[STAT-18] Nonparametric testing

Miguel-Angel Canela
Associate Professor, IESE Business School

Nonparametric testing

The tests of the preceding lectures are valid under normality assumptions and asymptotically valid without them. In general, a **nonparametric test** is one in which it is not assumed that the distribution of the variables involved is of a particular type. A trivial example of a nonparametric test, the **sign test**, based on the binomial distribution, is an alternative to the one-sample t test (and consequently to the paired data t test), with no distributional assumption.

Suppose a continuous distribution with median μ and a sample of size n, containing no zeros. The null is $H_0 : \mu = 0$ (for $\mu = \mu_0$, it suffices to subtract μ_0 and perform the test as presented here). Calling B^+ the number of positive observations and B^- that of negative observations, the test statistic is $B = \max(B^+, B^-)$. The test is based on the fact that, under the null, the probability of a positive result is 0.5. Then B^+ and B^- have a $\mathcal{B}(n, 0.5)$ distribution. The P-value is the double of the probability of the right tail associated to the actual value of B in $\mathcal{B}(n, 0.5)$ distribution. Some stat packages reportasymptotic P-values, derived from a norm al approximation.

¶ Zero observations are discarded in this test. This is not relevant as far as the continuity assumption, under which repetition is not expected, is tenable.

Example 1. For the `lwages` data set analyzed in lecture STAT-15, we find 443 cases in which the wages have been increased. The (two-tail) P-value gives about $P < 0.001$, consistent with the outcome of the t test. An asymptotic P-value can be based on the $\mathcal{N}(272.5, 272.5)$ distribution.

The Wilcoxon signed rank test

The **Wilcoxon signed rank test** is a second alternative to the one-sample t test. The distributional assumptions are the continuity and the symmetry with respect to the mean. I set, as in the preceding section, a null $\mu = 0$. The test statistic is obtained as follows:

- We sort the observations by absolute value. Let me assume first that there are no ties.
- We assign ranks. The first observation gets 1, the second 2, etc.
- Calling T^+ and T^- the sum of ranks of the positive and negative observations, respectively, we have $T^+ + T^- = n(n+1)/2$. In the version provided by Stata, the test statistic is T^+, but many textbooks use $T = \max(T^+, T^-)$ to simplify the use of the tables.

Under the null, T_+ has a symmetric (discrete) distribution, with mean and variance given by

$$\mathrm{E}[T_+] = \frac{n(n+1)}{4}, \qquad \mathrm{var}[T_+] = \frac{n(n+1)(2n+1)}{24}.$$

If there are ties in the absolute values, they get an average rank. The variance must be then corrected. Stat packages usually provide a correction for this case.To get exact significance levels for this test and those which follow, one should look at the corresponding tables or use a special package. What we usually get from generalist stat packages is an **asymptotic significance level**. The difference may be relevant for small sample experimental studies, but not the sample sizes that we usually find in Econometrics. In fact, the tables that we find in textbooks do no go beyond $n = 20$. Asymptotic P-values are based on a normal approximation with the mean and variance given by the above formulas.

Example 1 (continuation). For the `lwages` data set, we get $z = 15.9$ ($P < 0.001$).

The rank-sum test

The **Wilcoxon two-sample rank-sum test** is an alternative to the two-sample t test that only requires that the distributions compared are continuous and of the same type. There are two equivalent versions, that of Wilcoxon, that I present here, and that of Mann and Whitney, sometimes called **Mann-Whitney U test**.

More specifically, it is assumed here the PDF's compared are related by an equation

$$f_2(x) = f_1(x - \Delta).$$

$\Delta = \mu_1 - \mu_2$ is sometimes called the **treatment effect**. The null is $\Delta = 0$. The test applies to two independent samples of sizes n_1 and n_2, respectively. To simplify the notation, I assume here that $n_1 \le n_2$. The test statistic W is obtained as follows:

- The two samples are merged, and the the resulting sample (size $n_1 + n_2$) is sorted.
- We assign ranks to the observations, averaging ties.
- W is the sum of the ranks of the first sample.

Under the null, W has a symmetric (discrete) distribution, with

$$\mu = \frac{n_1\big(n_1 + n_2 + 1\big)}{2}, \qquad \sigma^2 = \frac{n_1 n_2\big(n_1 + n_2 + 1\big)}{12}.$$

As in the signed rank test, exact significance levels are usually extracted from tables, but only for small samples. For $n_2 > 10$, asymptotic levels are accepted.

Example 2. In the analysis of the `jobsat1` data set, to compare Chile and Mexico as in lecture STAT-16, we get $z = -2.224$ ($P = 0.026$), similar to the t test.

The Kruskal-Wallis test

The **Kruskal-Wallis test** is an extension of the rank-sum test to k samples, just as the one-way ANOVA F test is an extension of the two-sample t test. The assumptions are as in the rank-sum tests. The test statistic is obtained as follows:

- The samples are merged and the resulting sample is sorted, assigning ranks.
- The test statistic is

  $$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1),$$

  where R_i is the sum of the ranks of sample i and n is the total sample size ($n = n_1 + \cdots + n_k$).

For $n_i \ge 5$, the distribution of H can be approximated by a $\chi^2(k-1)$.

Example 2 (continuation). To compare the three countries of the `jobsat1` data set, we get $\chi^2(2) = 9.46$ ($P = 0.009$), with more significance than in lecture STAT-17.

Homework

A. Apply a nonparametric test to the `scapital` data set, and compare the result with that obtained in lecture STAT-16.

B. Apply a nonparametric test to the `jobsat2` data set, and compare the result with that obtained in lecture STAT-17.