INTRODUCTION

This report includes a diagnosis that shows how to use RStudio to perform the follow outcomes

• 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.

What is Non Parametric test?

Nonparametric statistics refers to a statistical method in which the data are not assumed to come from prescribed models that are determined by a small number of parameters; examples of such models include the normal distribution model and the linear regression model. Nonparametric statistics sometimes uses data that is ordinal, meaning it does not rely on numbers, but rather on a ranking or order of sorts. For example, a survey conveying consumer preferences ranging from like to dislike would be considered ordinal data (Investopedia., 2021).

Types of Nonparametric Tests

Some of the nonparametric tests are (Statistics., 2021):

- **1-sample sign test** - Use this test to estimate the median of a population and compare it to a reference value or target value.
- **1-sample Wilcoxon signed rank test** - With this test, you also estimate the population median and compare it to a reference/target value. However, the test assumes your data comes from a symmetric distribution (like the Cauchy distribution or uniform distribution).
- **Kruskal-Wallis test** - Use this test instead of a one-way ANOVA to find out if two or more medians are different. Ranks of the data points are used for the calculations, rather than the data points themselves.
- **Spearman Rank Correlation** - Use when you want to find a correlation between two sets of data.

Advantages and Disadvantages of Non parametric test

Compared to parametric tests, nonparametric tests have several advantages, including:

- More statistical power when assumptions for the parametric tests have been violated. When assumptions haven't been violated, they can be almost as powerful.
- Fewer assumptions (i.e., the assumption of normality doesn't apply).
- Small sample sizes are acceptable.
- They can be used for all data types, including nominal variables, interval variables, or data that has outliers or that has been measured imprecisely.

However, they do have their disadvantages. The most notable ones are (Statistics., 2021):

- Less powerful than parametric tests if assumptions haven't been violated.
- More labor-intensive to calculate by hand (for computer calculations, this isn't an issue).
- Critical value tables for many tests aren't included in many computers' software packages. This is compared to tables for parametric tests (like the z-table or t-table) which usually *are* included.

ANALYSIS SECTION

Libraries Used

```r
#Libraries used
library(BSDA)
library(stats)
```

First task Part One

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

a. State the hypotheses and identify the claim.

#Ho => Median = 3000

#Ha => Median! = 3000

b. Find the critical value(s).

Critical value = 5

c. Compute the test value.

```
##
##  One-sample Sign-Test
##
## data:  sample
## s = 20, p-value = 1.907e-06
## alternative hypothesis: true median is not equal to 0
## 95 percent confidence interval:
##  2724.426 3681.365
## sample estimates:
## median of x
##      2931.5
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level   L.E.pt   U.E.pt
## Lower Achieved CI     0.8847 2758.000 3540.000
## Interpolated CI       0.9500 2724.426 3681.365
## Upper Achieved CI     0.9586 2720.000 3700.000
```

d. Make the decision.

```
ifelse(result$p.value < alpha, "Reject the null hypothesis", "Fail to reject the null hypothesis")
```

```
## [1] "Reject the null hypothesis"
```

```
ifelse(testvalue > critical_value, "Reject the null hypothesis", "Fail to reject the null hypothesis")
```

```
## [1] "Reject the null hypothesis"
```

e. Summarize the results.

We have enough evidence to reject the null hypothesis that the median number for the paid attendance at 20 local football games is 3000.

First task Part Two

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?

a. State the hypotheses and identify the claim.

#Ho => Median = 200

#Ha => Median < 200

b. Find the critical value(s).

Critical value = -1.6844

c. Compute the test value.

z value = -1.423025

d. Make the decision.

```
ifelse(z > critical_value, "Fail to reject null hypothesis", "Reject null hypothesis")
```

```
## [1] "Fail to reject null hypothesis"
```

e. Summarize the results.

We do not have enough evidence to reject the null hypothesis. So, we can conclude that the median tickets sold per day is less than 200.

Second Task Part One

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				

a. State the hypotheses and identify the claim.

#Ho => There is no difference in the sentence received by each gender

#H1 => There is difference in the sentence received by each gender

b. Find the critical value(s).

Critical value = 1.959964

c. Compute the test value.

```
##
##  Wilcoxon rank sum test
##
## data:  Males and Females
## W = 113, p-value = 0.1357
## alternative hypothesis: true location shift is not equal to 0
```

d. Make the decision.

```
ifelse(result$p.value < alpha, "Reject the null hypothesis", "Fail to reject null hypothesis")
```

```
## [1] "Fail to reject null hypothesis"
```

e. Summarize the results.

We do not have enough evidence to reject the null hypothesis. Hence, there is no difference in the sentence received by each gender

Second Task Part Two

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

a. State the hypotheses and identify the claim.

#Ho => There is difference in number of wins.

#H1 => There is no difference in number of wins.

b. Find the critical value(s).

Critical value = 1.959964

c. Compute the test value.

```
##
##  Wilcoxon rank sum test
##
## data:  NL and AL
## W = 59, p-value = 0.6657
## alternative hypothesis: true location shift is not equal to 0
```

d. Make the decision.

```
ifelse(result$p.value < alpha, "Reject the null hypothesis", "Fail to reject null hypothesis")
```

```
## [1] "Fail to reject null hypothesis"
```

e. Summarize the results.

We do not have enough evidence to reject the null hypothesis. Hence, there is difference in number of wins.

Third Task Part One

$$w_s = 13,\ n = 15,\ \alpha = 0.01,\ \text{two-tailed}$$

a. Find the critical value(s).

Critical value = 16

b. Make the decision.

```
ifelse(ws > criticalvalue, "Fail to reject null hypothesis", "Reject the null hypothesis")
```

```
## [1] "Reject the null hypothesis"
```

Third Task Part Two

$$w_s = 32,\ n = 28,\ \alpha = 0.025,\ \text{one-tailed}$$

a. Find the critical value(s).

Critical value = 116

b. Make the decision.

```
ifelse(ws > criticalvalue, "Fail to reject null hypothesis", "Reject the null hypothesis")
```

```
## [1] "Reject the null hypothesis"
```

Third Task Part Three

$$w_s = 65,\ n = 20,\ \alpha = 0.05,\ \text{one-tailed}$$

a. Find the critical value(s).

Critical value = 60

b. Make the decision.

```
ifelse(ws > criticalvalue, "Fail to reject null hypothesis", "Reject the null hypothesis")
```

```
## [1] "Fail to reject null hypothesis"
```

Third Task Part Four

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

a. Find the critical value(s).

Critical value = 26

b. Make the decision.

```
ifelse(ws > criticalvalue, "Fail to reject null hypothesis", "Reject the null hypothesis")
```

```
## [1] "Reject the null hypothesis"
```

Fourth Task

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

a. State the hypotheses and identify the claim.

#Ho => There is no difference in the mean of math's score between three regions

#H1 => There is a difference in the mean of math's score between three regions

b. Find the critical value(s).

Critical value = 5.991465

c. Compute the test value.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  scores by group
## Kruskal-Wallis chi-squared = 4.1674, df = 2, p-value = 0.1245
```

d. Make the decision.

```
ifelse(result$p.value < alpha, "Reject the null hypothesis", "Fail to reject null hypothesis")
```

```
## [1] "Fail to reject null hypothesis"
```

e. Summarize the results.

We do not have enough evidence to reject the null hypothesis. Hence, there is no difference in the mean of math's score between three regions

Fifth Task

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

a. State the hypotheses and identify the claim.

#Ho => There is no relationship between the subway and rail commuter passengers

#H1 => There is a relationship between the subway and rail commuter passengers

b. Find the critical value(s).

Critical value = 0.886

c. Compute the test value.

```
##
##  Spearman's rank correlation rho
##
## data:  data$subway and data$rail
## S = 14, p-value = 0.2417
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.6
```

d. Make the decision.

```
ifelse(result$p.value < alpha, "Reject the null hypothesis", "Fail to reject null hypothesis")
```

```
## [1] "Fail to reject null hypothesis"
```

e. Summarize the results.

We do not have enough evidence to reject the null hypothesis. Hence, there is no relationship between the subway and rail commuter passengers.

Sixth Task Part One

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)

Result

One needs to pick an average of 7.225 boxes to get all four prizes. (The results may wary if try to run the code again as the sample are being picked randomly every time, we run the code)

Sixth Task Part Two

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)

Result

One needs to pick an average of 12.46667 tickets to win prizes. (The results may wary if try to run the code again as the sample are being picked randomly every time, we run the code)

CONCLUSION

Finally, I increased my understanding of non parametric testing by studying numerous types of tests such as the Sign test, Wilcoxon Rank Sum test, Wilcoxon Signed Rank test, Kruskal-Wallis test, Spearman Rank Correlation Coefficient test, and Run test for randomness while working on several problems. While working on this report, I also learnt about Monte Carlo tests. I utilized BSDA and the stats library to conduct the above-mentioned test, which helped me better understand the concepts while keeping them simple and easy to execute.

BIBLIOGRAPHY

1) Non Parametric Data and Tests (Distribution Free Tests). (2021, May 31). Statistics How To. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/parametric-and-non-parametric-data/

2) How Nonparametric Statistics Work. (2021, October 22). Investopedia. https://www.investopedia.com/terms/n/nonparametric-statistics.asp#:%7E:text=Nonparametric%20statistics%20refers%20to%20a,and%20the%20linear%20regression%20model.

3) Kotz, S.; et al., eds. (2006), Encyclopedia of Statistical Sciences, Wiley.

4) Lindstrom, D. (2010). Schaum's Easy Outline of Statistics, Second Edition (Schaum's Easy Outlines) 2nd Edition. McGraw-Hill Education