Overview of Statistical Hypothesis Testing

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps.

STEP 1: Formulate the null hypothesis H_0 (typically that the observations are the result of pure chance) and the alternative hypothesis H_a (typically that the observations show a real effect).

The hypotheses are about a population parameter P.

For instance we can test that a certain population average μ is equal to a certain value μ_0, or we can test that a proportion is equal to a certain value p_0. In regression analysis we test that a parameter β is equal to zero. So in all these cases for the generic population parameter P, we can write:

H0: $P = P_0$ (null value of interest)

Ha: $P \neq P_0$

STEP 2: Identify a test statistic that can be used to assess the truth of the null hypothesis. Each test has a different test statistic. The test statistic is computed from the sample.

The general expression of a test statistic is as follows:

$$\text{Test statistic} = \frac{\text{sample estimate of parameter - } P_0}{\text{standard error of estimate}}$$

For each test statistic we can specify its distribution under the assumption that the null hypothesis is true, and we can use this distribution to compute the probability of observing test statistic values. This will be used in step 3 to compute the p-value.

EXAMPLE: For a test on averages, the test statistic is the well-known: $t = \dfrac{\bar{x} - \mu_0}{st.dev./\sqrt{n}}$, where $\bar{x}$ is the sample estimate of the population average, and st.dev./sqrt(n) is the standard error of the estimate.

For a large sample size, the approximate distribution of the test statistic, if <u>the null hypothesis were true,</u> is the standard normal N(0,1). Thus under the null hypothesis

1. There is a high chance of observing a test statistic value close to zero.
2. And, the probability of observing a value of the test statistics t larger than 2 or smaller than -2 would be quite small (less than 0.05).

STEP 3: Compute the p-value, which is the probability of observing a value equal to the test statistic or even more extreme assuming that the null hypothesis is true. This p-value is computed using the distribution of the test-statistic if the null hypothesis were true. The smaller the P-value, the stronger the evidence against the null hypothesis.

STEP 4: Compare the p-value to an acceptable significance value alpha (where alpha is small and often =0.05 or 0.01). If $p \le$ alpha, then the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

Decision Rules:

- If p-value < 0.05, reject the null hypothesis at 5% level (test is significant), therefore alternative hypothesis is true.
- If p-value < 0.01, reject the null hypothesis at 1% level (test is highly significant), therefore alternative hypothesis is true.
- If p-value > 0.05, we cannot reject the null hypothesis at 5% level. Test is inconclusive. Data do not provide enough evidence to reject the null hypothesis.