

Finding Derivations: Significant

- A contrast-set for which at least two groups differ in their support is called **Significant**.
- Perform an statistical test (chi-square) for the contrast set:
- Null hypothesis: “The support for the contrast-set is the same across all groups”
- We build a 2 X k contingency table, k is the number of groups
- Compute the χ^2 statistics and check value in the chi-square distribution
- Must be less than a threshold α . (Typically, $\alpha = 0.05$)

c1: “*Product = burger \wedge Store = Fastfood*”

	Inner City	Industrious	Countryside
c1	944	776	172
\neg c1	2745	1851	4597

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \longrightarrow \text{Expected values}$$

$$E_{ij} = \frac{\sum_{i=1}^2 O_{ij} \sum_{j=1}^k O_{ij}}{N} \quad \longrightarrow \text{Observed values}$$



Consumer
Data
Research
Centre

An ESRC Data
Investment

KANTAR W^{ORLD}PANEL

Finding Derivations: Surprising

- A contrast-set for is considered to be **surprising** if the support of the contrast set in each group is different from its specialised nodes

Example

For the contrast set c1: “*Product = burger \wedge Store = Fastfood*”

$\text{support}(\text{Store} = \text{Fastfood}) = 13\%$

$\text{support}(\text{Product} = \text{burger} \wedge \text{Store} = \text{Fastfood}) = 12\%$

Subset support ratio = $\frac{12}{13} \approx 0.92$