Jaccard Coefficient Calculations

The table shows the pathological test results for three individuals.

Name	Gender	Fever	Cough	Test-1	Test-2	Test-3	Test-4
Jack	M	Y	N	P	N	N	A
Mary	F	Y	N	P	A	P	N
Jim	M	Y	P	N	N	N	A

Calculate Jaccard coefficient for the following pairs:

- (Jack, Mary)
- (Jack, Jim)
- (Jim, Mary)

The Jaccard similarity coefficient compares members for 2 sets to see which members are shared and which are distinct. It is a measure of percentage and the higher the number the more similar the two populations.

Jaccard Index = (the number in both sets) / (the number in either set) * 100

The same formula in notation is (Statistics How To, 2016):
$J(X,Y) = |X \cap Y| / |X \cup Y|$

To calculate the Jaccard coefficient, we first convert the asymmetric variables to binary values and re-write the table. Since Gender is a symmetric variable (that is, male, female have the same weight), it is not converted.

So let Y & P = 1; N & A = 0

$$Jaccard = \frac{f_{01}+f_{10}}{f_{01}+f_{10}+f_{11}}$$

- (Jack, Mary)

  F01 = 1

  F10 = 0

  F11 = 2

  (1/3)*100 = **33.33%**

- (Jack, Jim)

  F01 = 1

  F10 = 1

  F11 = 1

  (⅔)*100 = **66.67%**

- (Jim, Mary)

  F01 = 2

  F10 = 1

  F11 = 1

  (¾)*100 = **75%**

References

Statistics How To. (2016). Jaccard Index / Similarity Coefficient. Statistics How To. Available at: https://www.statisticshowto.com/jaccard-index/ [Accessed 25 January 2025].