Sean McLean

ALY 6015

Module 5 Assignment

Nonparametric Methods and Sampling



**Introduction**

The focus is to answer a set of selected questions with different types of nonparametric statistical methods. Each question starts with creating null and alternate hypotheses and identifying the claim. The critical value and test values are then computed using a nonparametric method and with the necessary codes from the weekly module references. The results of these tests will dictate whether the null hypothesis will be rejected or not.

**Analysis**

**Section 13-2**

**Question 6: Game attendance**

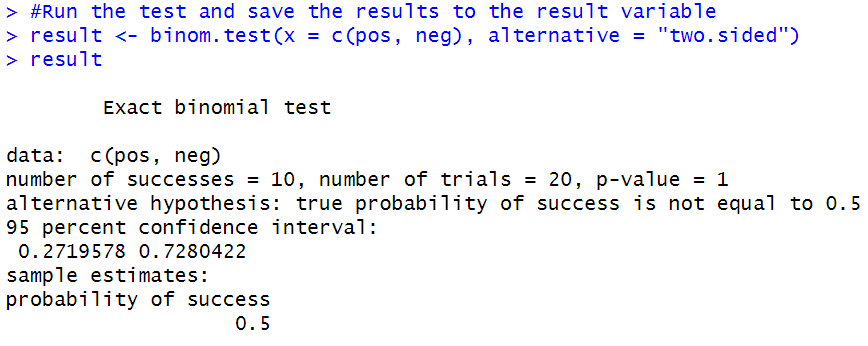
Hypotheses and claim:

* H0: median = 3000 (claim)
* H1: median !=3000

Critical value

* a=0.05, n=20, two-tailed test, critical value = 5

Test value



The p-value of 1.0 is higher than 0.05 so we will fail to reject the null hypothesis. There is not enough evidence to reject the claim that the paid attendance at 20 local football games is 3000. This could be a situation where correlation does not imply causation regarding the relationship between the factors. While the p-value is high (1.0), most values from the sample of games are close to the median of 3000. I would be comfortable then with printing this number on the programs for the games because of the close figures in the sample.

**Question 10: Lottery ticket sales**

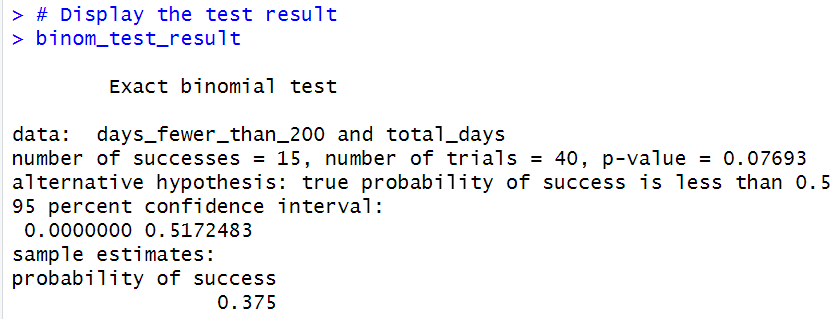
Hypotheses and claim

* H0: median = 200
* H1: median < 200 (claim)

Critical value

* n = 200 and alpha = 0.05, critical value = 88

Test value



The p-value of 0.07693 is higher than 0.05 so we will fail to reject the null hypothesis. There is not sufficient evidence to conclude that the median is below 200 lottery tickets.

**Section 13-3**

**Question 4: Lengths of Prison Sentences**

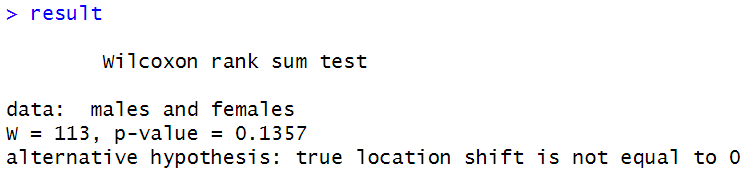
Hypotheses and claim

* H0: There is no difference in the sentence received by each gender. (claim)
* H1: There is a difference in the sentence received by each gender.

Critical value

* Df: 24, critical value = -2.063899, 2.063899

Test value



The p-value of 0.1357 is higher than 0.05 so we will fail to reject the null hypothesis. There is not enough evidence to support the claim that there is no difference in the sentences received by each gender at the prison.

**Question 8: Winning Baseball Games**

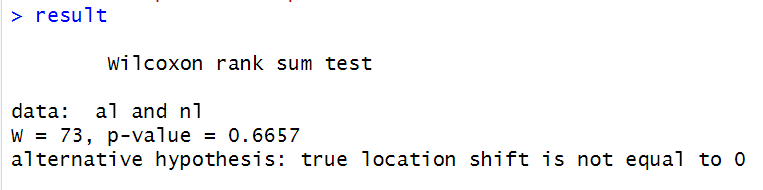
State hypotheses and claim

* H0: There is no difference in the number of wins between the leagues.
* H1: There is a difference in the number of wins between the leagues. (claim)

Critical value

* Df = 21, critical value = -2.063899, 2.063899

Test value



The p-value of 0.6657 is higher than 0.05 so we will fail to reject the null hypothesis. There is not enough evidence to support the claim that there is a difference in the amount of wins between the two leagues over that time frame.

**Section 13-4**

* Table K from textbook used to determine whether the null hypothesis should be rejected (Bluman, 2018).

**Question 5: ws = 13, n = 15, α = 0.01, two-tailed**

Critical value: 16

Reject the null hypothesis because the test value is less than or equal to the critical value.

**Question 6: ws = 32, n = 28, α = 0.025, one-tailed**

Critical value: 117

Reject the null hypothesis because the test value is less than or equal to the critical value.

**Question 7: ws = 65, n = 20, α = 0.05, one-tailed**

Critical value: 60

Fail to reject the null hypothesis because the test value is larger than or equal to the critical value.

**Question 8: ws = 22, n = 14, α = 0.10, two-tailed**

Critical value: 26

Reject the null hypothesis because the test value is less than or equal to the critical value.

**Section 13-5**

**Question 2: Mathematics Literacy Scores**

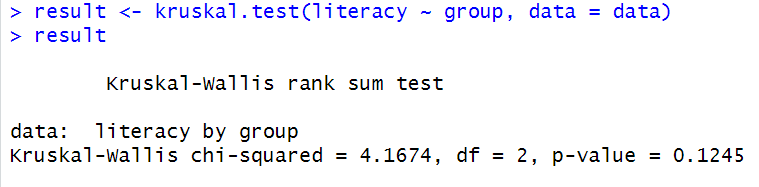
Hypotheses and claim

* H0: There is no difference in means of literacy scores between the three regions.
* H1: There is a difference in means of literacy scores between the three regions. (claim)

Critical value

* df: 3-1=2, a = 0.05, critical value = 5.991

Test value



The p-value of 0.1245 is higher than 0.05 so we will fail to reject the null hypothesis. There is not enough evidence to support the claim that there is a difference in the averages of mathematics literacy scores between the three regions.

**Section 13-6**

**Question 6: Subway and Commuter Rail Passengers**

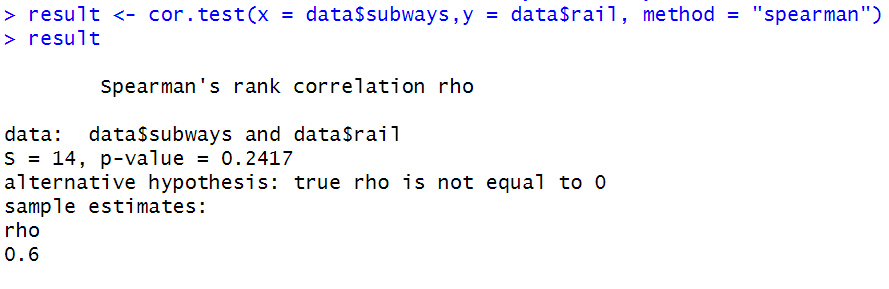
Hypotheses and claim

* H0: There is no correlation between the number of daily passenger trips for subways and commuter rail service in the six cities.
* H1: There is a significant correlation between the number of daily passenger trips for subways and commuter rail service in the six cities.

Critical value

* n=6, a = 0.05, critical value = 0.886

Test value

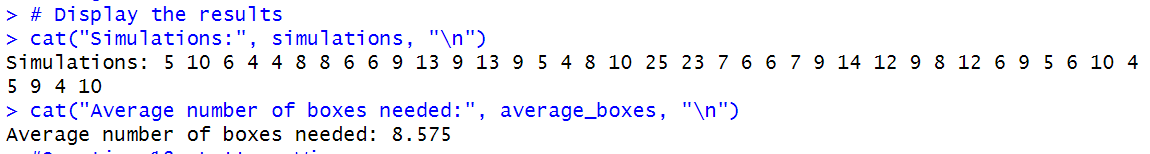


The p-value of 0.2417 is higher than 0.05 so we will fail to reject the null hypothesis. There is not enough evidence to support the claim that there is a significant correlation between the number of daily passenger trips for subways and commuter rail service in the six cities. This could be useful information for a transportation authority if there is a positive correlation between factors, potentially indicating that there is high demand for this type of transportation. It could lead to future investments and other resources being put into these modes of transportation if there is a high number of daily passenger trips.

**Section 14-3**

**Question 16: Prizes in Caramel Corn Boxes**

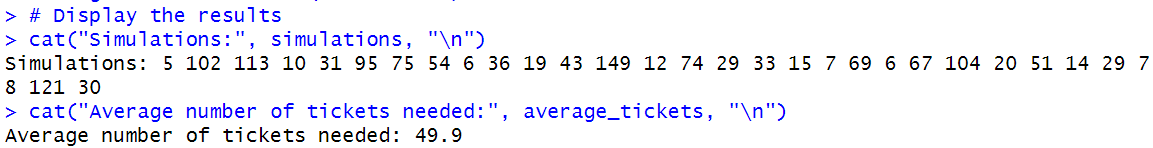
* Number of simulations: 40



The following R code simulates the process 40 times, calculates the number of boxes needed in each simulation, and then calculates the average. This will simulate the experiment and find the average number of boxes a person needs to buy to get all four prizes. From 40 simulations, the average number of boxes needed is 8.575.

**Question 18: Lottery Winner**

* Number of simulations: 30



The following R code simulates the process 30 times, calculates the number of tickets needed in each simulation, and then calculates the average. This will simulate the experiment and find the average number of tickets a person must buy to win the prize, considering the probabilities of obtaining each letter. From 30 simulations, the average number of tickets needed is 49.9.

**Conclusion**

The assignment questions used several nonparametric statistical methods, including the sign test, Wilcoxon rank sum test, signed-rank test, Kruskal-Wallis test, and the runs test. Each test that featured hypothesis testing had results that failed to reject the null hypothesis due to high p-values. This also was a common theme in signed-rank tests, as only one question had a result where the null hypothesis was rejected. The last two questions that used runs tests calculated an average value based on several simulations of the experiments. The module 5 assignment overall was a valuable opportunity to apply nonparametric statistical methods to several examples and practice more hypothesis testing.

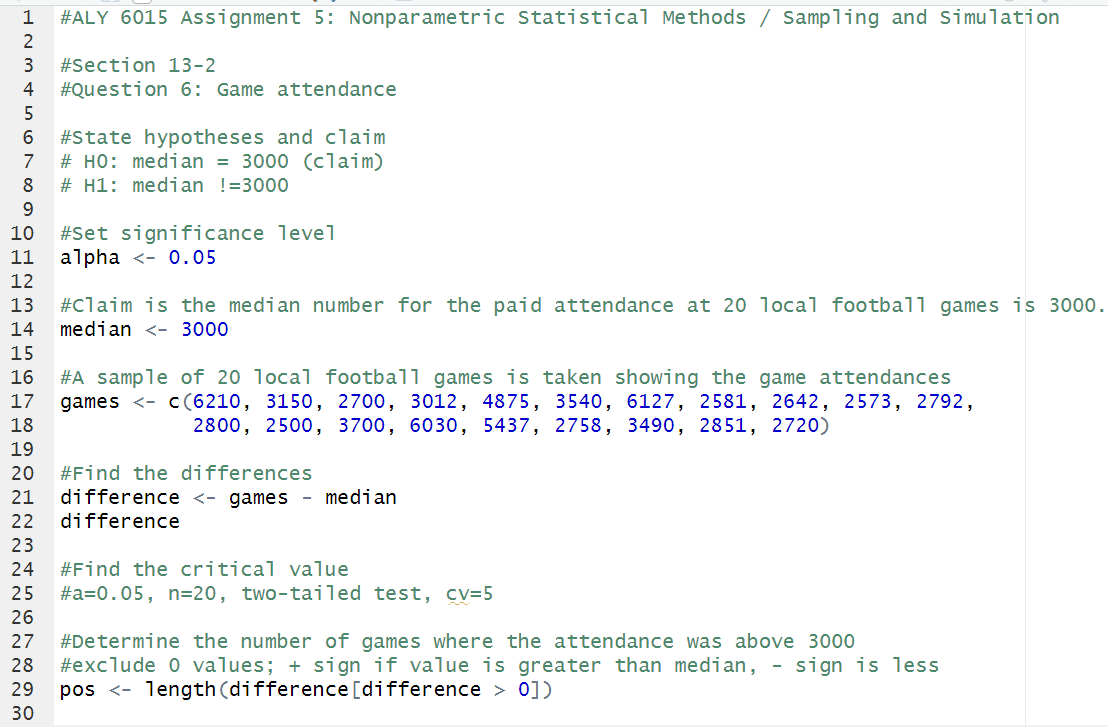
**References**

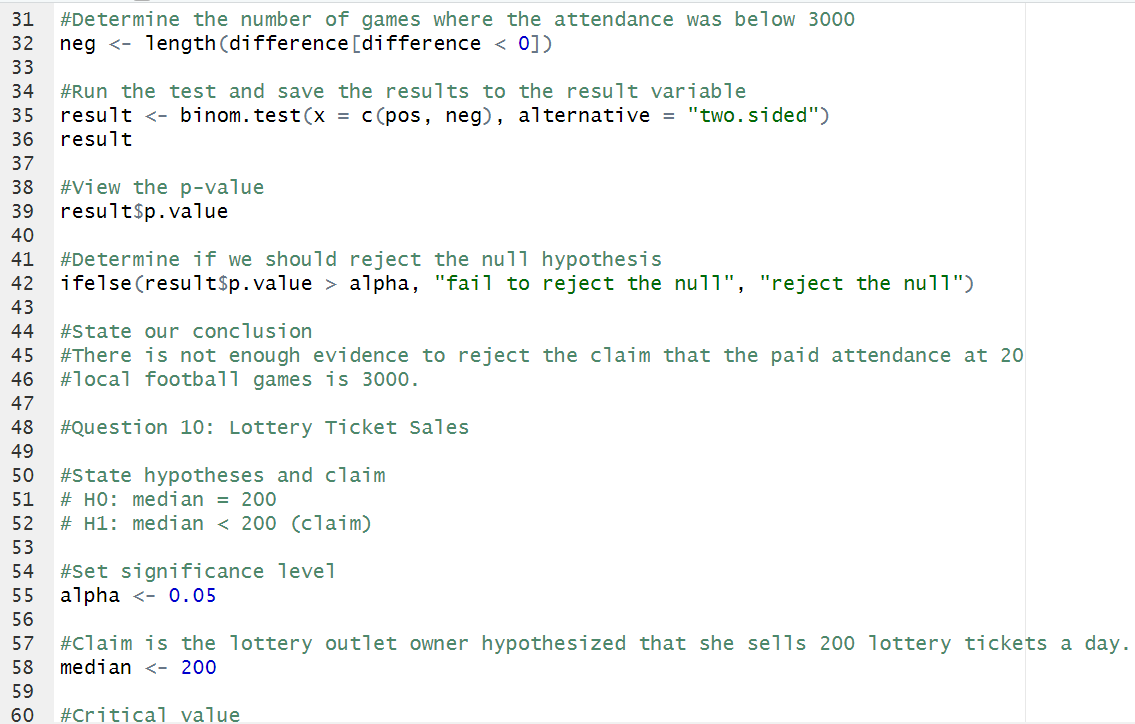
Bluman, A. (2018). Elementary statistics: Astep by step approach (10th ed.). McGraw Hill.

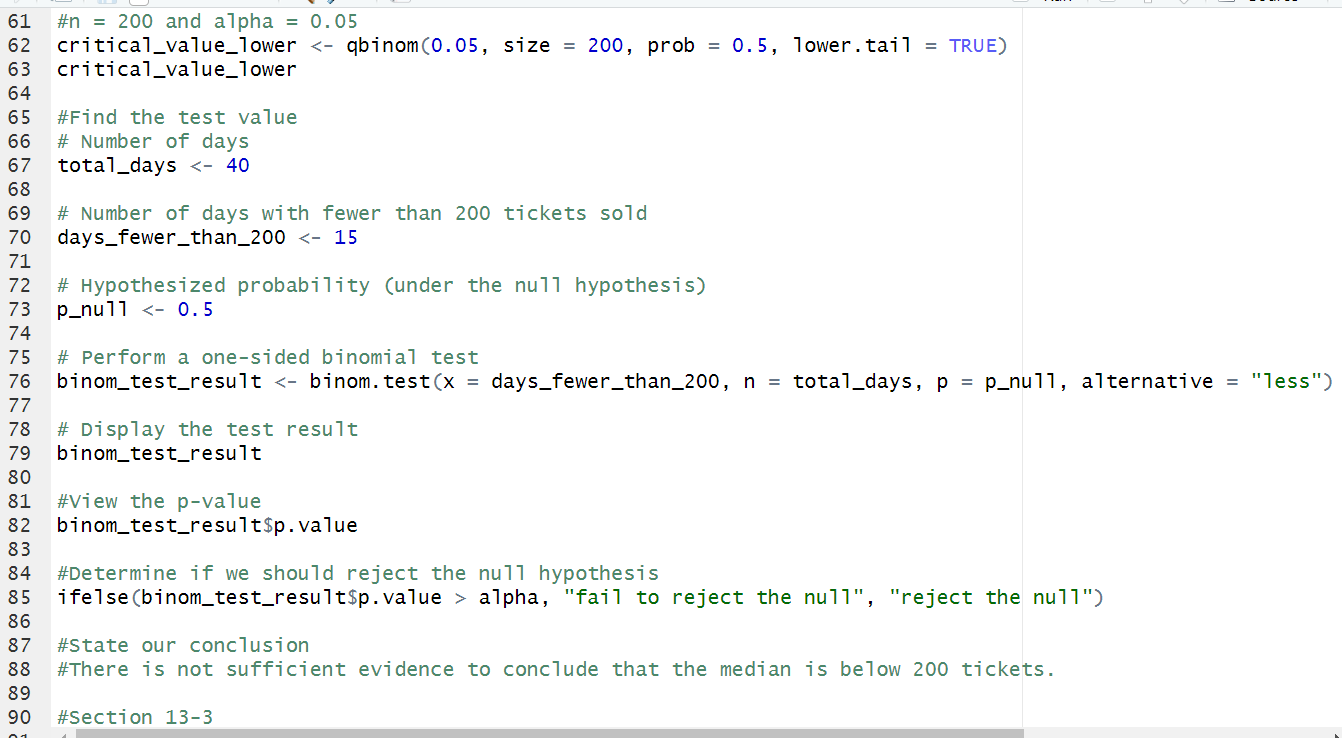
Chat GPT. (2023, December 10th). Default (GPT 3.5). < https://chat.openai.com/>

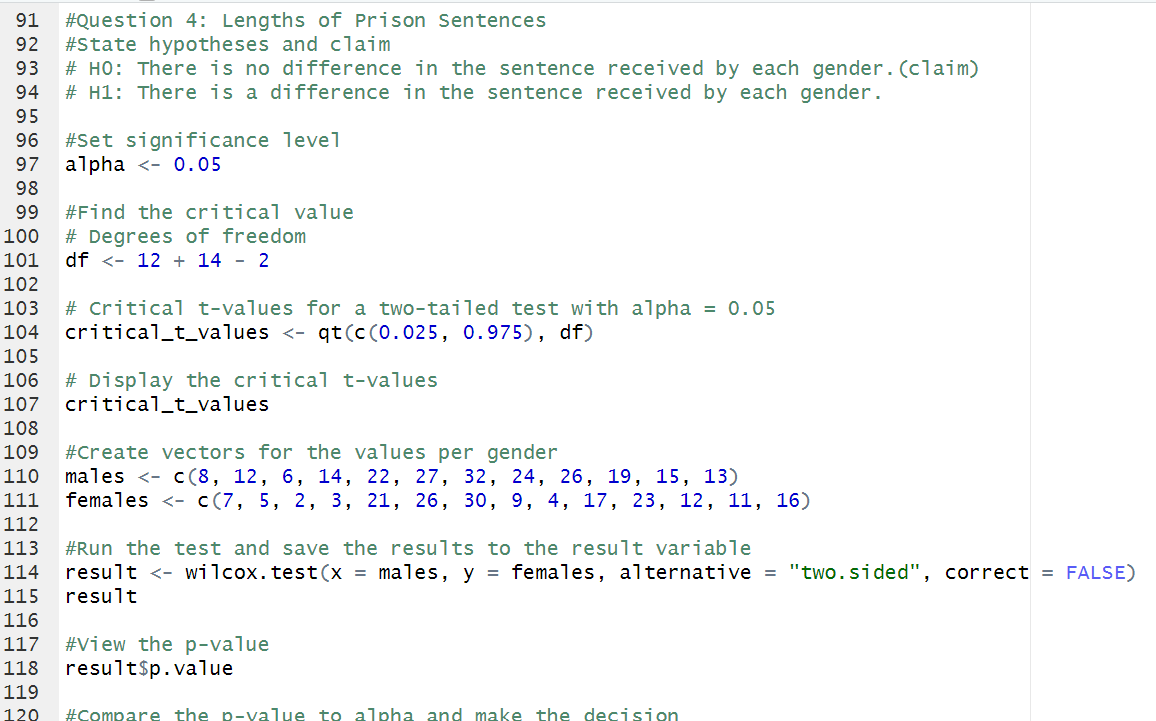
Kabacoff, R. I. (2022). R in action: Data analysis and graphics with R and tidyverse (3rd ed.). Manning Publications.

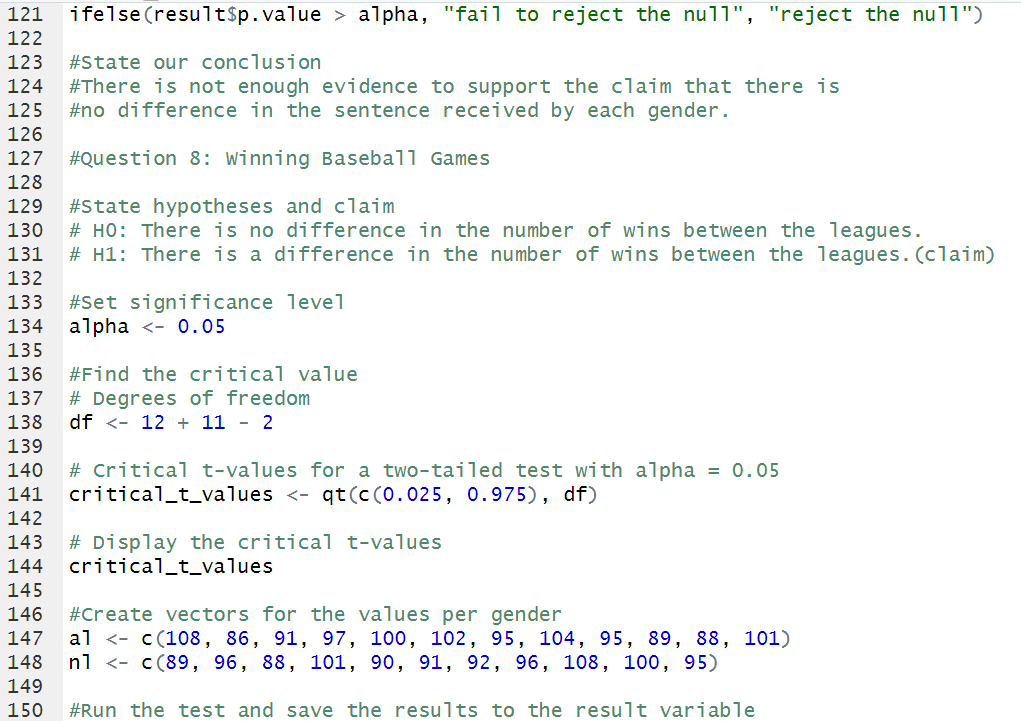
**Appendix**

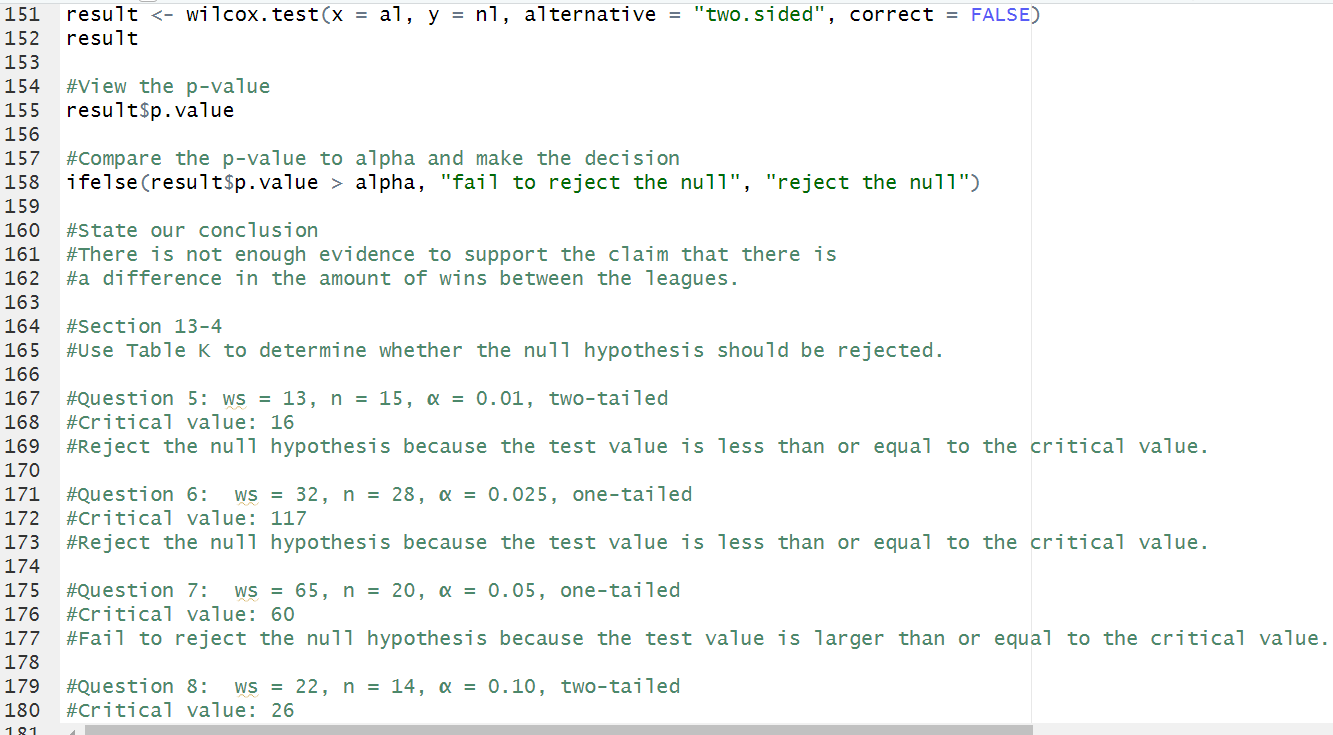


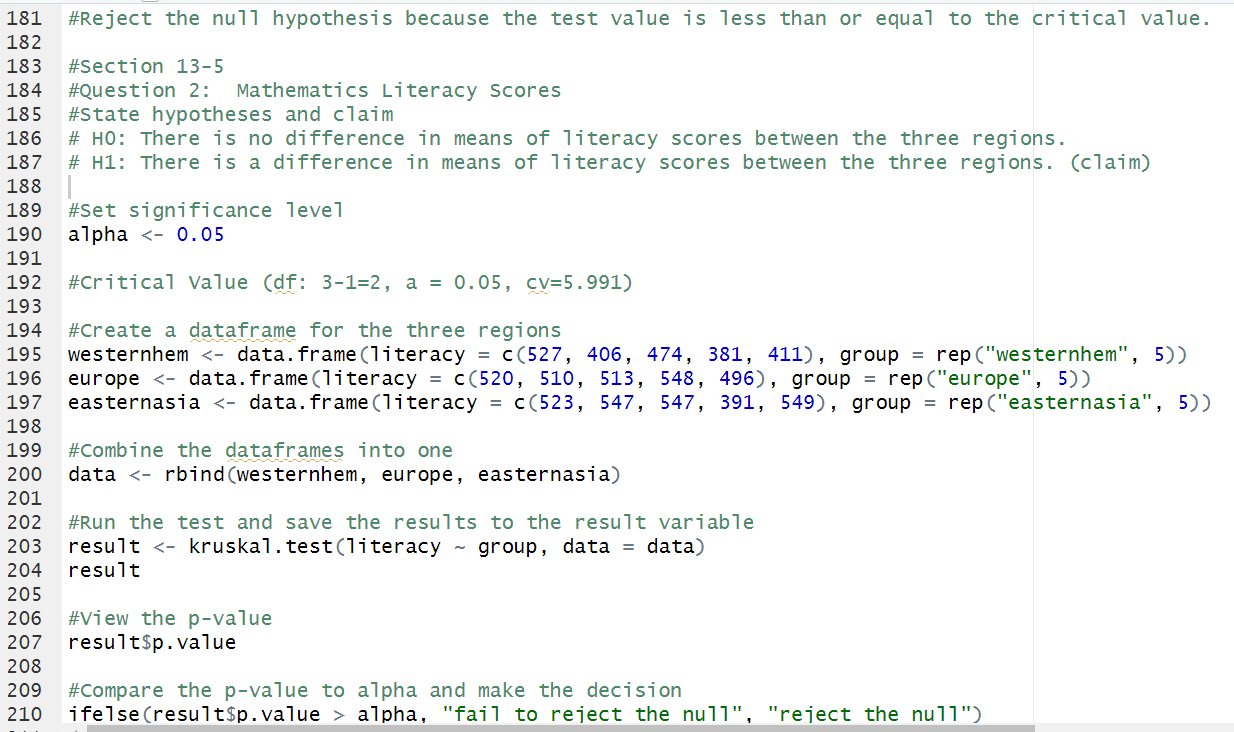


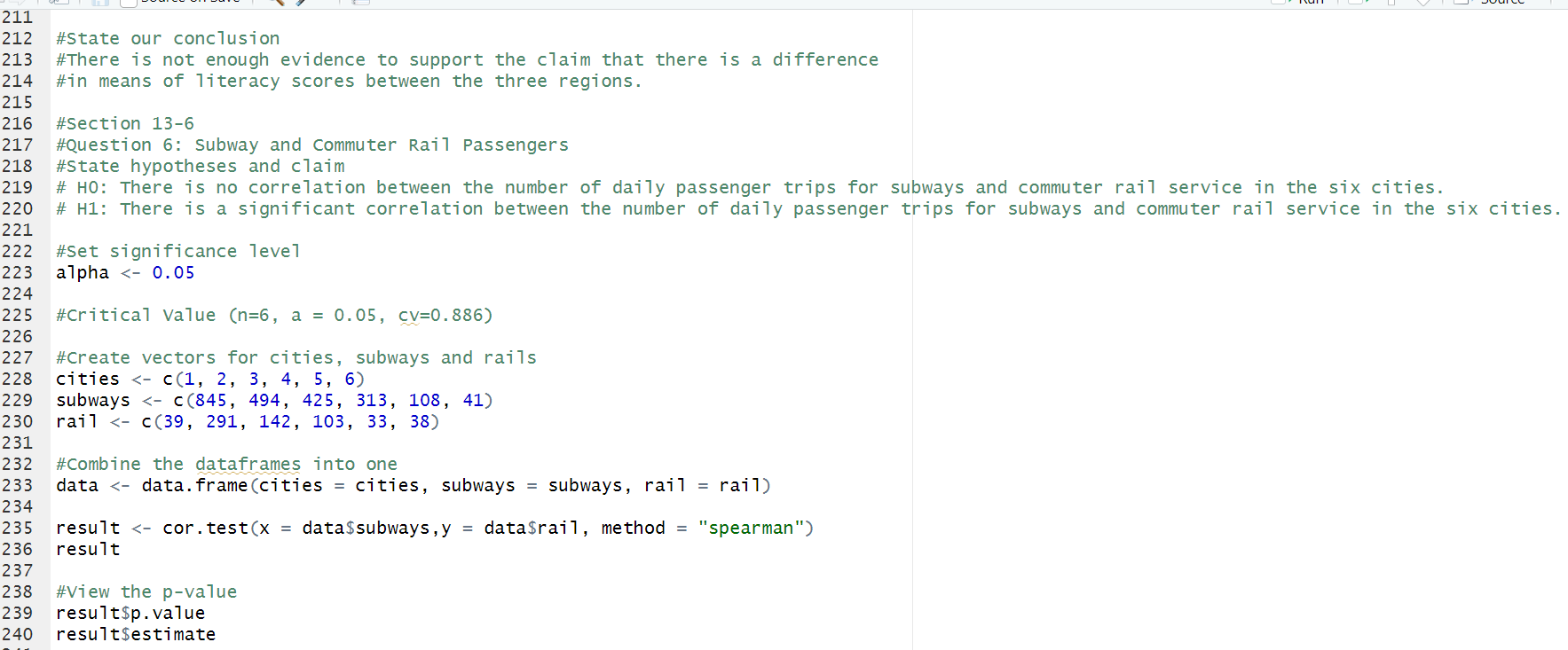


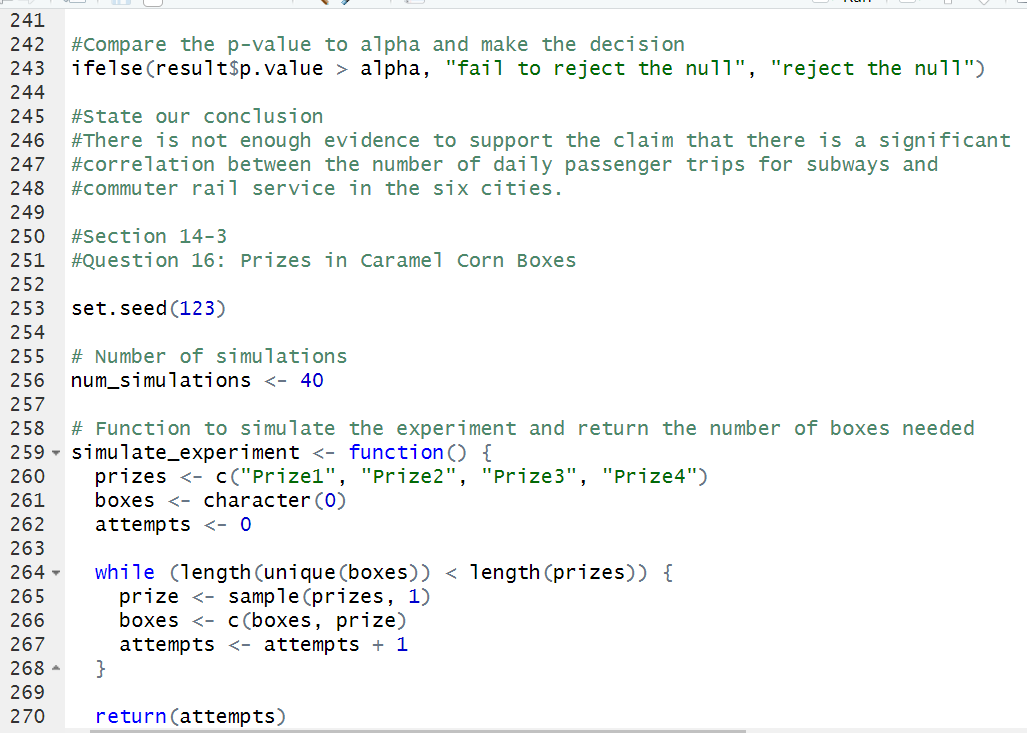


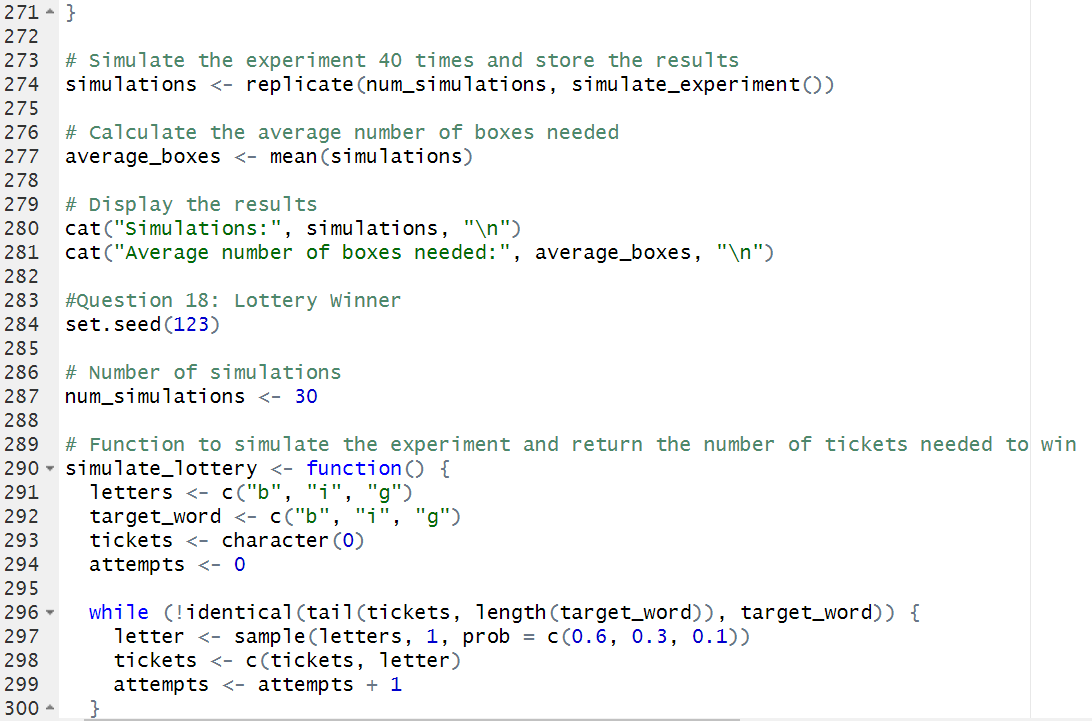


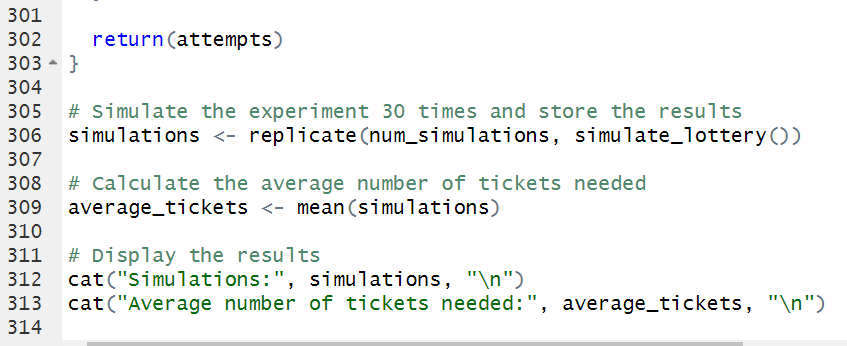












Nonparametric Methods and Sampling