

CS6530 Assignment 1 Report

Brandon Rodriguez

May 19, 2020

All answers are acquired from associated Python program, which was written from scratch for this assignment. See readme.md at project root for documentation on running it.

Question 1

0.1 Midterm Score Calculations

Rounded to nearest 2 decimal places.

Max: 100
Min: 37
Q1: 77.0
Median: 77.0
Q3: 77.0
Mean: 76.72
Mode: 77
Variance: 173.11

0.2 Final Score Calculations

Rounded to nearest 2 decimal places.

Max: 100
Min: 35
Q1: 89.0
Median: 89.0
Q3: 89.5
Mean: 87.08
Mode: 97
Variance: 119.11

Question 2

0.3 Compare Variance before and after Normalization

Rounded to nearest 4 decimal places.

Orig Midterm Score Variance: 173.1058

Z-Score Normalized Midterm Variance: 0.9999

Orig Final Score Variance: 119.1129

Z-Score Normalized Final Score Calculations: 1.0

0.4 Normalized value of an original score of 90

Rounded to nearest 4 decimal places.

Record 223 has both a original midterm and original final of 90.

Normalized, this comes to:

Midterm: 1.0097

Final: 0.2672

We have two different values because we're using z-score normalization, which normalizes around the mean and standard deviation.

If we used a different normalization, such as "Min-Max" normalization, then we'd have much closer normalized values. Note that because the minimum scores of "midterm" and "final" are different, "Min-Max" won't give identical normalized values for 90 either. But it will be a lot closer than z-score.

0.5 Pearson's Correlation Coefficient

Person's Correlation Coefficient: 0.544424742312412

0.6 Covariance Value

Covariance: 78.25419419419427

Question 3

0.7 Jaccard Coefficient

Jaccard Coefficient: 0.3222222222222224

0.8 Minkowski Distance

h of 1: 6152.0

h of 2: 715.3278968417211

h of ∞ : 170.0

0.9 Cosine Similarity

Cosine Similarity: 0.8414040256623078

0.10 Kullback-leibler Divergence

Kullback-Leibler Divergence: 0.19420536448348794

Question 4

Expected Table:

[8.94436519, 181.05563481]

[156.05563481, 3158.94436519]

Chi Stat: 2450.716326822006

Degrees of Freedom: 1

P-Value: 0.0

This indicates that there is a strong correlation, I think?