Two-sample problem: Wilcoxon rank-sum test (similar to Mann-Whitney U test)

Let $X_1, X_2, \cdots, X_n$ be a random sample from population 1 and $Y_1, Y_2, \cdots, Y_m$ be a random sample from population 2. Let all the $n + m$ observations are at least ordinal, and independent of one another. We want to test the hypotheses:

H_0: The two distributions are identical.

H_1: The two distributions are different.

We assign ranks to all the $n + m$ observations treating them to be one set. The smallest value gets rank 1 and so on. If two or more values are same, each of them gets a rank that is the average of the ranks they would have received if they had differed slightly. For example, when the observations are 7, 8, 9, 9 and 11, the respective ranks are 1, 2, 3.5, 3.5 and 5.

Let $T =$ sum of the ranks of the first sample.

$$E(T) = \mu = \frac{n(m+n+1)}{2}$$

$$V(T) = \sigma^2 = \frac{mn(m+n+1)}{12}$$

For moderately large m and n (both being greater than 7 should be sufficient), the distribution of T is normal with the above mean and variance. We can calculate the p-value from the Z table.

Example (Wikipedia)

Aesop is dissatisfied with his classic experiment in which one tortoise was found to beat one hare in a race, and decides to carry out a significance test to discover whether the results could be extended to tortoises and hares in general. He collects a sample of 7 tortoises and 6 hares, and makes them all run his race at once. The order in which they reach the finishing post is as follows, writing T for a tortoise (please do not confuse with the test statistic T) and H for a hare:

T H H H H T H T T T T T H

Can you conclude that tortoises and hares are different in racing?

Solution

Let us rank the animals by the time they take to complete the course. Then, the sum of the ranks achieved by the tortoises,

$$T = 1 + 6 + 8 + 9 + 10 + 11 + 12 = 57.$$

Do the rest part yourself. Note that this particular problem is a two-tailed test.

The runs test for randomness

A basic assumption in many statistical methods is that the data set is a random sample from some population. Sometimes the data are not generated by a truly random process. They follow a trend or a type of cyclic pattern. We often need to test the null hypothesis that a given data set constitutes a random sample.

Let each of the data values be either a 0 or a failure 1. That is, we assume that each data value can be dichotomized as being either a success or a failure. Any consecutive sequence of 1's (or 0's) is called a 'run'. For example, the data set

1 0 0 1 1 1 0 0 1 0 1 1 1 1 0 1 0 0 0 0 1 1

contains 11 runs — 6 runs of 1 and 5 runs of 0.

Let a data set contain m 1's and n 0's, where $n + m = N$. Let R denote the number of runs in the data set. If the null hypothesis is true, we should have

$$E(R) = \mu = \frac{2mn}{m+n} + 1$$

$$V(R) = \sigma^2 = \frac{(\mu - 1)(\mu - 2)}{m + n - 1}$$

When m and n are both large, R will approximately follow a normal distribution with the mean and variance given above.

If $R < \mu$,

p-value $= 2\,\Phi(R - \mu)/\sigma)$.

If $R > \mu$,

p-value $= 2\,(1 - \Phi(R - \mu)/\sigma))$.

Example

The following is the result of the last 30 games played by an athletic team, with ‘W’ signifying a win and ‘L’ a loss.

W W W LW W LW W LW LW W LW W W W LW LW W W LW LW L

Are these data consistent with pure randomness?

Solution

Note that the data, which consist of 20 W’s and 10 L’s, contain 20 runs. Thus, we have $m = 20$, $n = 10$ and $R = 20$.

run=20,N=30

Do the rest part yourself.