Logo

Description automatically generated

# ALY 6015: INTERMEDIATE ANALYTICS

## **Assignment 5: Nonparametric Statistical Methods/ Sampling and Simulation**

Submitted to

### Prof. Fatemeh Ahmadi Abkenari

Submitted by

Abhilash Dikshit

Mrityunjay Gupta

Siddharth Alashi

Smit Parmar

**Assignment 5: Feature Selection in R**

Abhilash Dikshit, Siddharth Alashi, Mrityunjay Gupta, Smit Parmar

*College of Professional Studies*

*Northeastern University*

*Vancouver, Canada*

**Introduction**:

In this assignment, we are using nonparametric statistical methods and sampling and simulation to perform the following steps:

1. State the hypotheses and identify the claim.
2. Find the critical value.
3. Compute the test value.
4. Make the decision.
5. Summarize the results.

**Problem#1: Winning Baseball Games**

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 α = 0.05, is there sufficient evidence to conclude a difference in the number of wins?

A picture containing table

Description automatically generated

**Answer:**

**Step 1: State the hypotheses and identify the claim.**

*H*0: There is no difference in the number of wins for each Eastern Division leagues. (claim)

*H*1: There is a difference in the number of wins for each Eastern Division leagues.

**Step 2: Find the critical value.**

Since α = 0.05 and the test is a two tailed test, we will use the critical values of -1.96 and 1.96 from Table E.

**Step 3: Compute the test value.**

1. **Combining the data from the two samples and arranging the combined data in ascending order, and later ranking each value.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample** | Value | Rank | Rank (Adjusted for ties) |
| A | 86 | 1 | 1 |
| A | 88 | 2 | 2.5 |
| N | 88 | 3 | 2.5 |
| N | 89 | 4 | 4.5 |
| A | 89 | 5 | 4.5 |
| N | 90 | 6 | 6 |
| N | 91 | 7 | 7.5 |
| A | 91 | 8 | 7.5 |
| N | 92 | 9 | 9 |
| N | 95 | 10 | 11 |
| A | 95 | 11 | 11 |
| A | 95 | 12 | 11 |
| N | 96 | 13 | 13.5 |
| N | 96 | 14 | 13.5 |
| A | 97 | 15 | 15 |
| A | 100 | 16 | 16.5 |
| N | 100 | 17 | 16.5 |
| N | 101 | 18 | 18.5 |
| A | 101 | 19 | 18.5 |
| A | 102 | 20 | 20 |
| A | 104 | 21 | 21 |
| A | 108 | 22 | 22.5 |
| N | 108 | 23 | 22.5 |

1. **Sum the ranks of the group with smaller size.**

The sum of ranks for sample N is:

RN=2.5+4.5+6+7.5+9+11+13.5+13.5+16.5+18.5+22.5=125*R*1​=2.5+4.5+6+7.5+9+11+13.5+13.5+16.5+18.5+22.5=125

and the sum of ranks of sample A is:

RA=1+2.5+4.5+7.5+11+11+15+16.5+18.5+20+21+22.5=151*R*2​=1+2.5+4.5+7.5+11+11+15+16.5+18.5+20+21+22.5=151

Hence, the test statistic is R= RN= 125.

1. **Use the formula to find the test value.**

|  |
| --- |
|  |

**Step 4: Make the decision.**

Since -0.431 lies outside of the critical region, we do not reject the null hypothesis (A = N). That is, our findings are statistically significant (at the significance level 0.05).

**Step 5: Summarize the results.**

It is concluded that the null hypothesis Ho *is not rejected.* Therefore, there is not enough evidence to claim that there is no difference in the number of wins for each Eastern Division leagues at the α=0.05 significance level.

**Problem#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 α = 0.05.

Table

Description automatically generated

**Answer:**

**Step 1: State the hypotheses and identify the claim.**

*H*0: There is no difference in total mathematics literacy scores (both   
genders) for selected countries in different parts of the world. (claim)

*H*1: There is a difference in total mathematics literacy scores (both   
genders) for selected countries in different parts of the world.

**Step 2: Find the critical value.**

Since α = 0.05, we will use the chi-square table, Table G, with d.f. = k-1 = 3-1 = 2 where k is number of groups, critical value is 5.991.

**Step 3: Compute the test value.**

1. **Arranging the data from lowest to highest and ranking each value.**

|  |  |  |
| --- | --- | --- |
| **Group** | **Value** | **Rank** |
| W | 381 | 1 |
| A | 391 | 2 |
| W | 406 | 3 |
| W | 411 | 4 |
| W | 474 | 5 |
| E | 496 | 6 |
| E | 510 | 7 |
| E | 513 | 8 |
| E | 520 | 9 |
| A | 523 | 10 |
| W | 527 | 11 |
| A | 547 | 12.5 |
| A | 547 | 12.5 |
| E | 548 | 14 |
| A | 549 | 15 |

1. **Find the sum of the ranks of each group**

RW = 1+3+4+5+11= 24

RE = 6+7+8+9+14= 44

RA = 2+10+12.5+12.5+15= 52

n = n1 + n2 +...+ nk = 5 + 5 + 5= 15

Mean Rank W = 24 / 5 = 4.8

Mean Rank E = 44 / 5 = 8.8

Mean Rank A = 52 / 5 = 10.4

1. **Substitute the Formula**

|  |  |
| --- | --- |
|  |  |

**Step 4: Make the decision.**

Since the test value of 4.16 is less than the critical value of 5.991, the decision is not to reject the null hypothesis. The mean scores of all groups assume to be equal. In other words, the difference between the mean ranks of all groups is not big enough to be statistically significant.

**Step 5: Summarize the results.**

There is not enough evidence to reject the claim that there is no difference in total mathematics literacy scores (both genders) for selected countries in different parts of the world. Hence, the differences are not significant at α = 0.05.

**Problem#3: 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 α = 0.05, test the claim that there is no difference in the sentence received by each gender. The data (in months) are shown here.

A picture containing text

Description automatically generated

**Answer:**

**Step 1: State the hypotheses and identify the claim.**

*H*0: There is no difference in the sentence received by males and females for certain type of crime. (claim)

*H*1: There is a difference in the sentence received by males and females for certain type of crime.

**Step 2: Find the critical value.**

Since α = 0.05 and the test is a two tailed test, we will use the critical values of -1.96 and 1.96 from Table E.

**Step 3: Compute the test value.**

1. **Arranging the data from lowest to highest and ranking each value.**

|  |  |  |
| --- | --- | --- |
| **Months** | **Groups** | **Ranks** |
| 2 | F | 1 |
| 3 | F | 2 |
| 4 | F | 3 |
| 5 | F | 4 |
| 6 | M | 5 |
| 7 | F | 6 |
| 8 | M | 7 |
| 9 | F | 8 |
| 11 | F | 9 |
| 12 | M | 10.5 |
| 12 | F | 10.5 |
| 13 | M | 12 |
| 14 | M | 13 |
| 15 | M | 14 |
| 16 | F | 15 |
| 17 | F | 16 |
| 19 | M | 17 |
| 21 | F | 18 |
| 22 | M | 19 |
| 23 | F | 20 |
| 24 | M | 21 |
| 26 | M | 22.5 |
| 26 | F | 22.5 |
| 27 | M | 24 |
| 30 | F | 25 |
| 32 | M | 26 |

Mean of Male = 18.17

Mean of Female = 13.29

1. **Find the sum of the ranks of each group**

RM= 5+7+10.5+12+13+14+17+19+21+22.5+24+26= 191

RF = 1+2+3+4+6+8+9+10.5+15+16+18+20+22.5+25=160

R = RM = 191

1. **Substitute the Formula**

|  |
| --- |
|  |

**Step 4: Make the decision.**

Since 1.49 is less than critical value 1.96 i.e., 1.49 < 1.96, we do not reject the null hypothesis.

**Step 5: Summarize the results.**

It is concluded that there is enough evidence to support the claim that there is no difference in the sentence received by each gender at the α=0.05 significance level.