

IEOR 165 – Lecture 18

Two-Sample Location Tests

1 Framework for Two-Sample Location Tests

In many cases, we are interested in deciding if the mean (or median) of one group of random variables is equal to the mean (or median) of another group of random variables. Even for this seemingly simple question, there are actually multiple different null hypothesis that are possible. And corresponding to each of these possible nulls, there is a separate hypothesis test. The reason that there are multiple nulls is that we might have different assumptions about the distribution of the data, the variances, or the possible deviations.

A canonical example of a two-sample location test is the following null hypothesis

$$H_0 : X_i, Y_i \sim \mathcal{N}(\mu, \sigma^2), \text{ where } X_i, Y_i \text{ are iid, and } \mu, \sigma^2 \text{ are unknown.}$$

For two-sample tests, we are typically interested in the case where both the mean (or median) and variance are unknown. The reason is that a common setup is for the two groups being compared to be a *control* group and a *treatment* group; and the purpose of including a control group is to be able to estimate the amount of effect without any intervention.

One of the most significant differences between various null hypothesis for two-sample tests are the concepts of *paired* and *unpaired* tests.

- The idea of a paired test is that each point in group 1 is related to each point in group 2. For instance, group 1 might be the blood pressure of patients before taking a drug, and group 2 might be the blood pressure of the same patients after taking a drug. Mathematically, an example of a null hypothesis for a paired two-sample test is

$$H_0 : X_i, Y_i \sim \mathcal{N}(\mu_i, \sigma^2), \text{ where } X_i, Y_i \text{ are independent, and } \mu_i, \sigma^2 \text{ is unknown.}$$

The idea of this null hypothesis is that the i -th measurement from each group (i.e., X_i and Y_i) have the same mean μ_i , and this mean can be different for each value of i .

- The idea of an unpaired test is that each point in group 1 is unrelated to each point in group 2. For instance, we might want to compare the satisfaction rate of two groups of cellphone customers, where group 1 consists of customers with Windows Phones, and group 2 consists of customers with Android Phones. Mathematically, an example of a null hypothesis for an unpaired two-sample test is

$$H_0 : X_i, Y_i \sim \mathcal{N}(\mu, \sigma^2), \text{ where } X_i, Y_i \text{ are independent, and } \mu, \sigma^2 \text{ is unknown.}$$

The idea of this null hypothesis is that the i -th measurement from each group (i.e., X_i and Y_i) have the same mean μ .

2 Two-Sample Unpaired Z -Test

Suppose we have n_x measurements of X_i and n_y measurements of Y_i . There are three possible concepts of extreme deviations, each with a corresponding equation. Suppose σ_x^2, σ_y^2 are known constants. If $Z \sim \mathcal{N}(0, 1)$, then the three tests are

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_X, \sigma_x^2), Y_i \sim \mathcal{N}(\mu_Y, \sigma_y^2)$, where X_i, Y_i are independent and $\mu_X \geq \mu_Y$:

$$p = \mathbb{P}\left(Z < \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}}\right)$$

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_X, \sigma_x^2), Y_i \sim \mathcal{N}(\mu_Y, \sigma_y^2)$, where X_i, Y_i are independent and $\mu_X \leq \mu_Y$:

$$p = \mathbb{P}\left(Z > \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}}\right)$$

- Two-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_0, \sigma^2), Y_i \sim \mathcal{N}(\mu_0, \sigma^2)$, where X_i, Y_i are independent:

$$p = \mathbb{P}\left(|Z| > \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}} \right|\right)$$

The intuition behind these tests is that, by the properties of iid Gaussian distributions, we know that $\bar{X} - \bar{Y} \sim \mathcal{N}(0, \sigma_x^2/n_x + \sigma_y^2/n_y)$. This test is sometimes used even if the distribution in the null hypothesis is not Gaussian. The reason is again the central limit theorem, which allows us to approximate the p -value with the above equations when we have a “large” amount of data.

3 Two-Sample Unpaired t -Test

Suppose we have n_x measurements of X_i and n_y measurements of Y_i . There are three possible concepts of extreme deviations, each with a corresponding equation. Suppose σ^2 is an unknown constant, and let

$$s^2 = \frac{1}{n_x + n_y - 2} \left(\sum_{i=1}^{n_x} (X_i - \bar{X})^2 + \sum_{i=1}^{n_y} (Y_i - \bar{Y})^2 \right).$$

If t has a Student's t -distribution with $n_x + n_y - 2$ degrees of freedom, then the three tests are

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_X, \sigma^2), Y_i \sim \mathcal{N}(\mu_Y, \sigma^2)$, where X_i, Y_i are independent and $\mu_X \geq \mu_Y$:

$$p = \mathbb{P}\left(t < \frac{\bar{X} - \bar{Y}}{s \sqrt{1/n_x + 1/n_y}}\right)$$

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_X, \sigma^2), Y_i \sim \mathcal{N}(\mu_Y, \sigma^2)$, where X_i, Y_i are independent and $\mu_X \leq \mu_Y$:

$$p = \mathbb{P}\left(t > \frac{\bar{X} - \bar{Y}}{s\sqrt{1/n_x + 1/n_y}}\right)$$

- Two-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_0, \sigma^2), Y_i \sim \mathcal{N}(\mu_0, \sigma^2)$, where X_i, Y_i are independent:

$$p = \mathbb{P}\left(|t| > \left| \frac{\bar{X} - \bar{Y}}{s\sqrt{1/n_x + 1/n_y}} \right|\right)$$

The intuition behind these tests is that we have $n_x + n_y - 2$ degrees of freedom because we are using the sample means \bar{X}, \bar{Y} in our estimate of the sample variance s^2 . This test is sometimes used even if the distribution in the null hypothesis is not Gaussian. The reason is again the central limit theorem, which allows us to approximate the p -value with the above equations when we have a “large” amount of data.

4 Two-Sample Paired Z -Test

Suppose we have n measurements of X_i and Y_i , and let σ^2 be a known constant. There are three possible concepts of extreme deviations, each with a corresponding equation. If $Z \sim \mathcal{N}(0, 1)$, then the three tests are

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_{X,i}, \sigma^2), Y_i \sim \mathcal{N}(\mu_{Y,i}, \sigma^2)$, where X_i, Y_i are independent and $\mu_{X,i} \geq \mu_{Y,i}$:

$$p = \mathbb{P}\left(Z < \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{2/n}}\right)$$

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_{X,i}, \sigma^2), Y_i \sim \mathcal{N}(\mu_{Y,i}, \sigma^2)$, where X_i, Y_i are independent and $\mu_{X,i} \leq \mu_{Y,i}$:

$$p = \mathbb{P}\left(Z > \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{2/n}}\right)$$

- Two-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_i, \sigma^2), Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, where X_i, Y_i are independent:

$$p = \mathbb{P}\left(|Z| > \left| \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{2/n}} \right|\right)$$

The intuition behind these tests is that, by the properties of iid Gaussian distributions, we know that the difference of a single pair of measurements $X_i - Y_i \sim \mathcal{N}(0, 2\sigma^2)$, and so $\bar{X} - \bar{Y} \sim \mathcal{N}(0, 2\sigma^2/n)$. This test is sometimes used even if the distribution in the null hypothesis is not Gaussian. The reason is again the central limit theorem, which allows us to approximate the p -value with the above equations when we have a “large” amount of data. The difference between this test and the two-sample unpaired Z -test is in the assumption in the null hypothesis regarding what is known about the mean.

5 Two-Sample Paired t -Test

Suppose we have n measurements of X_i, Y_i , and let σ^2 be an unknown constant. There are three possible concepts of extreme deviations, each with a corresponding equation. Let

$$\Delta_i = X_i - Y_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\Delta_i - \bar{\Delta})^2.$$

If t has a Student's t -distribution with $n-1$ degrees of freedom, then the three tests are

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_{X,i}, \sigma^2), Y_i \sim \mathcal{N}(\mu_{Y,i}, \sigma^2)$, where X_i, Y_i are independent and $\mu_{X,i} \geq \mu_{Y,i}$:

$$p = \mathbb{P}\left(t < \frac{\bar{\Delta}}{s\sqrt{1/n}}\right)$$

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_{X,i}, \sigma^2), Y_i \sim \mathcal{N}(\mu_{Y,i}, \sigma^2)$, where X_i, Y_i are independent and $\mu_{X,i} \leq \mu_{Y,i}$:

$$p = \mathbb{P}\left(t > \frac{\bar{\Delta}}{s\sqrt{1/n}}\right)$$

- Two-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_i, \sigma^2), Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, where X_i, Y_i are independent:

$$p = \mathbb{P}\left(|t| > \left| \frac{\bar{\Delta}}{s\sqrt{1/n}} \right| \right)$$

The intuition behind these tests is that we have $n-1$ degrees of freedom because we are using the differences Δ_i to construct our estimate of the sample variance s^2 . This test is sometimes used even if the distribution in the null hypothesis is not Gaussian. The reason is again the central limit theorem, which allows us to approximate the p -value with the above equations when we have a “large” amount of data.

6 Mann-Whitney U Test

Just as in the case of one-sample tests where we had a nonparametric test, a similar test exists for the two sample case. Suppose the null hypothesis is given by

$$H_0 : \text{median}(X_i) = \text{median}(Y_i) = m, \text{ where } X_i, Y_i \text{ are iid and } F_x(u) = F_y(u).$$

In fact, this test is not applicable if the distributions have different variances but the same means or medians.

The test works as follows. The data X_i, Y_i is placed into a single list that is sorted by ascending order, and the rank of a data point is defined as its ordinal position in the single list. Next, the sum of the ranks for all data points X_i is computed, and call this value R . The test statistic U is defined as

$$U = \min\{R - n_x(n_x + 1)/2, n_x n_y - R + n_x(n_x + 1)/2\}.$$

The p -value is then computed using the distribution of U . The computation of this is beyond the scope of this class, but there are two points of intuition. The first is that the distribution of U can be approximated by $\mathcal{N}(n_x n_y / 2, n_x n_y (n_x + n_y + 1) / 12)$. The second is that U is a measure of how evenly mixed the data is, with more extreme values indicating non-even mixing. If X_i, Y_i have identical distributions, then you would expect the data to be evenly mixed.