CSDS 313/413: Introduction to Data Analysis
Homework 3: Pairwise Association
Solutions

Your Name

November 8, 2025

1 Task 1: Association Between Genomic Variants

1.1 Part (a): Two Variables from p1a.csv

1.1.1 Test Statistics and P-values

Mutual Information (MI)

- Observed MI value:
- Number of permutations (N):
- P-value:
- Selected significance level (α):
- Conclusion:

Jaccard Index (JI)

- Observed JI value:
- Number of permutations (N):
- P-value:
- Selected significance level (α):
- Conclusion:

Pearson's Chi-squared (χ^2)

- Observed χ^2 value:
- Degrees of freedom:
- P-value:
- Selected significance level (α):
- Conclusion:

1.1.2 Analysis and Interpretation

Statistical Significance:

Strength of Association:

Agreement Between Statistics:

Explanation of Any Discrepancies:

1.1.3 Visualization

1.2 Part (b): 105 Variable Pairs from p1b.csv

1.2.1 Test Statistics and P-values

Number of Permutations (N):

Selected Significance Level (α):

Multiple Hypothesis Correction Method:

1.2.2 Results Summary

Mutual Information (MI)

- Number of significantly associated pairs:
- List of significant pairs:

Jaccard Index (JI)

- Number of significantly associated pairs:
- List of significant pairs:

Pearson's Chi-squared (χ^2)

- Number of significantly associated pairs:
- List of significant pairs:

1.2.3 Comparison of Statistics

Overlap Between Statistics:

- MI and JI overlap:
- MI and χ^2 overlap:
- JI and χ^2 overlap:
- All three statistics overlap:

Which Two Statistics Are Most Similar:

Preferred Test Statistic and Justification:

Impact of Using Only Preferred Statistic:

1.2.4 Visualizations

2 Task 2: Association Between Continuous Variables

2.1 Part (a): Variable Pair from p2a.csv

2.1.1 Test Statistics and P-value

Pearson Correlation (r_a):

- Correlation coefficient:
- P-value (p_a):
- Selected significance level (α):
- Sample size:

2.1.2 Analysis and Interpretation

Statistical Significance:

Magnitude of Association:

Direction of Association:

2.1.3 Visualization

2.2 Part (b): Comparison of p2a.csv and p2b.csv

2.2.1 Test Statistics and P-value for p2b.csv

Pearson Correlation (r_b):

- Correlation coefficient:
- P-value (p_b):
- Selected significance level (α):
- Sample size:

2.2.2 Comparison Analysis

Comparison of Correlations (r_a vs r_b):

Stronger Association Based on Correlations:

Comparison of P-values (p_a vs p_b):

Stronger Association Based on P-values:

Agreement Between Correlations and P-values:

Explanation of Any Discrepancy:

Visual Assessment from Scatter Plots:

Agreement with Statistical Measures:

2.2.3 Visualizations

2.3 Part (c): Comparison of p2a.csv and p2c.csv

2.3.1 Test Statistics and P-value for p2c.csv

Pearson Correlation (r_c):

- Correlation coefficient:
- P-value (p_c):
- Selected significance level (α):
- Sample size:

2.3.2 Comparison Analysis

Comparison of Correlations (r_a vs r_c):

Stronger Association Based on Correlations:

Comparison of P-values (p_a vs p_c):

Stronger Association Based on P-values:

Agreement Between Correlations and P-values:

Explanation of Any Discrepancy:

Visual Assessment from Scatter Plots:

Agreement with Statistical Measures:

2.3.3 Visualizations

3 Code Appendix