Analysis of Categorical Data

The Goodness of Fit Test

Goodness of Fit

- Purpose: to examine the fit of a one-way frequency distribution for X, the variable of interest, to the distribution expected under the null hypothesis. Example.

- Notations:
  - ♦ O_i - the observed value at level i;
  - ♦ E_i - $n\pi_i$, the expected value at level i under the null hypothesis;
  - ♦ n - total number of observations;
  - ♦ π_i - the population prob. of level i.

Measurements of Goodness of Fit

- $\sum (O_i - E_i)$

- $\sum (O_i - E_i)^2$

- $\sum (O_i - E_i)^2 / E_i$

Example1

Example2

of Intervals Used for Continuous Data

Cochran, 1952

Table 1: Guideline for the number of intervals to be used with a continuous variable

Sample Size	# of Intervals	Sample Size	# of Intervals
200	15	400	20
600	24	800	27
1000	30	1500	35
2000	39		