Shelf Review Substitutability Score

Step 1 - define Inputs

$HHDs(P_1)$ = # of households buying product P_1

$HHDs(P_2)$ = # of households buying product P_2

$HHDs(P_1P_2)$ = # of households buying both product P_1 and product P_2

$HHDs(All)$ = # of households shopping the category

$HHDs(P_1P_2$ distinct baskets$)$ = # of households who buy both P_1 and P_2 but only in distinct baskets - so if household at any point in time (during the time period) buys P_1 and P_2 in the same basket then the household is not counted

Step 2 - calculate the expected number of households buying P_1 and P_2 in the time period -* expected household given indepedence of P_1 and P_2

$$E\left[HHDs(P_1P_2)\right] = \frac{HHDs(P_1) * HHDs(P_2)}{HHDs(All)}$$

to see why this is the expected number of households buying P_1 and P_2 given independence of P_1 and P_2: Multiply each side by $1/HHDs(All)$

$$\frac{E\left[HHDs(P_1P_2)\right]}{HHDs(All)} = \frac{HHDs(P_1) * HHDs(P_2)}{HHDs(All) * HHDs(All)}$$

Reduce household counts into probabilities

$$P(P_1P_2) = P(P_1) * P(P_2)$$

From Bayes theorem:

$P(A \cap B) = P(A|B) * P(B) = P(B|A) * P(A)$

$P(HHDs(P_1)) * P(HHDs(P_2)) = P(HHDs(P_1P_2))$

only if P_1 and P_2 are independent

Step 3 - caculate the expected # buying products P_1 and P_2 in distinct baskets

calculate the average percentage of household buying any two products in a category in distincts baskets

i) For each product pair calculate the proportion of households buying P_1 and P_2 that are purchased in distinct baskets

$$\%\,HHDs(P_1P_2 \text{ distinct baskets}) = \frac{HHDs(P_1P_2 \text{ distinct baskets})}{HHD_1(P_1P_2)}$$

ii) Calculate the average proportion in the category

$$\text{average distinct basket proportion} = \text{mean}(\%\,HHDs(P_iP_j \text{ distinct baskets}))$$

across all product pairs

iii) $E[HHDs(P_1P_2 \text{ different baskets})] = E[HHD_s(P_1P_2)] * \text{mean}(\%\,HHD_s(P_iP_j \text{ different baskets}))$

$E[HHDs(P_1P_2 \text{ different basket})]$ is the expected number of households purchasing P_1 and P_2, assuming independence of P_1 and P_2 mutipled by the average proportion of households that only buy P_1 and P_2 in distinct baskets so the resulting Expectation is the expected # of households buy P_1 and P_2 adjusted by the average proportion of distinct baskets in the commodity. This # will be used as the base line to compare actual distinct basket product pair household counts

Step 4 - calculate a partial index for each product pair

$$\text{Partial Index}(P_1P_2) = \frac{HHDs(P_1P_2 \text{ different baskets})}{E[HHDs(P_1P_2 \text{ different baskets})]}$$

Step 5 - calculate Substitutability Index for each product pair

i) calculate the median for all partial indices

ii) $$\text{Sub index}(P_1P_2) = \frac{\text{Partial Index}(P_1P_2)}{\text{Median}(\text{Partial Index}(P_iP_j))}$$

Step 6 - Calculate the chi squared statistic for each over indexing product pair

$$\chi^2 = \frac{(X - M)^2}{M}$$

why calculate sub index if we don't use it?

We exclude any product pair with partial index less than or equal to one (because less than one can not be a substitute)

$$\text{Chi Stat}(P_1 P_2) = \frac{\left(\text{HHDs}(P_1 P_2 \text{ different baskets}) - E\left[\text{HHDs}(P_1 P_2 \text{ different baskets})\right]\right)^2}{E\left[\text{HHDs}(P_1 P_2 \text{ different baskets})\right]}$$

Step 7 - calculate the Chi square percentage for P_1 with each potential substitute (only products that have partial index greater than one)

maybe uses the sub index

$$\%\text{ chi}(P_1 P_2) = \frac{\text{Chi-Stat}(P_1 P_2)}{\Sigma \text{ all Chi stat for P1 subs}}$$

So for example if products P_2, P_3, and P_4 were the over indexing substitutes for product P_1 then the %chi for P_1 with P_2 would be

$$\%\text{Chi}(P_1 P_2) = \frac{\text{Chistat}(P_1 P_2)}{\text{Chistat}(P_1 P_2) + \text{Chistat}(P_1 P_3) + \text{ChiStat}(P_1 P_4)}$$

Substitute pairs are determined by Step 5 in the sub index and Step 7 attemps to measure the statistical strength of the relationship between substitute pairs using the chi-squared statistic