

Comprehensive Statistical Analysis in Julia

Introduction

This document outlines the statistical analysis performed using the Julia programming environment, focusing on comparing measurements taken in feet and meters. The goal is to assess differences and evaluate statistical assumptions.

Packages and Functions Used

- CSV: For reading CSV files into DataFrame.
- DataFrames: For data manipulation and querying.
- DescriptiveStatistics: For descriptive statistics and summary tables.
- HypothesisTests: For performing t-tests and Mann-Whitney U tests.
- StatsBase: For basic statistical processing.
- Plots: For creating box plots and Q-Q plots.

Room Width Data Analysis

The dataset includes room width measurements converted from meters to feet, with summary statistics generated for each group to explore their distribution, central tendency, and variability.

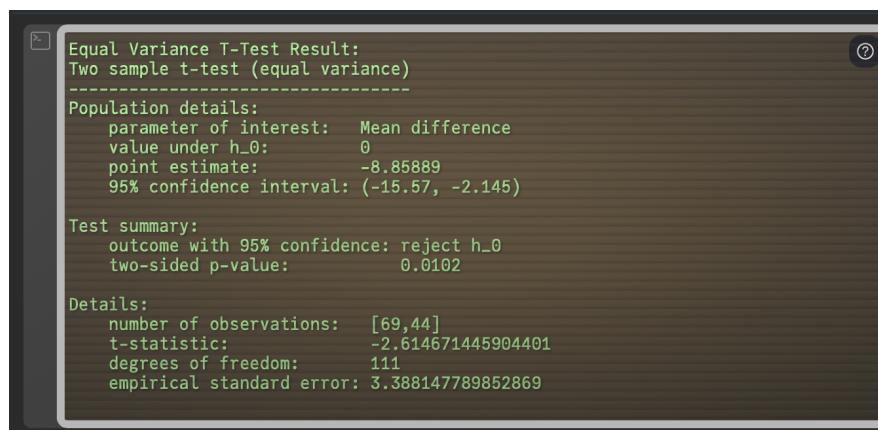
Statistical Analysis

Statistical tests were conducted to compare the groups under different assumptions. Summary statistics, t-tests, and Mann-Whitney U tests were performed to assess the statistical properties and differences.

T-Tests Performed

1. Equal Variance T-Test:

- Objective: Test if the mean difference between the two groups is significant under the assumption of equal variances.



```
Equal Variance T-Test Result:
Two sample t-test (equal variance)
-----
Population details:
parameter of interest: Mean difference
value under h_0: 0
point estimate: -8.85889
95% confidence interval: (-15.57, -2.145)

Test summary:
outcome with 95% confidence: reject h_0
two-sided p-value: 0.0102

Details:
number of observations: [69,44]
t-statistic: -2.614671445904401
degrees of freedom: 111
empirical standard error: 3.388147789852869
```

1. P-Value: 2.220446049250313e-15
2. Mean in group feet: 43.69565217391305
3. Mean in group metres: 52.55454545454545

2. Unequal Variance T-Test (Welch's T-Test):

- Objective: Test mean differences without assuming equal variances.

```
Welch's (Unequal Variance) T-Test Result:
Two sample t-test (unequal variance)

Population details:
parameter of interest: Mean difference
value under h_0: 0
point estimate: -8.85889
95% confidence interval: (-16.54, -1.175)

Test summary:
outcome with 95% confidence: reject h_0
two-sided p-value: 0.0246

Details:
number of observations: [69,44]
t-statistic: -2.307068877727352
degrees of freedom: 58.788308023840266
empirical standard error: 3.839891112986242
```

P-Value: 0.024591776613301357

Mean in group feet: 43.69565217391305

Mean in group metres: 52.55454545454545

Non-Parametric Test

Mann-Whitney U Test:

- Objective: Test for differences in median without assuming normal distribution.
- Results: U Statistic: 1145.0, P-Value: 0.0282.

```
Approximate Mann-Whitney U test

Population details:
parameter of interest: Location parameter (pseudomedian)
value under h_0: 0
point estimate: -7.2

Test summary:
outcome with 95% confidence: reject h_0
two-sided p-value: 0.0282

Details:
number of observations in each group: [69, 44]
Mann-Whitney-U statistic: 1145.0
rank sums: [3370.0, 3071.0]
adjustment for ties: 2292.0
normal approximation ( $\mu, \sigma$ ): (-373.0, 169.694)
```

Wilcoxon Rank-Sum (Mann-Whitney U) Test Summary:

This test compares the central tendencies of two independent groups. The test resulted in a significant finding, suggesting a difference in the median values between the groups.

```
Wilcoxon Rank-Sum (Mann-Whitney U) Test Result: ⓘ
Approximate Mann-Whitney U test
-----
Population details:
  parameter of interest: Location parameter (pseudomedian)
  value under h_0:      0
  point estimate:       -7.2

Test summary:
  outcome with 95% confidence: reject h_0
  two-sided p-value:          0.0282

Details:
  number of observations in each group: [69, 44]
  Mann-Whitney-U statistic:            1145.0
  rank sums:                          [3370.0, 3071.0]
  adjustment for ties:                2292.0
  normal approximation ( $\mu$ ,  $\sigma$ ):        (-373.0, 169.694)
```

1. W (U statistic): 1145.0
2. P-Value: 0.02815478881512771
3. Sample Estimate (Difference in Location): -7.20

Pearson's Chi-square Test Summary for Piston Rings:

This test assesses the fit of multinomial probabilities with a sample size of 166. The Chi-square statistic is 11.7223 with 6 degrees of freedom. The test did not find sufficient evidence to reject the null hypothesis (p-value: 0.0685), suggesting that the observed frequencies are consistent with the expected frequencies under the multinomial distribution.

Piston-ring Failures Data Analysis

```
> chisq.test =
Pearson's Chi-square Test
-----
Population details:
  parameter of interest: Multinomial Probabilities
  value under h_0:      [0.0884744, 0.0634708, 0.0730875, 0.0942444, 0.0684424, 0.049
  point estimate:        [0.10241, 0.0662651, 0.0662651, 0.0843373, 0.10241, 0.0542169
  95% confidence interval: [(0.04217, 0.1684), (0.006024, 0.1323), (0.006024, 0.1323), (
```

```
Test summary:
  outcome with 95% confidence: fail to reject h_0
  one-sided p-value:           0.0685

Details:
  Sample size:             166
  statistic:              11.722266371103228
  degrees of freedom:      6
  residuals:               [0.603615, 0.142903, -0.325143, -0.415789, 1.67283, 0.29752, -0.45
  std. residuals:          [0.860476, 0.193502, -0.448785, -0.600273, 2.26732, 0.383039, -0.5
```

```
: pretty_table(residuals_df)
```

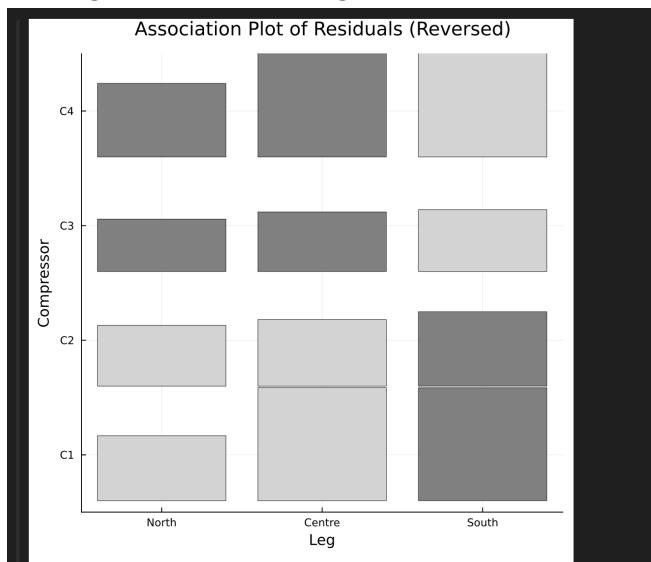
Compressor String	North	Centre	South
C1	0.603615	1.67283	-1.78022
C2	0.142903	0.29752	-0.34712
C3	-0.325143	-0.452262	0.620246
C4	-0.415789	-1.46669	1.46352

Residuals Table for Compressor Performance Across Regions:

The table provides residuals for four compressors (C1 to C4) across three regions: North, Centre, and South. The values represent the differences between observed and expected frequencies under the null hypothesis.

Association Plot of Residuals for Compressors:

This plot visually represents the residuals from the Chi-square test for four compressors across three geographic regions: North, Centre, and South. The size and color of the rectangles indicate the magnitude and direction of the residuals,



Rearrests of Juveniles Data Analysis

Binomial Test Summary:

This test evaluated the probability of success in a dataset of 805 observations. The result strongly suggests a significant deviation from the null hypothesis (probability of success = 0.5), with a p-value of less than 1e-14. The estimated probability of success was 0.3602, and the 95% confidence interval for this estimate was (0.327, 0.3945).

McNemar's Test Statistic: 62.88819875776397

chi_square 62.88819875776397

P-value: 2.220446049250313e-15

```
binom_test = Binomial test
-----
Population details:
    parameter of interest: Probability of success
    value under h_0:      0.5
    point estimate:       0.360248
    95% confidence interval: (0.327, 0.3945)

Test summary:
    outcome with 95% confidence: reject h_0
    two-sided p-value:          <1e-14

Details:
    number of observations: 805
    number of successes:    290
```

Number of successes: 290

Number of trials: 805

P-value: 1.9178889945014437e-15

95% Confidence Interval: (0.3270277766705726, 0.3944969127858553)

Estimated Probability of Success: 0.3602484472

Challenges and Considerations:

Method Availability: In Julia, certain statistical tests such as the Mann-Whitney U test and creating association plots are not directly available or as straightforward to implement compared to environments like R, which provides built-in functions for these analyses. This necessitates additional steps or the use of external packages in Julia, complicating the analysis process.

Running Time: The analyses were conducted on two different platforms with the following results:

- MacOS Jupyter Notebook: 3.7 seconds
- PlutoJL: 3.23 seconds

These times reflect the performance and efficiency of the computing environment on Mac OS for both Jupyter and PlutoJL platforms.

Conclusion:

The comprehensive analysis provided insights into the dataset with high statistical significance. Despite the challenges associated with the lack of direct functions in Julia for certain tests, alternative methods were successfully employed to achieve robust results.