Comparison of Proportions

Summary

This procedure is designed to compare the observed proportions of an event in k samples. It performs a chi-squared test to determine whether or not there are significant differences amongst the proportions. It also performs an analysis of means (ANOM) to determine which proportions are significantly different than the overall average.

## Sample StatFolio: *compareprops.sgp*

Sample Data:

The file *houses.sgd* contains the results of a survey performed on k = 9 different streets within a town. On each street, the number of houses with and without air conditioners was recorded. The data, from Guttman, Wilks and Hunter (1982), is shown below:

Street	With Air Conditioners	Without Air Conditioners
1	5	18
2	8	35
3	18	25
4	3	38
5	17	24
6	11	31
7	25	17
8	19	20
9	18	18

Data Input

The data required by this procedure is a numeric column containing the *k* proportions.
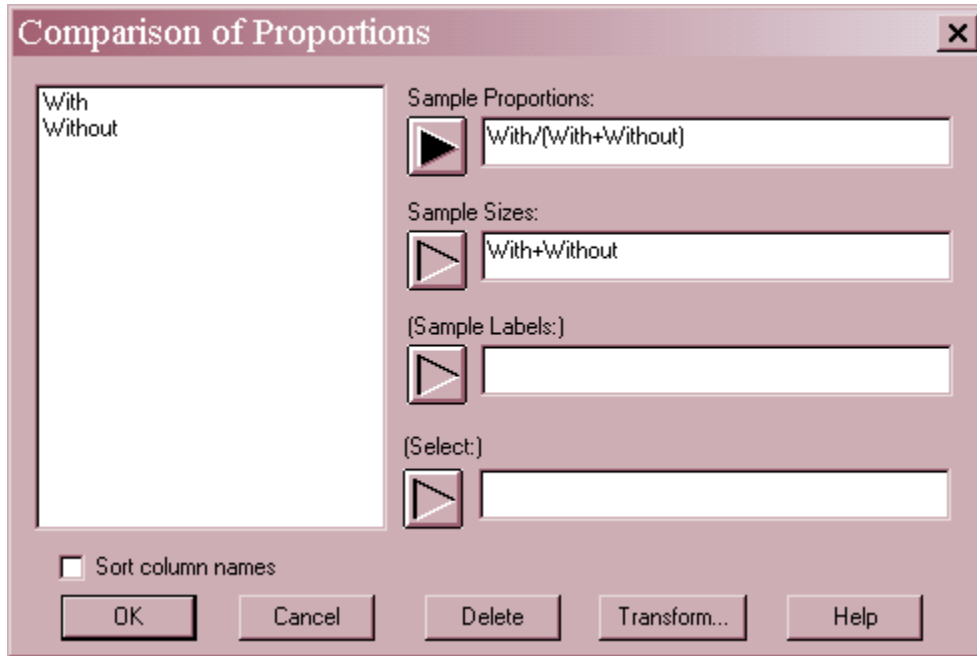
- **Sample Proportions:** a numeric column containing the proportions p_i in each sample. Note that an expression can be used to calculate the proportions "on-the fly", as in the example above.

- **Sample Sizes:** a numeric column containing the sample sizes n_i.

- **Sample Labels:** optional labels for each sample.

- **Select:** subset selection.

Analysis Summary

This pane summarizes the data.

Comparison of Proportions - With/(With+Without)
Data variables: With/(With+Without)
Number of samples = 9
Average sample size = 38.8889
Mean proportion = 0.354286

Chi-Squared Test

Chi-squared	Df	P-Value
41.38	8	0.0000

Included are:

- **Number of samples:** the number of samples *k*.

- **Average sample size:** the average sample size

$$\overline{n} = \frac{\sum_{i=1}^{k} n_i}{k} \tag{1}$$

- **Mean Proportion:** the weighted average of the proportions:

$$\overline{p} = \frac{\sum_{i=1}^{k} n_i p_i}{\sum_{i=1}^{k} n_i} \tag{2}$$

- **Chi-Squared Test:** the result of a chi-squared test of the hypotheses:

   *Null hypothesis*: all proportions are equal

   Alternative hypothesis: not all proportions are equal

   This is the same test that would have been performed has the data been analyzed using the *Contingency Tables* procedure. Small P-values for the test (less than 0.05 is operating at the 5% significance level) lead to a rejection of the hypothesis that all proportions are equal.

For the sample data, the P-Value is very small, leading to the conclusion that the proportions of houses with air conditioners is *not* the same on all 9 streets.

Analysis of Means (ANOM) Plot

To compare the sample proportions, an analysis of means plot is created:
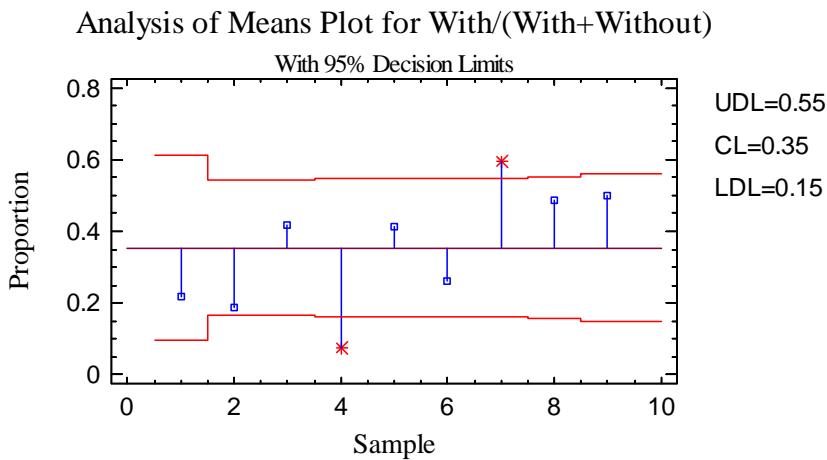
The plot shows each sample proportion, with a vertical line drawn from to centerline, located at the mean proportion $\overline{p}$. Decision limits are drawn above and below the centerline at

$$\overline{p} \pm h \sqrt{\frac{\overline{p}(1-\overline{p})(k-1)}{n_i k}} \qquad (3)$$

where h is the critical value of a multivariate t distribution for k groups with a very large number of degrees of freedom.

Any proportions beyond the decision limits are statistically significantly different than $\overline{p}$ at the selected confidence level. In this case, the average proportion of houses with air conditioners $\overline{p} = 0.354$. The proportion of houses with air conditioners is significantly lower than average on street 4 and significantly higher than average on street 7.

Analysis of Means (ANOM) Table

The ANOM table shows the comparison of proportions in tabular form:

ANOM Report

95% Decision Limits
UDL = 0.55
Centerline = 0.35
LDL = 0.15
Number of samples beyond limits = 2

Sample	Size	Proportion
1	23	0.217391
2	43	0.186047
3	43	0.418605
4	41	0.0731707*
5	41	0.414634
6	42	0.261905
7	42	0.595238*
8	39	0.487179
9	36	0.5

* = Beyond Limits

Any sample proportions outside the decision limits are indicated by an asterisk (*).

Pane Options
The ANOM plot and ANOM report share the same options:
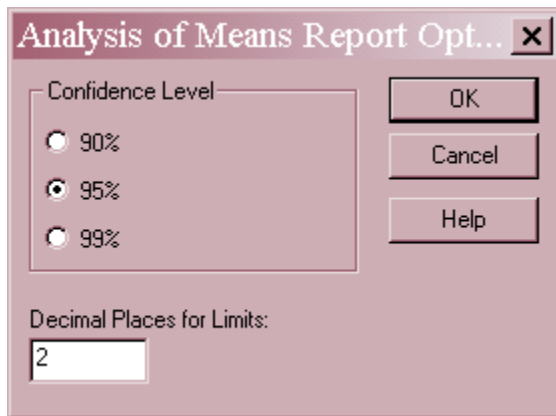
- **Confidence Level:** percentage used to define the decision limits.

- **Decimal Places for Limits:** number of decimal places used to display the decision limits.