Northeastern University
College of Professional Studies

Nonparametric Statistical Methods / Sampling and Simulation

Overview and Rationale

In this assignment, you will use your knowledge of nonparametric statistical methods and sampling to solve various types of problems.

Course Outcomes

This assignment is directly linked to the following key learning outcomes from the course syllabus:

- State the advantages and disadvantages of nonparametric methods.
- Test hypotheses, using the following methods: sign test, Wilcoxon rank sum test, signed-rank test, Kruskal-Wallis test, and the runs test.
- Compute the Spearman rank correlation coefficient.
- Demonstrate a knowledge of the four basic sampling methods.
- Recognize faulty questions on a survey and other factors that can bias responses.
- Solve problems, using simulation techniques.

Assignment Summary

Complete the following problems using R and/or MS Excel. Be sure to show your work and include the hypothesis tests, the critical values, the computed test values, and the resulting decisions where applicable.

Section 13-2

Perform these steps:
- a. State the hypotheses and identify the claim.
- b. Find the critical value(s).
- c. Compute the test value.
- d. Make the decision.
- e. Summarize the results.

Use the traditional method of hypothesis testing unless otherwise specified. Assume all assumptions are met.

6. **Game Attendance** An athletic director suggests the median number for the paid attendance at 20 local football games is 3000. The data for a random sample are shown. At $\alpha = 0.05$, is there enough evidence to reject the claim? If you were printing the programs for the games, would you use this figure as a guide?

6210	3150	2700	3012	4875
3540	6127	2581	2642	2573
2792	2800	2500	3700	6030
5437	2758	3490	2851	2720

Source: Pittsburgh Post Gazette.

10. **Lottery Ticket Sales** A lottery outlet owner hypothesizes that she sells 200 lottery tickets a day. She randomly sampled 40 days and found that on 15 days she sold fewer than 200 tickets. At $\alpha =$ 0.05, is there sufficient evidence to conclude that the median is below 200 tickets?

Section 13-3
Use the Wilcoxon rank sum test. Assume that the samples are independent. Also perform each of these steps.

   a. *State the hypotheses and identify the claim.*
   b. *Find the critical value.*
   c. *Compute the test value.*
   d. *Make the decision.*
   e. *Summarize the results.*

Use the traditional method of hypothesis testing unless otherwise specified. Assume all assumptions are valid.

4. **Lengths of Prison Sentences** A random sample of men and women in prison was asked to give the length of sentence each received for a certain type of crime. At $\alpha = 0.05$, test the claim that there is no difference in the sentence received by each gender. The data (in months) are shown here.

Males	8 12 6 14 22 27 32 24 26
Females	7 5 2 3 21 26 30 9 4
Males	19 15 13
Females	17 23 12 11 16

Source: Based on information from the National Highway Traffic Safety Administration.

8. **Winning Baseball Games** For the years 1970–1993 the National League (NL) and the American League (AL) (major league baseball) were each divided into two divisions: East and West. Below are random samples of the number of games won by each league's Eastern Division. At $\alpha = 0.05$, is there sufficient evidence to conclude a difference in the number of wins?

NL	89 96 88 101 90 91 92 96 108 100 95
AL	108 86 91 97 100 102 95 104 95 89 88 101

Source: World Almanac.

Section 13-4
Use Table K to determine whether the null hypothesis should be rejected.

5. $w_s = 13$, $n = 15$, $\alpha = 0.01$, two-tailed

6. $w_s = 32$, $n = 28$, $\alpha = 0.025$, one-tailed

7. $w_s = 65$, $n = 20$, $\alpha = 0.05$, one-tailed

8. $w_s = 22$, $n = 14$, $\alpha = 0.10$, two-tailed

Section 13-5

Use the Kruskal-Wallis test and perform these steps.

  a. *State the hypotheses and identify the claim.*
  b. *Find the critical value.*
  c. *Compute the test value.*
  d. *Make the decision.*
  e. *Summarize the results.*

2. **Mathematics Literacy Scores** Through the Organization for Economic Cooperation and Development (OECD), 15-year-olds are tested in member countries in mathematics, reading, and science literacy. Listed are randomly selected total mathematics literacy scores (i.e., both genders) for selected countries in different parts of the world. Test, using the Kruskal-Wallis test, to see if there is a difference in means at $\alpha = 0.05$.

Western Hemisphere	Europe	Eastern Asia
527	520	523
406	510	547
474	513	547
381	548	391
411	496	549

Source: www.nces.ed.gov

Section 13-6

Perform these steps.

  a. *Find the Spearman rank correlation coefficient.*
  b. *State the hypotheses.*
  c. *Find the critical value. Use $\alpha = 0.05$.*
  d. *Make the decision.*
  e. *Summarize the results.*

6. **Subway and Commuter Rail Passengers** Six cities are randomly selected, and the number of daily passenger trips (in thousands) for subways and commuter rail service is obtained. At $\alpha = 0.05$, is there a relationship between the variables? Suggest one reason why the transportation authority might use the results of this study.

City	1	2	3	4	5	6
Subway	845	494	425	313	108	41
Rail	39	291	142	103	33	38

Source: American Public Transportation Association.

Northeastern University
College *of* Professional Studies

Use random numbers to simulate the experiments. The number in parentheses is the number of times the experiment should be repeated.

16. **Prizes in Caramel Corn Boxes** A caramel corn company gives four different prizes, one in each box. They are placed in the boxes at random. Find the average number of boxes a person needs to buy to get all four prizes. (40)

18. **Lottery Winner** To win a certain lotto, a person must spell the word *big*. Sixty percent of the tickets contain the letter *b*, 30% contain the letter *i*, and 10% contain the letter *g*. Find the average number of tickets a person must buy to win the prize. (30)

Assignment Rubric

Criteria	Above Standards	Meets Standards	Approaching Standards	Below Standards
Analysis	Incorporates R code and the outputs. Uses the correct statistical test for the problem and obtains the correct results. Provides detailed analysis of the output focusing on significance results. Uses visualizations to make major points.	Provides all R code and the outputs. Uses the correct statistical method for the problem, performs the steps correctly. Includes interpretation of the output, graphs, figures, charts, and tables and the significance of the results in the analysis.	Provides R codes and outputs, but the R code does not match the outputs or is missing some code or outputs. Uses the correct statistical test for the problem, but does not perform steps correctly or obtains incorrect results. Includes limited interpretations, charts, and tables and the significance of the results in the analysis.	Does not use the correct statistical test for the problem. The conclusion does not summarize or attempt to make sense of the results. Conclusions do not reflect an understanding or reflect a misunderstanding of the material.
Interpretation	Wraps up the findings in a conclusion that provides an answer to the question(s) posed in the introduction. Makes specific recommendations based on the data presented.	The conclusion summarizes and makes sense of the results, making good points that reflect clear understanding of the assignment material.	The conclusion summarizes and makes sense of the results, making good points that reflect a basic understanding of the assignment material.	Does not provide R code or its outputs or minimal R code is provided. Includes few interpretations, charts, or tables. Does not identify the significance of the results in the analysis.
Data Visualizations	Data visualizations are appropriate for the level and type of analysis. Graphs, figures and tables communicate insights and significance to the reader. Graphs are well formatted and labeled.	Data visualizations are useful for the level and type of analysis. Graphs, figures and tables communicate the significance of the results to the reader.	Data visualizations are useful for the level and type of analysis, but graphs, figures and tables do not clearly communicate the significance of the results to the reader.	Data visualizations are used minimally or not at all. If graphs, figures and tables are used, it is unclear what they are intended to communicate or why.

Writing Mechanics, Title Page, & References	There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references.	There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references.	There are very few errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references.	There are more than five errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are incomplete.