AIM Algorithms

The Attribute Importance (AIM) procedure performs tests to find out if the groups are homogeneous.

Notation

The following notation is used throughout this chapter unless otherwise stated:

G	Number of groups.
C	Number of categories in the categorical variable.
n_{ij}	Number of cases in the *j*th category in the *i*th group, $i = 1, \ldots, G$ and $j = 1, \ldots, C$. Assume that $n_{ij} \geq 0$.
n_i	Number of cases in the i-th group. $n_i = \Sigma_{j=1}^{C} n_{ij}$
n	Overall number of cases. $n = \Sigma_{i=1}^{G} n_i$. Assume $n>0$.
p_j	Overall proportion of cases in the *j*th category. $p_j = \frac{1}{n}\Sigma_{i=1}^{G} n_{ij}$
$\overline{x}_i$	Mean of the continuous variable in the *i*th group.
s_i	Standard deviation of the continuous variable in the *i*th group. Assume that $s_i \geq 0$.
$\overline{x}$	Overall mean of the continuous variable. $\overline{x} = \frac{1}{n}\sum_{i=1}^{G} n_i x_i$

Test of Homogeneity of Proportions

This test is performed only for categorical variables. The null hypothesis is that the proportion of cases in the categories in the *i*th group is the same as the overall proportion. If $C > 1$, the Chi-square statistic for the *i*th group is computed as follows:

$$\chi^2 = \sum_{j=1}^{C} \frac{(n_{ij} - n_i p_j)^2}{n_i p_j}$$

The degrees of freedom is $C-1$. The significance is the probability that a Chi-square random variate with this degrees of freedom will have a value greater than the χ^2 statistic.

If $C<1$, the Chi-square statistic is always 0 with zero degrees of freedom, and the significance value is undefined.

Test of Equality of Means

This test is performed only for continuous variables. The null hypothesis is that the mean (of a continuous variable) in the *i*th group is the same as the overall mean. If $n_i > 1$ and $s_i > 0$, the Student's *t* statistic for the *i*th group is computed as follows:

$$t = \frac{(\overline{x}_i - \overline{x})}{s_i/\sqrt{n_i}}$$

The degrees of freedom is $n_i - 1$. The significance is the probability that a Student's *t* random variate with this degrees of freedom will have a value greater than the *t* statistic.

When $n_i > 1$ but $s_i = 0$, this implies that the continuous variable is constant in the *i*th group. In this case, the Student's *t* statistic is infinity with positive degrees of freedom $n_i - 1$, and the significance value is zero.

If $n_i \leq 1$, then s_i is undefined. In this case, the Student's *t* statistic is undefined, the degrees of freedom is 0, and the significance value is undefined.

Graphical Display of Significance

Since significance values are often very small numbers, the negative common logarithm ($-\log_{10}$) of significance values are displayed instead in the bar charts.