Note: If $r = c = 2$ we have a 2×2 table, and another **exact** test can be used which does not rely on the large sample approximation

Fisher's Exact Test

- another test for independence of assignment of the row and column factor levels
- test statistic and null distribution are complicated (based on the **hypergeometric distribution**)
- SPSS computes test statistic and p-value.

Example: Juvenile Delinquency and Spectacle Wearing.

Is there any association between the two factors ?

$$A \quad : \quad \text{Spectacle Wearing (Yes/No)}$$
$$B \quad : \quad \text{Juvenile Delinquent (Yes/No)}$$

		Delinquent		
		Yes	No	$n_{i.}$
Spectacles	Yes	1	5	6
	No	8	2	10
	$n_{.j}$	9	7	16

Example: Juvenile Delinquency and Spectacle Wearing.

Chi-squared Test:

$$X^2 = 6.112$$

Compare with Chi-squared$((r-1)(c-1))$ = Chi-squared(1); we have

$$\text{Chi-squared}_{0.05}(1) = 3.841$$

and a p-value of 0.013. Therefore we **reject** H_0.

Fisher's Exact Test: p-value is 0.035 (1-sided) or 0.024 (2-sided).

Thus we reject H_0 and we have evidence of association between the factors.

Case-Control Studies

A **case-control** study is an observational study where participants are selected for the study with regard to their **disease status**.

- a sample of **cases** (disease sufferers)
- a sample of **controls** (healthy patients)

We investigate the possible association between disease status and a factor that takes two levels. A 2×2 table of counts is formed for all combinations of disease status/factor level.

Example: BCG Vaccination and Leprosy.

Disease Status : Leprosy Sufferer (Yes/No)

Factor : Vaccination Scar (Yes/No)

		Disease Status: Case	Disease Status: Control	
		Yes	No	$n_{i.}$
Scar	Yes	101	554	655
	No	159	446	605
	$n_{.j}$	260	1000	1260

Is there an association ? Does vaccination induce leprosy ?

The Chi-squared test is potentially not valid here because of the design. An alternative test statistic is based on the **odds ratio**

$$\text{O.R.} = \frac{n_{11} n_{22}}{n_{12} n_{21}} = \widehat{\psi}$$

say. The test statistic is

$$Z = \frac{\log \widehat{\psi}}{\text{s.e.}(\log \widehat{\psi})}$$

where

$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

That is,

$$Z = \frac{\log n_{11} + \log n_{22} - \log n_{12} - \log n_{21}}{\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}}$$

Under

$$H_0 \ : \ \text{No association between factor and disease status}$$

it follows that

$$Z \dot{\sim} N(0,1)$$

Here log means ln or natural log.

Example: BCG Vaccination and Leprosy.

$$n_{11} = 101, n_{12} = 554, n_{21} = 159, n_{22} = 446$$

Therefore

$$\widehat{\psi} = \frac{n_{11} n_{22}}{n_{12} n_{21}} = 0.511 \qquad \log \widehat{\psi} = -0.671$$

and

$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} = 0.142$$

so

$$Z = \frac{-0.671}{0.142} = -4.717$$

For a text at $\alpha = 0.05$, the two-sided critical value is ± 1.96, so we

Reject H_0.

Example: Smoking and Lung Cancer.

$$n_{11} = 647, n_{12} = 622, n_{21} = 2, n_{22} = 27$$

Therefore

$$\log \widehat{\psi} = \log \frac{647 \times 27}{2 \times 622} = 2.642$$

and

$$\text{s.e.}(\log \widehat{\psi}) = \sqrt{\frac{1}{647} + \frac{1}{2} + \frac{1}{622} + \frac{1}{27}} = 0.735$$

so

$$Z = \frac{2.642}{0.735} = 3.590$$

For a text at $\alpha = 0.05$, the two-sided critical value is ± 1.96, so we

Reject H_0

and report evidence for association.

Single Population Tests

We seek non-parametric or distribution-free tests for hypotheses relating to single samples, the equivalents of one-sample Z- or T-tests, which rely on the **normality** of the samples.

Normally these tests are formulated in terms of **ranks** of the data to give

Rank Tests

For example, if the data are

0.24 3.16 1.97 2.10 0.92

we sort them into **ascending** order, and assign ranks in order

	0.24	0.92	1.97	2.10	3.16
Rank	1	2	3	4	5

The tests depend on the behaviour of statistics computed in terms of the ranks, and rely on a **large sample** justification.