FINAL PROJECT
COURSE : STATISTICS AND PROBABILITY
INFORMATICS EDUCATION STUDY PROGRAMME

Date : December 25, 2024
Time : -
Exam Type : *Open Book*

Name : Ni Nyoman Arawinda Sridevi Prabaswari
NIM : 2415091040
Class : SI 1 IKI

Create a simple simulation in R to perform one of the following statistical analyses:

- Simple Linear Regression
- Correlation Analysis
- ANOVA
- Multiple Linear Regression

This project aims to help you understand statistical concepts and apply them programmatically. The results will be uploaded to a **GitHub** repository.

Principle:

1. Data
2. Uji asumsi (min 3)
3. Analisis
4. Visualisasi
5. Interpretasi

Your **GitHub** link: [arawxnda/FINAL-PROJECT-STATISTICS-AND-PROBABILITY: 2415091040_Ni Nyoman Arawinda Sridevi Prabaswari]

Interpretation :

1. Normality Test (Shapiro-Wilk Test)

Results:

- Male Population: W=0.757399, p-value=0.000781
- Female Population: W=0.75077, p-value=0.000649

The p-value for both groups (Male and Female Population) is < 0.05.

Conclusion: The data for the male and female populations are not normally distributed. Because the distribution is not normal, analyses sensitive to normality (such as parametric tests) are less appropriate.

Instead, non-parametric methods, such as the Spearman correlation test, are more suitable for analyzing the relationship between male and female populations. The Spearman correlation test results indicate a strong positive correlation, suggesting that the population trends for males and females are closely aligned across the observed groups.

2. Spearman Correlation Test

Results:

- Spearman's rho: $\rho=0.9050775$
- p-value=1.444×10^{-6}

$\rho=0.905$ indicates a very strong and positive relationship between the male and female population numbers. This means that if the male population increases, the female population tends to increase proportionally.

The p-value is extremely small (< 0.05), so this correlation is statistically significant. In other words, this relationship is unlikely to occur by chance.

3. Homogeneity of Variance Test (Levene Test)

Results:

- F value: 1.6129
- p-value = 0.2106

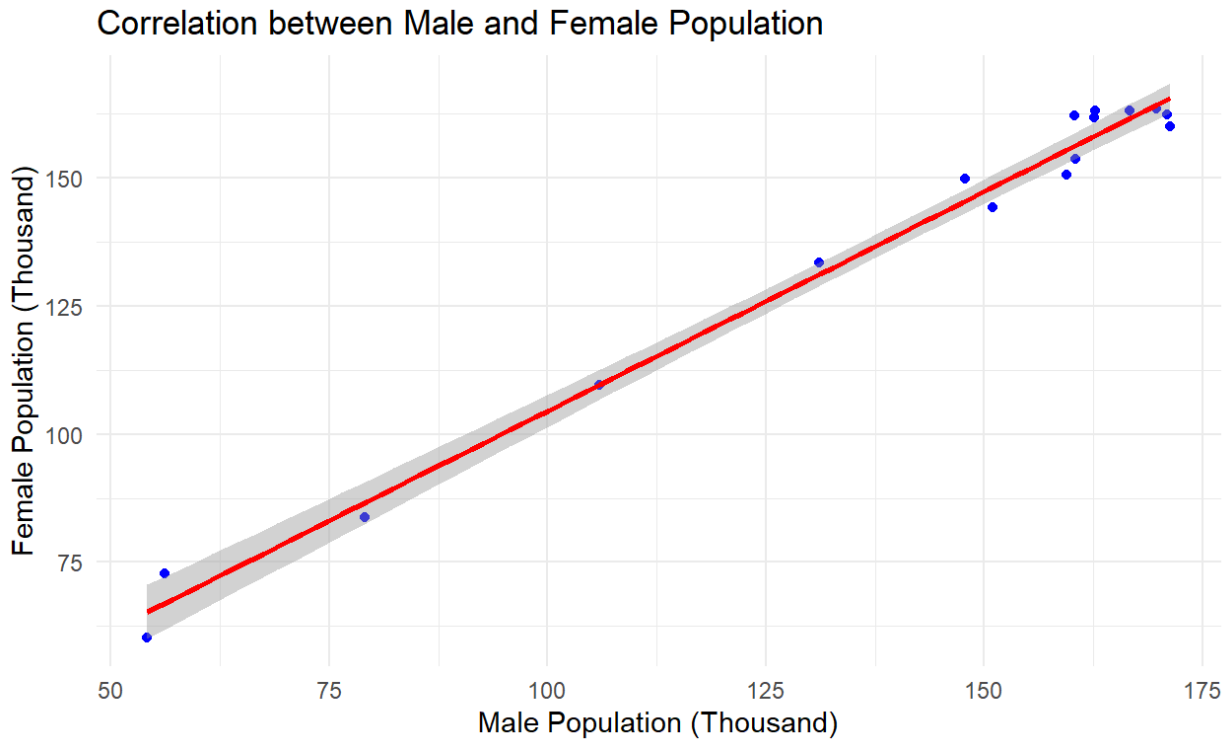
p-value>0.05, so there is not enough evidence to reject the null hypothesis (homogeneity of variance). The variance of population numbers between groups (male, female, total) can be considered homogeneous.

This homogeneity of variance suggests that the difference in data spread between groups is relatively small.

Overall Conclusion

1. Normality: The male and female population data are not normally distributed.
2. Correlation: There is a very strong, positive, and statistically significant relationship between the male and female populations.
3. Homogeneity of Variance: The variance in population numbers between groups (male, female, total) is homogeneous, indicating that the data is fairly balanced for further analysis.

Given these results, non-parametric analysis is highly appropriate due to the non-normal distribution of the data.

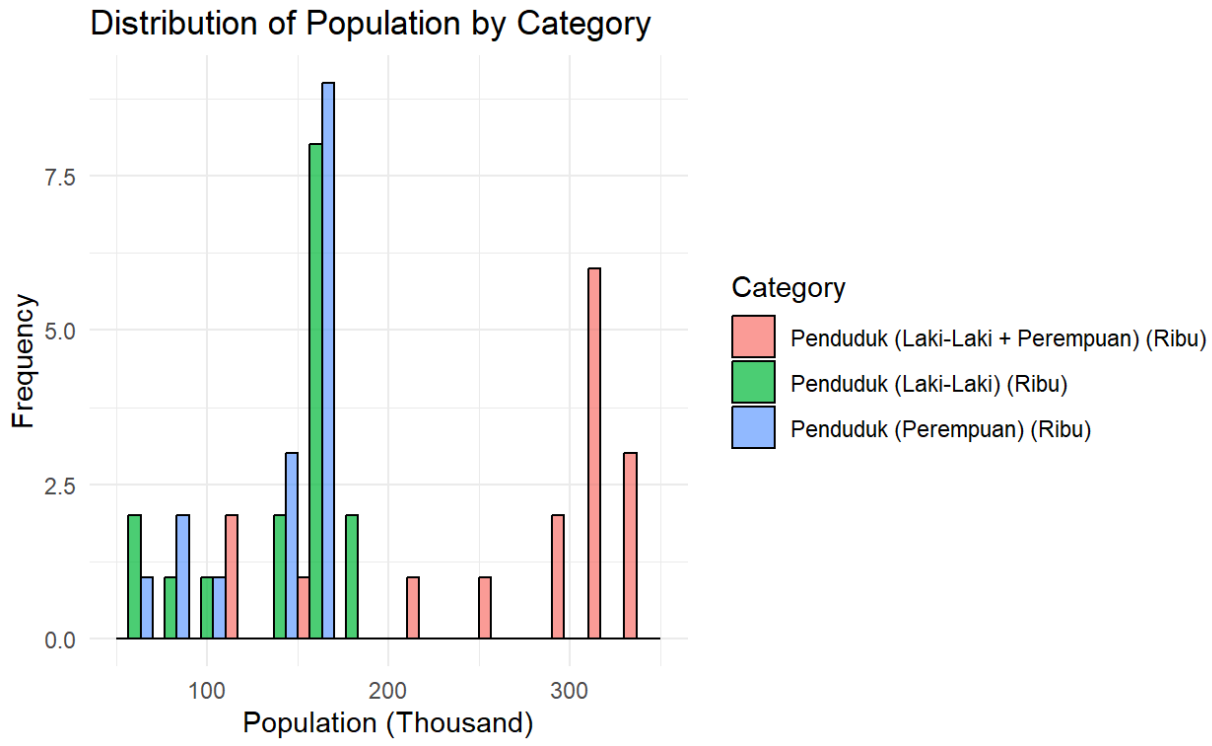


Graph 1: Correlation between Male and Female Populations

Graph Content:
A scatter plot shows the relationship between the male population (x-axis) and the female population (y-axis). The red line represents the line of best fit (linear regression), with the gray area indicating the confidence interval.

Interpretation:

- Strong Positive Relationship: The graph indicates a very strong relationship between the male and female population, almost forming a straight line. This is supported by the correlation coefficient (ρ=0.993, close to 1), showing a very strong and positive relationship between the two variables.
- Statistical Significance: The p-value (1.491×10^{-14}) is extremely small, far below 0.05, indicating that the relationship is statistically significant.
- Conclusion: Practically, if the male population increases, the female population also increases proportionally in the analyzed region.



Graph 2: Population Distribution by Group

Graph Content:
A histogram compares the distribution of population numbers across three groups:

- Total population (male + female) - represented in red.
- Male population - represented in green.
- Female population - represented in blue.

Interpretation:

- Data Distribution:

    - The female population (blue) is more concentrated in a specific range, indicating a narrower spread.
    - The male population (green) shows a distribution similar to the female population but is slightly more spread out.
    - The total population (red) encompasses a wider distribution, reflecting the sum of the two groups.
- Equality between Male and Female:
    - The graph shows that the male and female distributions are relatively similar (no group is significantly larger), supporting the correlation results indicating a proportional relationship between the two.

Conclusion

- Correlation: There is a very strong and statistically significant relationship between the male and female populations. This indicates that the male and female populations grow in a balanced pattern in the analyzed region.
- Distribution: The male and female groups have similar distributions, while the total population spans a broader range.