Calculate Jaccard Coefficient

The Jaccard coefficient, also known as the Jaccard similarity or Jaccard index, is a statistic used for gauging the similarity and diversity of sample sets. It measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. This is applicable for asymmetric data, where we only care for the presence of binary attributes. An asymmetric binary attribute is one in which outcomes are not of equal importance.

Jaccard= *f*01 + *f*11​​ / *f*01​+*f*10​+*f*11​

So let **Y & P** = 1; ***N & A*** = 0

**Jack, Mary**

Jack = {M, Fever, Negative, Test-1p, Test-2n, Test-3n, Test4a}

Mary = {F, Fever, Negative, Test-1p, Test-2a, Test3p, Test4a}

Intersection (f11): Fever (Y), Test-1p

Union (f01 + f10 + f11): Fever (Y), Cough (both N, thus ignored), Test-1 (P), Test-2 (A for Mary), Test-3 (P for Mary), Test-4 (A for both, thus ignored)

Jack: [1, 0, 1, 0, 0, 0]

Mary: [1, 0, 1, 1, 1, 0]

Jaccard (Jack, Mary) = **1/3 or 0.33**

**Jack, Jim**

Jack = {M, Fever, Negative, Test-1p, Test-2n, Test-3n, Test4a}

Jim = {M, Fever, Negative, Test-1p, Test-2n, Test-3n. Test-4a}

Intersection: Fever(Y)

Union: Fever(Y), Cough(P for Jim), Test-1 (P for Jack), Test-2 (both N thus ignored), Test-3 (both N thus ifnored), Test-4(A for both thus ignored)

Jaccard (Jack, Jim) = **2/3 or 0.67**

**Jim, Mary**

Jim = {M, Fever, Negative, Test-1p, Test-2n, Test-3n. Test-4a}

Mary = {F, Fever, Negative, Test-1p, Test-2a, Test3p, Test4a}

Intersection:

Union:

Jaccard (Jim, Mary) = **3/4 or 0.75**