STA 104 Applied Nonparametric Statistics

Chapter 3: Two-Sample Methods

Xiner Zhou

Department of Statistics, University of California, Davis

In this chapter the data consist of two random samples, a sample from the control population and an independent sample from the treatment population.

On the basis of these samples, we wish to investigate the presence of a treatment effect that results in a shift of location.

Table of contents

Two-Sample Permutation Test

We begin with a simple example. Suppose UCD is trying to decide whether to augment its traditional classroom instruction with hybrid mode. Seven students are selected for a trial. Four are randomly assigned to the new method of hybrid instruction, and the other three are given the traditional instruction. A test is given afterward to compare the two methods.

New Method	Smith (37), Lin (49), Neal (55), Zedillo (57)
Traditional Method	Johnson (23), Green (31), Zook (46)

A two-sample t-test:

A two-sample t-test of the null hypothesis of no difference between the two methods versus the one-sided alternative hypothesis that the mean of the new method is greater than the mean of the traditional method gives $t = 2.08$ and $p = .046$.

Thus, we conclude that the new method produces a significantly higher mean test score than the traditional method at the 5% level of significance.

A two-sample t-test:

The application of the t-test comes at a price by requiring assumptions:

- the observations are independent;
- the populations have normal distributions $\Rightarrow$ null distribution of test statsitics
- the variances of the two populations are the same.

In this case, there is no guarantee that the assumptions of the t-test are met.

If we apply the t-test anyway, we run the risk of misstating the p-value of the statistical test and therefore of declaring a result to be statistically significant when it is not.

Intuition of permutation test

Key question: What is null distribution without any parametric assumptions like t-test?

Intuition of permutation test

If there is no difference between the two methods, then all data sets obtained by randomly assigning four of these scores to the new method and the other three to the traditional method would have an equal chance of being observed in the study. There are

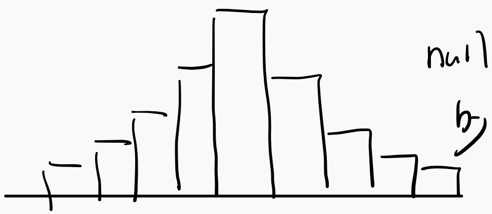
$$\binom{7}{4} = \frac{7!}{4!3!} = 35 \quad \text{permutations}$$

such two-sample data sets. Most data sets among the 35 have both large and small scores assigned to each treatment. These are the types of data sets we would expect to observe if the two treatments were not different (equally effective).

Any appropriate statistic that quantifies the difference between groups can work:

t-test statistic

	New Method	Traditional Method	Difference between means
1	46 49 55 57	23 31 37	21.4
2*	37 49 55 57	23 31 46	16.2
3	37 46 55 57	23 31 49	14.4
4	37 46 49 57	23 31 55	10.9
5	37 46 49 55	23 31 57	9.8
6	31 49 55 57	23 37 46	12.7
7	31 46 55 57	23 37 49	10.9
8	31 46 49 57	23 37 55	7.4
9	31 46 49 55	23 37 57	6.3
10	31 37 55 57	23 46 49	5.7
11	31 37 49 57	23 46 55	2.2
12	31 37 49 55	23 46 57	1.0
13	31 37 46 57	23 49 55	0.4
14	31 37 46 55	23 49 57	-0.8
15	31 37 46 49	23 55 57	-4.3
16	23 49 55 57	31 37 46	8.0
17	23 46 55 57	31 37 49	6.3
18	23 46 49 57	31 37 55	2.8
19	23 46 49 55	31 37 57	1.6
20	23 37 55 57	31 46 49	1.0
21	23 37 49 57	31 46 55	-2.5
22	23 37 49 55	31 46 57	-3.7
23	23 37 46 57	31 37 55	-4.3
24	23 37 46 55	31 49 57	-5.4
25	23 37 46 49	31 55 57	-8.9
26	23 31 55 57	37 46 49	-2.5
27	23 31 49 57	37 46 55	-6.0
28	23 31 49 55	37 46 57	-7.2
29	23 31 46 57	37 49 55	-7.8
30	23 31 46 55	37 49 57	-8.9
31	23 31 46 49	37 55 57	-12.4
32	23 31 37 57	46 49 55	-13.0
33	23 31 37 55	46 49 57	-14.2
34	23 31 37 49	46 55 57	-17.7
35	23 31 37 46	49 55 57	-19.4



null distribution by permutation

The procedure we have just described is called a two-sample permutation test, since it is based on permuting the observations among two groups in the original sample.

The distribution of the 35 differences of means is called the permutation distribution for the difference between two means.

- The permutation principle states that the permutation distribution is an appropriate reference distribution for determining the p-value of a test and deciding whether or not a statistical test is statistically significant.
- Under nonparametric permutation approach, the researcher is free to choose a statistic that he or she feels best describes the difference between the two groups and then use the permutation approach to determine whether or not the statistic is significant.

Prior to the computer age, the practical use of permutations tests was limited by prohibitive computations.

For instance, if two treatments each have 8 observations, then the number of possible two-sample data sets that can be obtained by permuting the 16 observations, 8 to a treatment, is

$$\binom{16}{8} = 12,870$$

If just two more observations are added to each treatment, then this number increases more than tenfold to

$$\binom{20}{10} = 184,756$$

Simple way to obtain an approximate permutation distribution:

Rather than using all the permutations, we take a random sample of the permutations and perform the steps involved in a permutation test on the randomly sampled permutations, say 1000.

Procedure

Summary of Steps Used in a Two-Sample Permutation Test

- Permute the $m+n$ observations between the two treatments so that there are m observations for treatment 1 and n observations for treatment 2 . Obtain all possible permutations. The number of possibilities are

$$\begin{pmatrix} m+n \\ m \end{pmatrix} = \frac{(m+n)!}{m!n!}$$

- For each permutation of the data, compute the test statistics. The permutation distribution of the test statistics characterizes the null distribution if the null hypothesis is true.
- Use the permutation distribution to calculate p-value or critical value. For example, upper-tail test

$$P_{\text{upper tail}} = \frac{\text{number of } D\text{ 's } \geq D_{\text{obs}}}{\begin{pmatrix} m+n \\ m \end{pmatrix}}$$

For a two-sided test, perform a similar procedure on the absolute values

$$P_{\text{two wail}} = \frac{\text{number of } \mid D'\text{ 's} \mid \geq |D_{\text{obs}} \mid}{\begin{pmatrix} m+n \\ m \end{pmatrix}}$$

- If a predetermined level of significance has been set, declare the test to be statistically significant if the p-value is less than or equal to this level.