Lecture 17
Hypothesis Testing of Means and Variances

Chao Song

College of Ecology

Lanzhou University

November 18, 2024

Procedure of hypothesis testing

Hypothesis testing involves the following steps:

- Forming a null (H_0) and alternative hypothesis (H_A);
- Calculating test statistics whose distribution we know;
- Determine a rejection region with a pre-specified significance level α;
- Determine whether the test statistic falls within the rejection region or compare p-value to the significance level.

The ubiquity of normal distribution in practice and central limit theorem mean that hypothesis tests based on underlying normal distributions has widespread application.

Test about one mean

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$. A hypothesis about the mean is typically

$$\mathrm{H}_0 : \mu = \mu_0$$
$$\mathrm{H}_\mathrm{A} : \mu \neq \mu_0$$

If the σ^2 is known, we can derive a test statistic based on the properties of normal distributions as

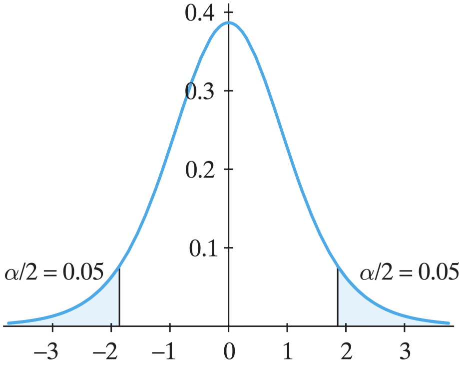
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

The term $\sigma/\sqrt{n}$ is the standard deviation of the sample mean. It is often referred to as the **standard error of the mean**.

Test about one mean

the p-value of the test is calculated as

$$p = P\Big(Z \geqslant \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}} \text{ or } Z \leqslant -\frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}}\Big) = 2P\Big(Z \geqslant \frac{|\overline{X} - \mu_0|}{\sigma/\sqrt{n}}\Big)$$



Test about one mean

In practice, we rarely know the value of σ^2. Accordingly, we take a more realistic position and assume that the variance is unknown. Recall the properties of a normal distribution that

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

and the definition of t-distribution

$$T = \frac{X}{\sqrt{v/k}}$$

where X has a standard normal distribution and v has a chi-square distribution with k degrees of freedom. X and v are independent.

Test about one mean

The test statistic is

$$T = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

and p-value of the test is calculated as

$$p = P\Big(T \geqslant \frac{|\overline{X}-\mu_0|}{s/\sqrt{n}} \text{ or } T \leqslant -\frac{|\overline{X}-\mu_0|}{s/\sqrt{n}}\Big) = 2P\Big(T \geqslant \frac{|\overline{X}-\mu_0|}{s/\sqrt{n}}\Big)$$

Test about one mean

Example: A paper firm has taken a number of measures to reduce the oxygen consuming power of their waste water discharge. They want to test if the measures they took has effectively changed the previous mean of 500. They monitored their waste water for each of the 25 days and observed a sample mean of 308.8 and a standard deviation 115.15.

The null hypothesis tested here is H_0: $\mu = 500$. Given $\overline{X} = 308.8$, $s = 115.15$ and $n = 25$, The test statistic is

$$T = \frac{\overline{X} - \mu_o}{s/\sqrt{n}} = -8.30$$

Here T follows a t-distribution with $25 - 1 = 24$ degrees of freedom. Thus,

$$p = P(T \leqslant -8.3 \; or \; T \geqslant 8.3) = 1.63 \times 10^{-8}$$

Test about one mean

In this example, as the paper firm is only interested in testing whether their measures have reduced the oxygen consuming power of the waste water, it makes more sense to perform a one-tailed test, with H_A: $\mu_0 < 500$.

The test statistic remains the same, but we only count the left tail of the probability when calculating p-value. That is

$$p = P(T < -8.3) = 8.17 \times 10^{-9}$$

Test of the equality of two means

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be two independent random samples of sizes n and m from two normal distributions $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$. The test on whether the means are equal has the hypothesis

$$\text{H}_0\colon \mu_X = \mu_Y$$
$$\text{H}_\text{A}\colon \mu_X \neq \mu_Y$$

The sample mean from a normal distribution also has a normal distribution, i.e., $\overline{X} \sim N(\mu_X, \sigma_X^2/n)$ and $\overline{Y} \sim N(\mu_Y, \sigma_Y^2/m)$. Because X and Y are independent, $\overline{X} - \overline{Y} \sim N(\mu_X - \mu_Y, \sigma_X^2/n + \sigma_Y^2/m)$. When σ_X^2 and σ_Y^2 are known, the test statistic

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \sim N(0, 1)$$

Test of the equality of two means

Realistically, the variances are typically not known. Thus, the standard normal distribution based test is rarely applicable in practice.

If we can assume that $\sigma_X = \sigma_Y$, then

$$T = \frac{\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}}{\sqrt{\frac{(n-1)s_X^2}{\sigma_X^2} + \frac{(m-1)s_Y^2}{\sigma_Y^2} \Big/ (n+m-2)}}$$

$$= \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim t(n+m-2)$$

Test of the equality of two means

If we cannot assume $\sigma_X = \sigma_Y$, we use the asymptotic normal distribution of maximum likelihood estimate to derive the test statistic

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\hat{\sigma}_X^2}{n} + \frac{\hat{\sigma}_Y^2}{m}}} \sim N(0, 1)$$

where $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ are maximum likelihood estimates of the respective variances.

Instead of normal distribution, Bernard Welch suggested that the test statistic is better approximated by a t-distribution with degrees of freedom

$$v \approx \frac{(\frac{s_X^2}{n} + \frac{s_Y^2}{m})^2}{\frac{s_X^2}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}}$$

This test is commonly referred to as **Welch's unequal variance t-test**.

Test of the equality of two means

Do we assume equal variance or not?

- When we assume equal variance but the variance are actually not equal, the type I error rate can be either inflated or reduced. The deviation of type I error rate from α is not substantial if sample sizes are similar, but can be very severe if sample sizes are vastly different.
- When the variance are actually equal but we do not assume they are equal, Welch's t-test has lower statistical power than the t-test assuming equal variance, but the loss of power is usually not substantial.
- If sample sizes are very different, Welch's t-test is a safer choice. But if sample sizes are similar and the sample variances look similar, equal variance t-test gives you more power and not much risk of wrong type I error rate.

Test about proportions

Suppose we have observed X success in n independent Bernoulli trials. We want to test the hypothesis H_0: $p = p_0$ vs H_A: $p \neq p_0$.

Recall from central limit theorem, we know that binomial distribution can be approximated by a normal distribution, i.e., X is approximately normal with mean np and variance $np(1-p)$. Thus a test statistic for H_0: $p = p_0$ is

$$Z = \frac{X - np_o}{\sqrt{np_0(1-p_0)}} = \frac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$$

Test of the equality of two means

So far, the test comparing two means assumes independence between X and Y. In many cases, we have a paired design and thus the independence assumption is not met. In this case, we may use a **paired t-test**.

Example: We took the pulse rate for 10 people before and after taking a particular treatment. Let X and Y be the pulse rates before and after the treatments. Clearly, X and Y are not independent because pulse rate of the same person is likely to be similar. In this case, we cannot use a two-sample t-test to compare the means of X and Y.

Test of the equality of two means

If X and Y are dependent, let $W = X - Y$, and the hypothesis that $\mu_X = \mu_Y$ would be replaced with the hypothesis $\mu_W = 0$. If we can assume that the distribution of W is $N(\mu_W, \sigma_W^2)$, we can test the hypothesis $\mu_W = \mu_0$ with the test statistic

$$T = \frac{\overline{W} - \mu_0}{s_W/\sqrt{n}}$$

where n is the number of pairs of X and Y. The test statistic $T \sim t(n-1)$

This test is referred to as **paired t-test**. It is essentially a one sample t-test on the differences of two variables.

Test of the equality of two proportions

Let X_1 and X_2 represent the number of observed success in n_1 and n_2 independent trials with success probability p_1 and p_2. To test the hypothesis H_0: $p_1 = p_2$,

Using the normal approximation of binomial distribution and the properties of normal distributions, we have

$$Z = \frac{X_1/n_1 - X_2/n_2 - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \sim N(0,1)$$

Under H_0, $p_1 = p_2$. We thus replace p_1 and p_2 with the maximum likelihood estimate $\hat{p} = (X_1 + X_2)/(n_1 + n_2)$ to have the test statistic

$$Z = \frac{X_1/n_1 - X_2/n_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}} \sim N(0,1)$$

Test about one variance

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$. Test the hypothesis H_0: $\sigma^2 = \sigma_0^2$.

For a normally distributed random variable, we have

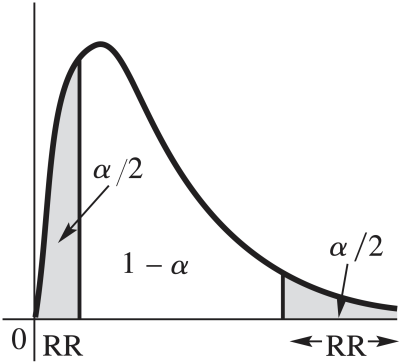
$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Thus, we can derive a test statistic

$$T = \frac{(n-1)s^2}{\sigma_0^2}$$

Test about one variance

If H_0 is not true, the test statistic tends to be either very small or very large. We thus reject the null hypothesis if the test statistic falls in either tail of a χ^2 distribution with $n - 1$ degrees of freedom.



Test of the equality of two variances

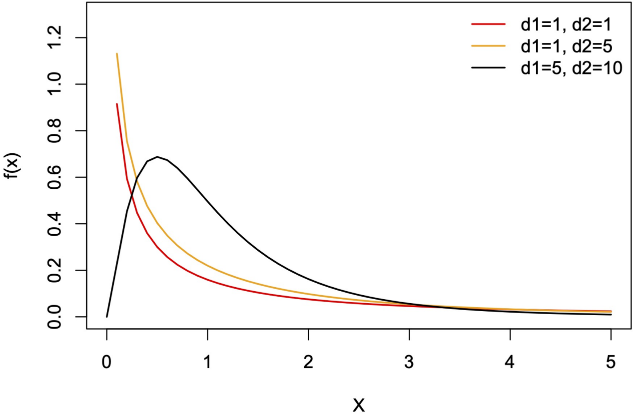
Let X_1 and X_2 be two normally distributed random variables with variance σ_1^2 and σ_2^2 respectively. Test the hypothesis H_0: $\sigma_1^2 = \sigma_2^2$.

To address this problem, we first introduce a new continuous distribution, **F-distribution**: Let $S_1 \sim \chi^2(d_1)$, $S_2 \sim \chi^2(d_2)$, and the two random variables are independent, then the variable defined below follows a $F-$distribution with degrees of freedom d_1 and d_2

$$F = \frac{S_1/d_1}{S_2/d_2}$$

Test of the equality of two variances

The F-distribution has two degrees of freedom as its parameters. These two parameters determine the shape of the PDF.



Test of the equality of two variances

To test the equality of two variances from normal distributions, we use the fact that $(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$. Thus the test statistic is

$$T = \frac{\frac{(n_1-1)s_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-1)s_2^2}{\sigma_2^2}/(n_2-1)} = \frac{s_1^2}{s_2^2} \sim F_{(n_1-1,n_2-1)}$$

using the fact that $\sigma_1 = \sigma_2$ under the null hypothesis.

The null hypothesis is rejected when the test statistic falls in either the lower or the upper tail of the F-distribution.