Comparison of Rates

Summary

This procedure is designed to compare the observed rates of an event amongst *k* samples. It performs a dispersion test and a likelihood ratio test to determine whether or not there are significant differences amongst the rates. It also performs an analysis of means (ANOM) to determine which rates are significantly different than the overall average.

Sample StatFolio: *comparerates.sgp*

Sample Data:

The file *aircraft.sgd* contains the data on the failure rates of air conditioning equipment on *k* = 13 aircraft. The data, from Cox and Lewis (1966), are shown below:

Aircraft	Failures	Flying Hours
1	6	493
2	23	2201
3	29	2422
4	15	1819
5	14	1832
6	30	1788
7	27	2074
8	24	1539
9	9	1800
10	6	639
11	2	623
12	12	1297
13	16	1312

Data Input

The data required by this procedure is a numeric column containing the *k* rates.
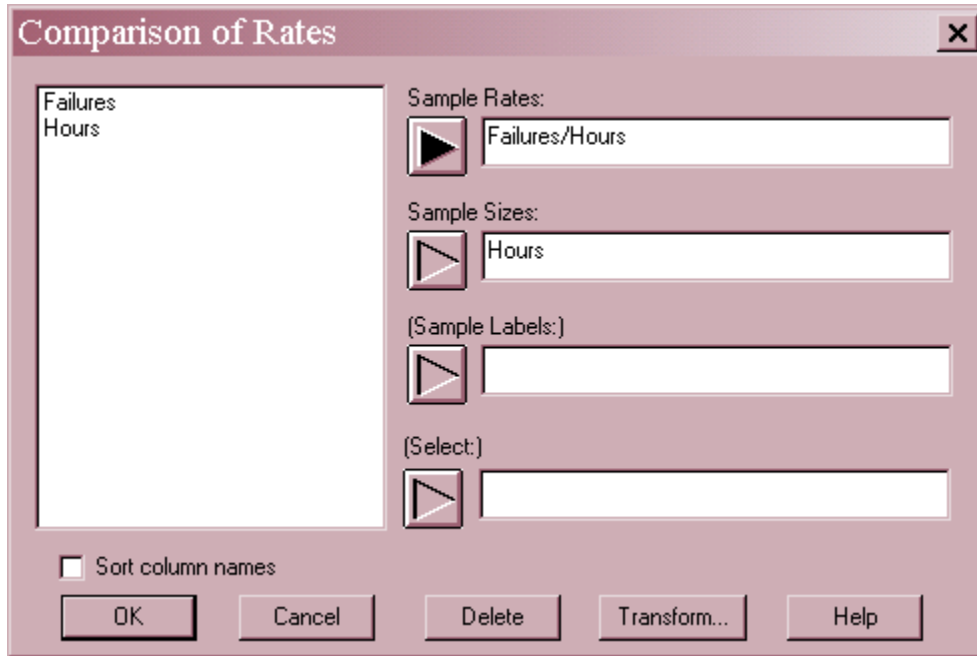
- **Sample Rates:** a numeric column containing the rates $\hat{\lambda}_i$ in each sample. Note that an expression can be used to calculate the rates "on-the fly", as in the example above.

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

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

- **Select:** subset selection.

Analysis Summary

This pane summarizes the data.

Comparison of Rates - Failures/Hours
Data variables: Failures/Hours
Number of samples = 13
Mean rate = 0.0107364

Tests for Equality of Rates

Test	Chi-Squared	Df	P-Value
Dispersion index	23.05	12	0.0273
Likelihood ratio	24.57	12	0.0170

Included are:

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

- **Mean Rate:** the weighted average of the rates:

$$\hat{\lambda} = \frac{\sum_{i=1}^{k} T_i \hat{\lambda}_i}{\sum_{i=1}^{k} T_i} \tag{1}$$

- **Dispersion Index Test:** This test compares the variance of the sample counts

$$n_i = \hat{\lambda}_i T_i \tag{2}$$

to their mean, which should be the same if the data come from a Poisson distribution. It tests the hypotheses:

Null hypothesis: all rates are equal

Alternative hypothesis: not all rates are equal

The test statistic is calculated by

$$d = \sum_{i=1}^{k} \frac{\left(\hat{\lambda}_i T_i - \hat{\lambda} T_i\right)^2}{\hat{\lambda} T_i} \tag{3}$$

which is compared to a chi-squared distribution with $k - 1$ degrees of freedom. 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 rates are equal.

- **Likelihood Ratio Test:** This test is similar to the dispersion index test. It is calculated by

$$H = 2\left\{ \sum_{i=1}^{k} n_i \log(n_i / T_i) - \left(\sum_{i=1}^{k} n_j\right) \log\left(\sum_{i=1}^{k} n_i / \sum_{i=1}^{k} T_i\right)\right\} \tag{4}$$

and is also compared to a chi-squared distribution with $k - 1$ degrees of freedom. Small P-values lead to a rejection of the hypothesis that all rates are equal.

For the sample data, both P-Values are below 0.05, leading to the conclusion that the rates differ significantly among the 13 aircraft (at the 5% significance level).

Analysis of Means (ANOM) Plot

To compare the sample rates, an analysis of means plot is created:

Analysis of Means Plot for Failures/Hours

With 95% Decision Limits

UDL=0.0181
CL=0.0107
LDL=0.0034
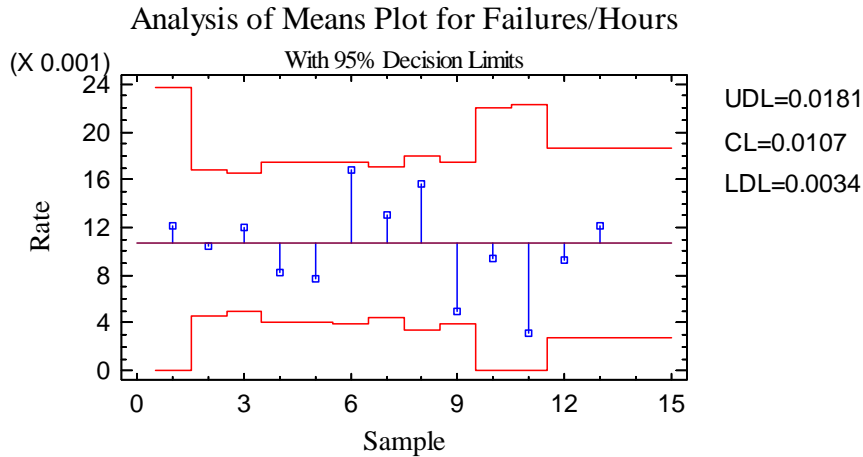
The plot shows each sample rate, with a vertical line drawn from to the centerline, located at the mean rate $\hat{\lambda}$. Decision limits are drawn above and below the centerline at

$$\hat{\lambda} \pm h\sqrt{\frac{\hat{\lambda}(k-1)}{kT_i}} \tag{5}$$

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

Any rates beyond the decision limits are statistically significantly different than the mean rate at the selected confidence level. In this case, no individual aircraft appears to be significantly different than the mean, despite the significant tests results.

Analysis of Means (ANOM) Table

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

ANOM Report

95% Decision Limits
UDL = 0.0181
Centerline = 0.0107
LDL = 0.0034
Number of samples beyond limits = 0

Sample	Rate
1	0.0121704
2	0.0104498
3	0.0119736
4	0.00824629
5	0.00764192
6	0.0167785
7	0.0130183
8	0.0155945
9	0.005
10	0.00938967
11	0.00321027
12	0.00925212
13	0.0121951

* = Beyond Limits

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

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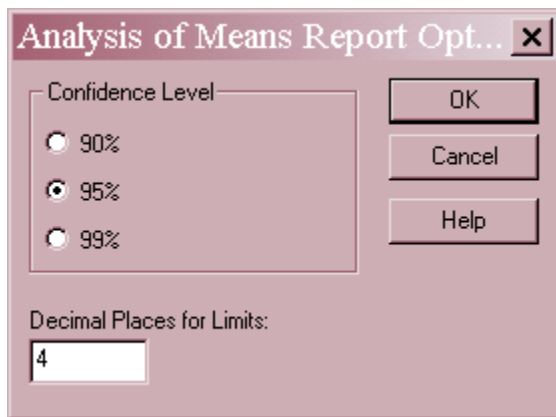
- **Confidence Level:** percentage used to define the decision limits.

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