Hypothesis Testing

In This Issue

1. Discover the reason for using hypothesis testing.
2. Learn how hypothesis testing and bootstrap simulation differ.

"How can you determine if the means and variances of two distributions are statistically identical or statistically different from one another?"

Theory

A hypothesis test is performed when testing the means and variances of two distributions to determine if they are statistically identical or statistically different from one another; that is, to see if the differences between the means and variances of two different forecasts that occur are based on random chance or if they are, in fact, statistically significantly different from one another.

This analysis is related to bootstrap simulation with several differences. Classical hypothesis testing uses mathematical models and is based on theoretical distributions. This means that the precision and power of the test is higher than bootstrap simulation's empirically based method of simulating a simulation and letting the data tell the story. However, classical hypothesis tests are only applicable for testing two distributions' means and variances (and by extension, standard deviations) to see if they are statistically identical or different. In contrast, nonparametric bootstrap simulation can be used to test for any distributional statistics, making it more useful, but the drawback is its lower testing power. **Risk Simulator** provides both techniques from which to choose.

Procedure

- Run a simulation.
- Select *Risk Simulator | Tools | Hypothesis Testing*.
- Select the two forecasts to test, select the type of hypothesis test you wish to run, and click OK (Figure 1).

Figure 1. Hypothesis Testing
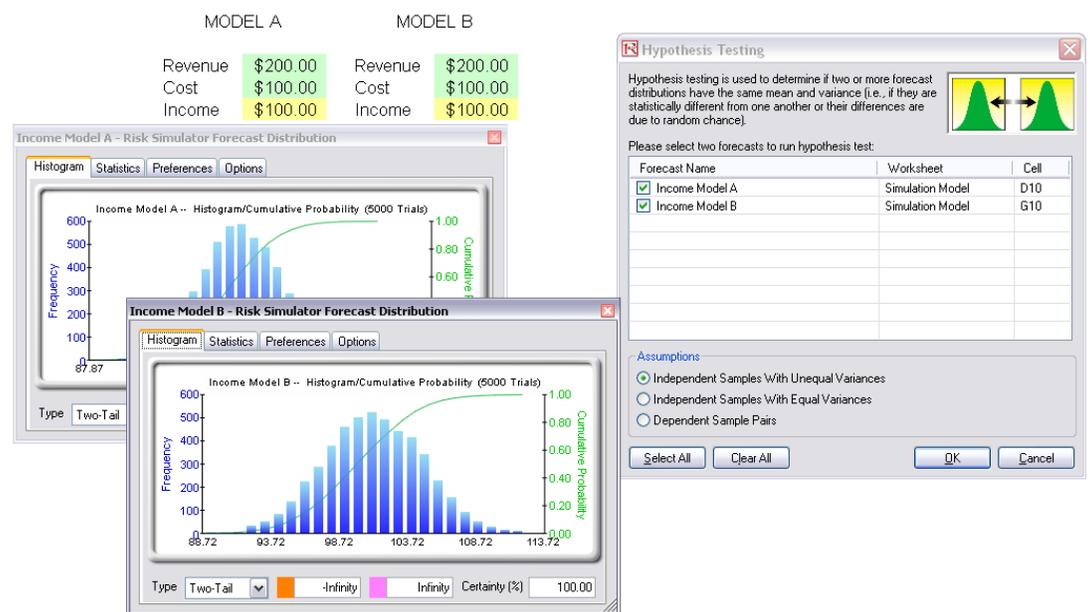
Contact Us

Real Options Valuation, Inc.

4101F Dublin Blvd., Ste. 425, Dublin, California 94568 U.S.A.

admin@realoptionsvaluation.com
www.realoptionsvaluation.com
www.rovusa.com

Results Interpretation

A two-tailed hypothesis test is performed on the null hypothesis (H_o) such that the two variables' population means are statistically identical to one another. The alternative hypothesis (H_a) is such that the population means are statistically different from one another. If the calculated p-values are less than or equal to 0.01, 0.05, or 0.10 alpha test levels, it means that the null hypothesis is rejected, which implies that the forecast means are statistically significantly different at the 1%, 5%, and 10% significance levels. If the null hypothesis is not rejected when the p-values are high, the means of the two forecast distributions are statistically similar to one another. The same analysis is performed on variances of two forecasts at a time using the pairwise F-test. If the p-values are small, then the variances (and standard deviations) are statistically different from one another; otherwise, for large p-values, the variances are statistically identical to one another. The example file used was *Hypothesis Testing and Bootstrap Simulation*.

Hypothesis Test on the Means and Variances of Two Forecasts

Statistical Summary

A hypothesis test is performed when testing the means and variances of two distributions to determine if they are statistically identical or statistically different from one another. That is, to see if the differences between two means and two variances that occur are based on random chance or they are in fact different from one another. The two-variable t-test with unequal variances (the population variance of forecast 1 is expected to be different from the population variance of forecast 2) is appropriate when the forecast distributions are from different populations (e.g., data collected from two different geographical locations, two different operating business units, and so forth). The two-variable t-test with equal variances (the population variance of forecast 1 is expected to be equal to the population variance of forecast 2) is appropriate when the forecast distributions are from similar populations (e.g., data collected from two different engine designs with similar specifications, and so forth). The paired dependent two-variable t-test is appropriate when the forecast distributions are from similar populations (e.g., data collected from the same group of customers but on different occasions, and so forth).

A two-tailed hypothesis test is performed on the null hypothesis Ho such that the two variables' population means are statistically identical to one another. The alternative hypothesis is that the population means are statistically different from one another. If the calculated p-values are less than or equal to 0.01, 0.05, or 0.10, this means that the hypothesis is rejected, which implies that the forecast means are statistically significantly different at the 1%, 5% and 10% significance levels. If the null hypothesis is not rejected when the p-values are high, the means of the two forecast distributions are statistically similar to one another. The same analysis is performed on variances of two forecasts at a time using the pairwise F-Test. If the p-values are small, then the variances (and standard deviations) are statistically different from one another, otherwise, for large p-values, the variances are statistically identical to one another.

Result

Hypothesis Test Assumption:	Unequal Variances:
Computed t-statistic:	-0.32947
P-value for t-statistic:	0.74181
Computed F-statistic:	1.026723
P-value for F-statistic:	0.351212

Figure 2. Hypothesis Testing Results

Notes

The two-variable t-test with *unequal* variances (the population variance of forecast 1 is expected to be different from the population variance of forecast 2) is appropriate when the forecast distributions are from different populations (e.g., data collected from two different geographical locations, or two different operating business units, and so forth). The two-variable t-test with *equal* variances (the population variance of forecast 1 is expected to be equal to the population variance of forecast 2) is appropriate when the forecast distributions are from similar populations (e.g., data collected from two different engine designs with similar specifications, and so forth). The paired dependent two-variable t-test is appropriate when the forecast distributions are from exactly the same population and subjects (e.g., data collected from the same group of patients before an experimental drug was used and after the drug was applied, and so forth).