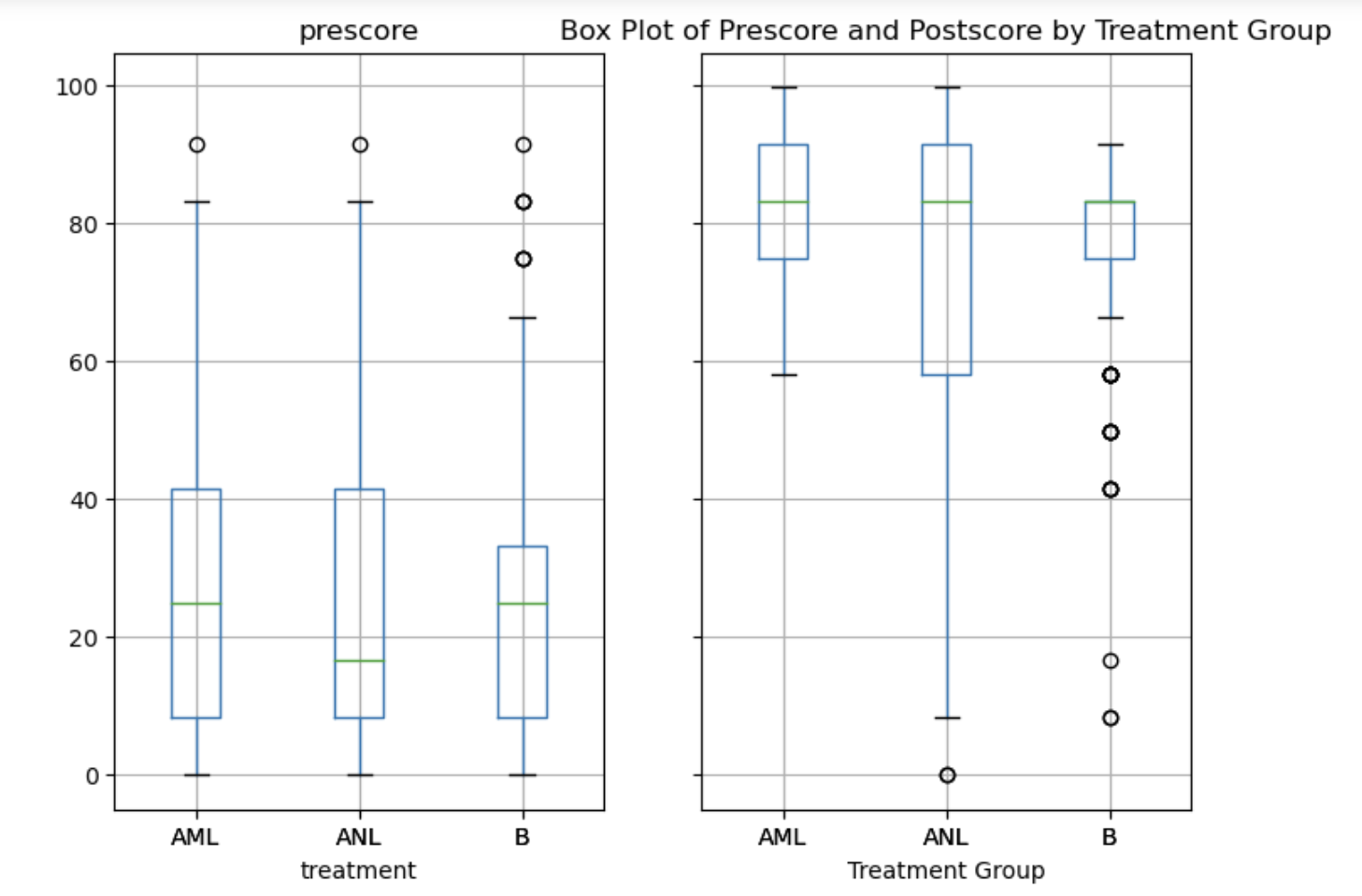
In a well-executed RCT, you are expected to see a relatively even distribution of subjects across treatment groups, similar summary statistics for baseline measures, and no significant differences in baseline covariates between groups. Any deviations from these expectations may raise questions about the randomization process or the quality of the trial.

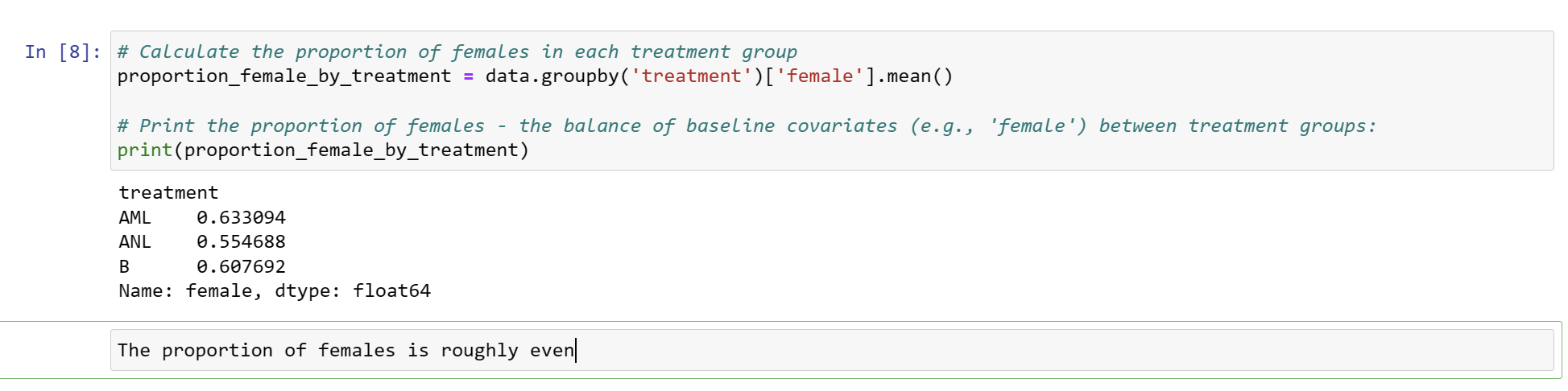


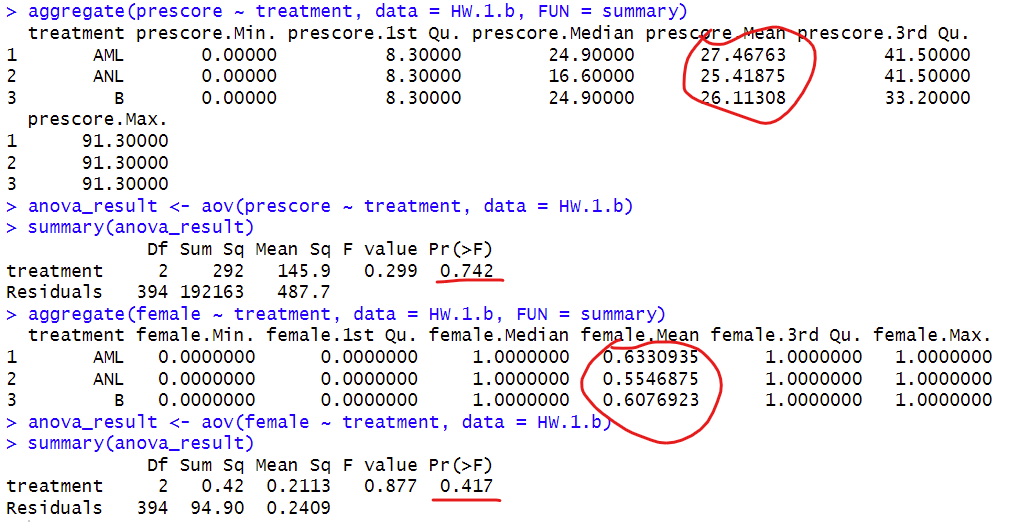






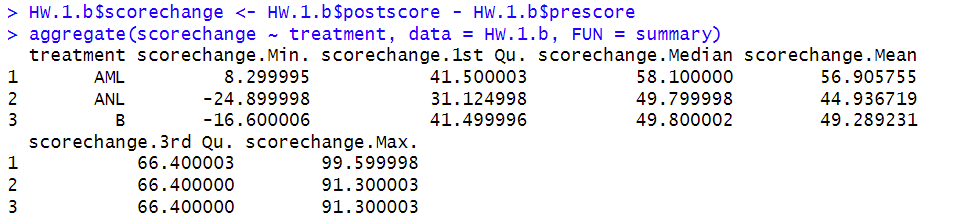




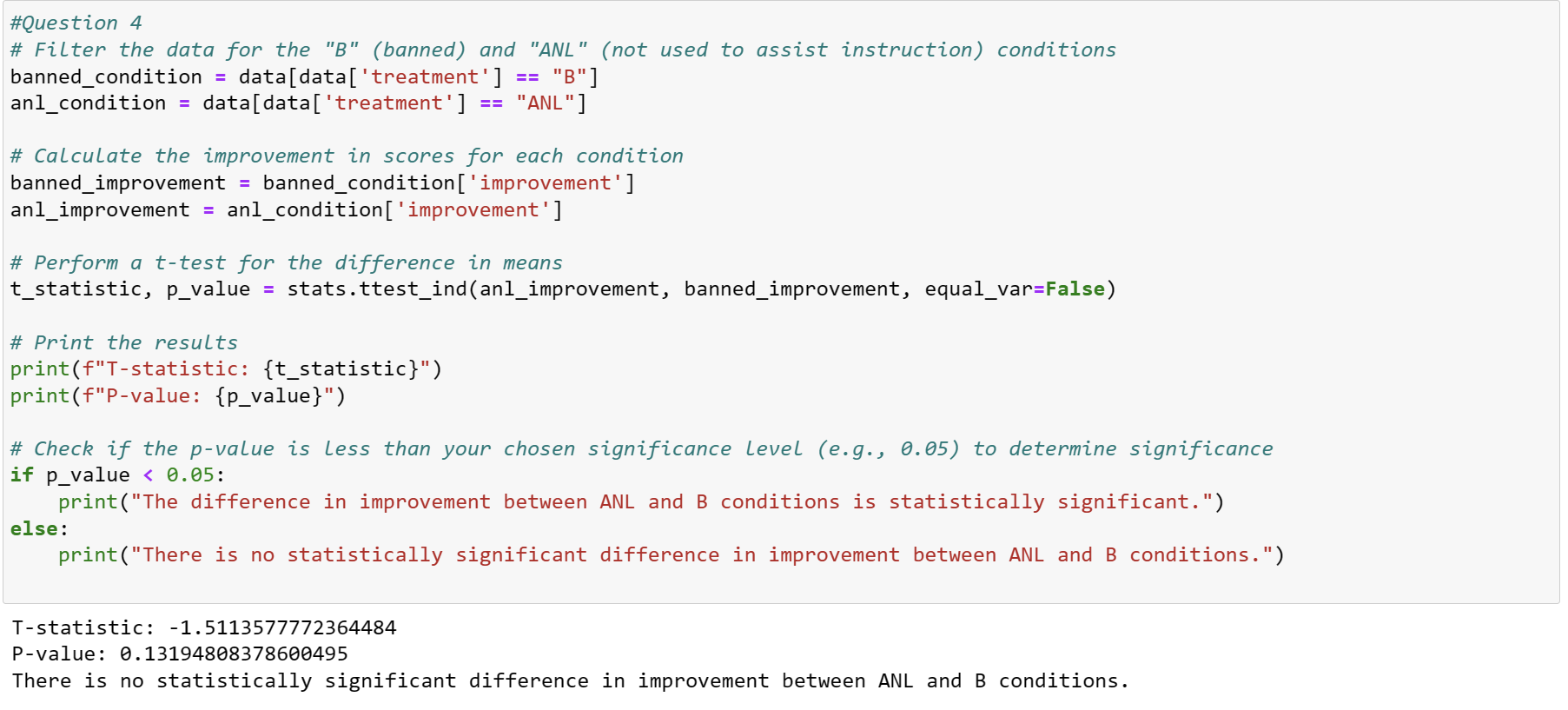


1. Show descriptive statistics for the improvement in test scores from the pre-test to the post-test

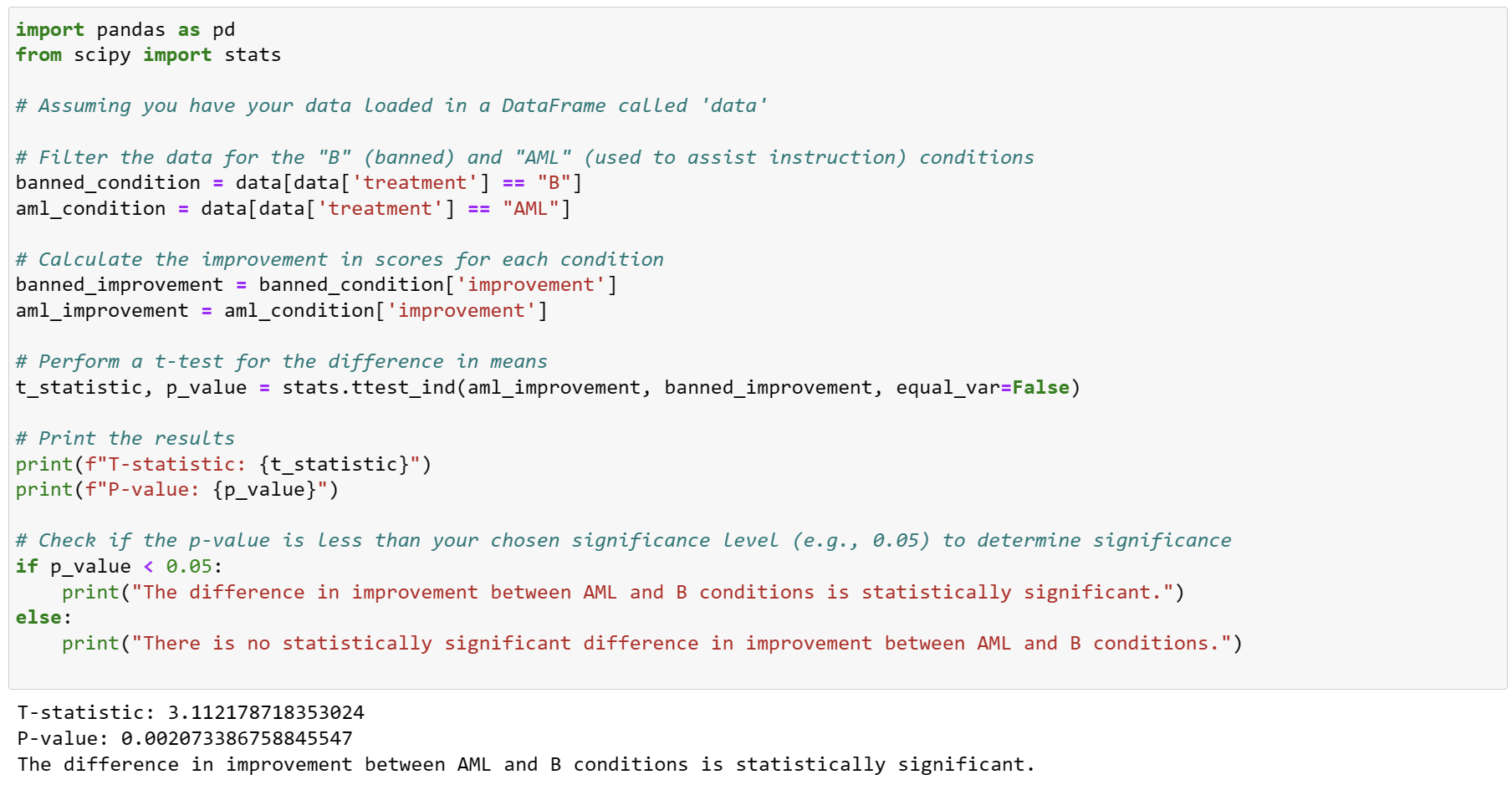




d) Show whether the improvement in the scores (as defined above) reduced when smartphones were allowed into the classroom and not used to assist instruction (compared to when they were banned from the classroom)?

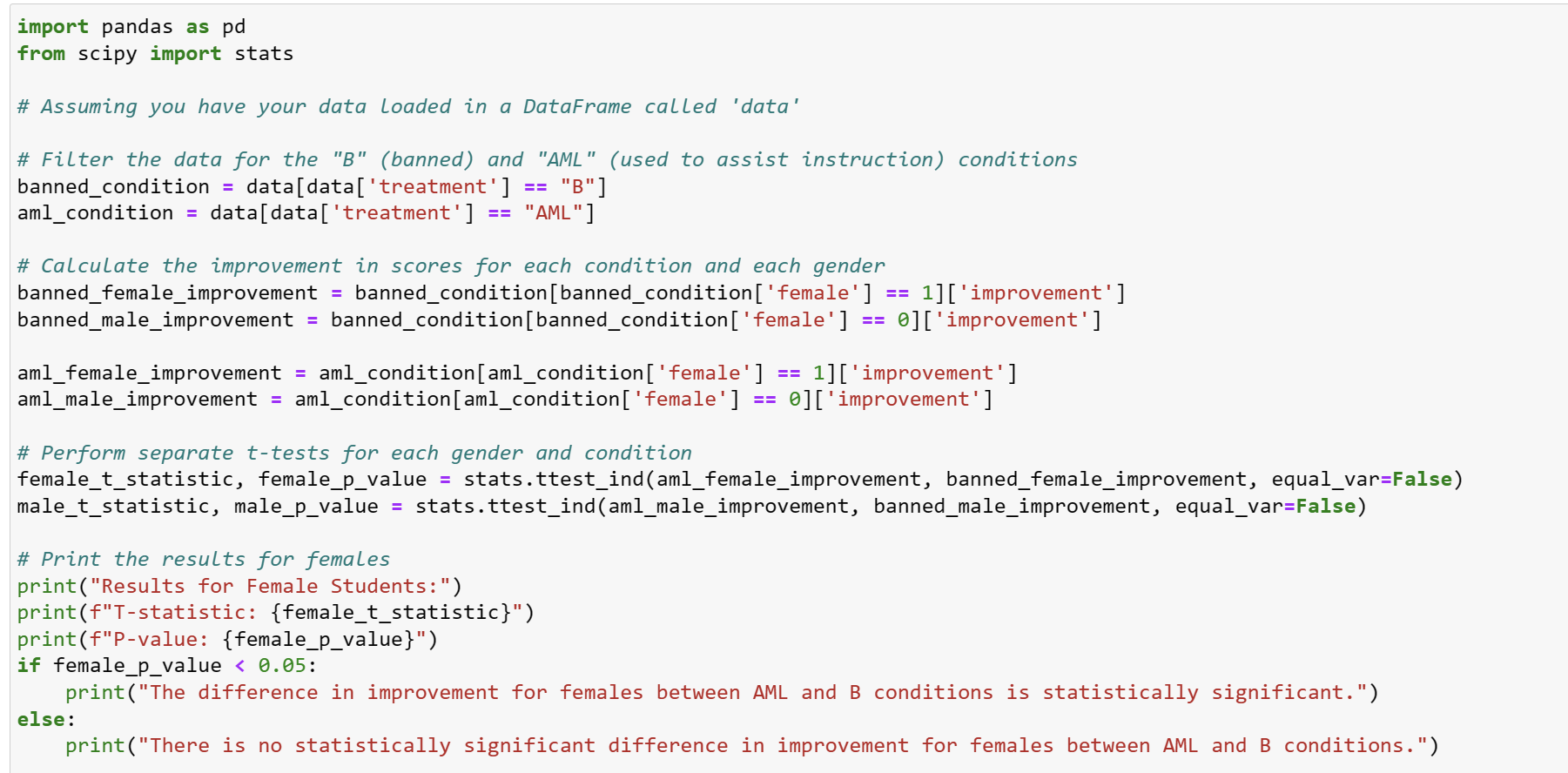


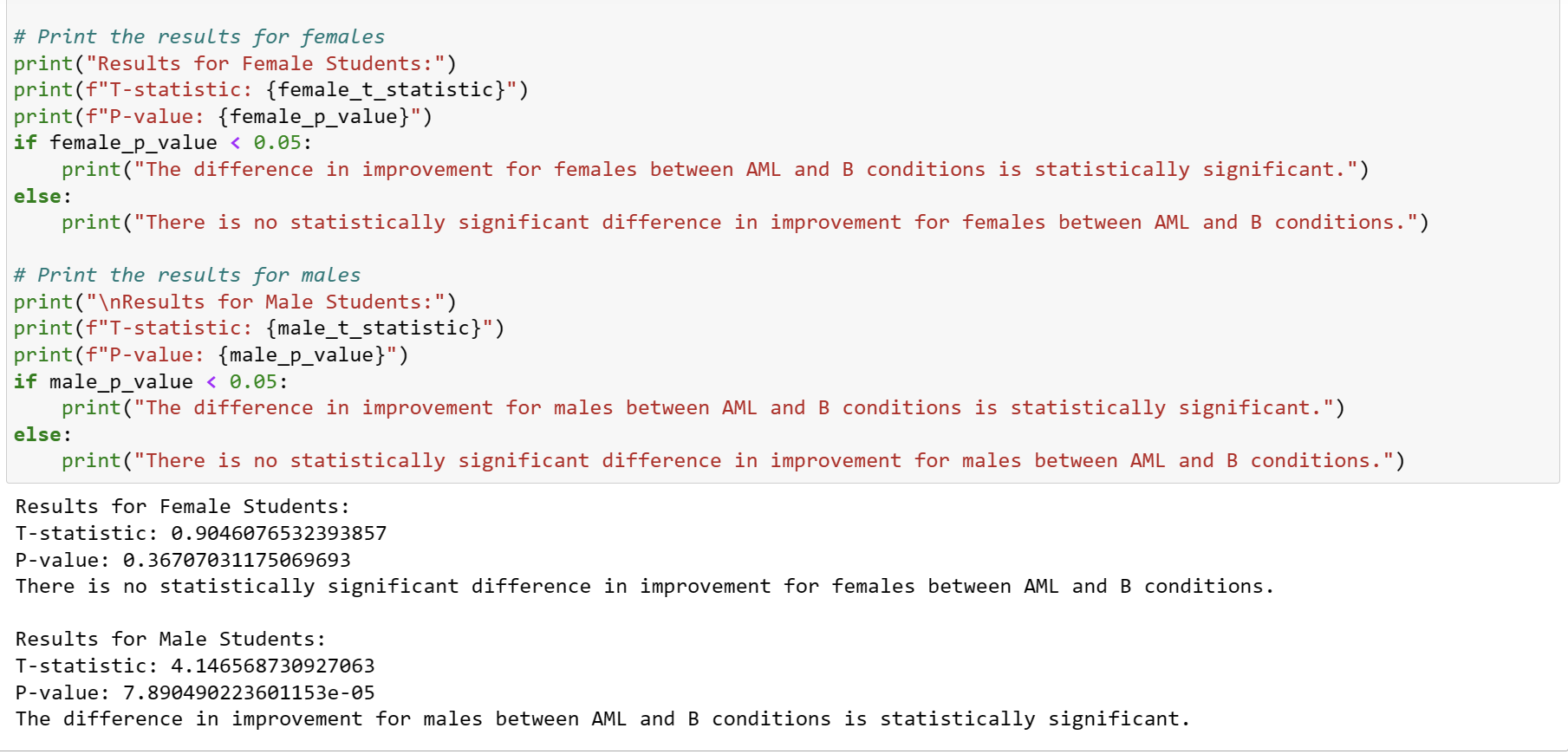
The difference in improvement scores between ANL and B conditions is not statistically significant and hence, we cannot confirm that the improvement in scores reduced when smartphones were allowed into the classroom and not used to assist instruction.

e) Show whether the improvement in the scores (as defined above) increased when smartphones were allowed into the classroom and used to assist instruction (compared to when they were banned from the classroom)?

The difference in improvement scores between AML and B conditions is statistically significant and hence, the improvement in scores increased when smartphones were allowed into the classroom and used to assist instruction.

f) Show whether the results obtained in d) and e) above are different for female and male students





The improvement in scores between AML and B conditions is statistically significant for males, but it is not statistically significant for females

References:

* Lecture notes
* ChatGPT-4 for coding in R: Linear regression, Chi-Squared Test for Independence
* ChatGPT-4 for helping with code on t-tests in python