B365 HW4

Elliot Rubinstein

October 6, 2025

Question 1

Given vectors

$$u = [1, 1, 1, 0, 0, 0, 0, 0, 0, 0], \quad v = [1, 0, 1, 0, 0, 0, 0, 0, 0, 0].$$

Compute the following distances/similarities:

1. Simple Matching Coefficient (SMC)
2. Jaccard similarity
3. Minkowski distances: L_1, L_2, L_∞

Solution:

$$\text{SMC} = \frac{\text{number of matches}}{\text{total elements}} = \frac{9}{10}$$

$$\text{Jaccard} = \frac{|u \cap v|}{|u \cup v|} = \frac{2}{2+1+0} = \frac{2}{3}$$

$$L_1 = \sum |u_i - v_i| = 1$$

$$L_2 = \sqrt{\sum (u_i - v_i)^2} = 1$$

$$L_\infty = \max |u_i - v_i| = 1$$

Question 2

Suppose you roll a red number cube and a blue number cube.

- Event A: Roll a 5 on the red cube.
- Event B: Roll a 1 or 2 on the blue cube.

Compute:

$$P(A), \quad P(B), \quad P(A \cap B)$$

Then determine whether A and B are independent.

Solution:

$$P(A) = \frac{1}{6}, \quad P(B) = \frac{2}{6} = \frac{1}{3}$$

$$P(A \cap B) = \frac{1}{18}$$

Check independence:

$$P(A)P(B) = \frac{1}{18} = P(A \cap B) \quad \Rightarrow \quad A \text{ and } B \text{ are independent.}$$

Question 3

A survey in Indiana produced the following data:

	dog	**cat**
rural	276	245
urban	439	351

(a) Estimate probabilities

$$P(\text{dog}|\text{rural}) = \frac{276}{276 + 245} = 0.530$$

$$P(\text{dog}|\text{urban}) = \frac{439}{439 + 351} = 0.560$$

(b) 95% Confidence Intervals

Use $z^* = 1.96$.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Compute for each case:

$$\text{Rural: } n_r = 521, \quad \hat{p}_r = \frac{276}{521}, CI = 0.530 \pm 1.96\sqrt{\frac{0.530(1-0.530)}{521}} = (0.487, 0.573)$$

$$\text{Urban: } n_u = 790, \quad \hat{p}_u = \frac{439}{790}, CI = 0.560 \pm 1.96\sqrt{\frac{0.560(1-0.560)}{790}} = (0.521, 0.590)$$

(c) Interpretation

Would it be reasonable to conclude that dogs are more popular in urban than rural areas? Discuss using the CI overlap and sample proportions.

The two confidence intervals significantly overlap so it is not reasonable to conclude that dogs are more popular in urban than rural areas

Question 4

Compute the Chi-squared statistic for the same table.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where

$$E = \frac{(\text{row total})(\text{column total})}{\text{grand total}}.$$

	O	E	$(O-E)^2/E$
Rural, Dog	276	284.38	0.247
Rural, Cat	245	236.62	0.297
Urban, Dog	439	430.62	0.163
Urban, Cat	351	359.38	0.195

Finally,

$$\chi^2 = 0.902, df = 1$$