Hypothesis testing concepts

Hypothesis testing

A hypothesis test requires a null hypothesis and an alternative hypothesis.

- **Null hypothesis**: the hypothesis that there is no statistically significant observation
- **Alternative hypothesis**: the hypothesis that states something statistically significant is happening.

Sometimes it helps to think of a null hypothesis as the "old theory" and the alternative hypothesis as the "new theory".

The goal of a hypothesis test is to *provide evidence against the null hypothesis*.

Note that this is *not* the same as proving the alternative hypothesis true.

Null hypothesis versus alternative hypothesis

Example

Suppose we wanted to test if the average height of men is different to the average height of women. The null hypothesis is "the average heights of men and women are the same". The alternative hypothesis is "the average heights of men and women differ".

Using symbols, we would write:

Let μ_1 denote the mean height of men and let μ_2 denote the mean height of women. We are testing the hypotheses

$$H_0\colon \mu_1 = \mu_2 \text{ versus } H_1\colon \mu_1 \neq \mu_2$$

This is an example of a *two-sided test*. We would be testing to see if the data provides evidence to suggest that the two means are *not the same*.

Null hypothesis versus alternative hypothesis

Example

Suppose we wanted to test if men are on average *taller* than women. The null hypothesis is "men are on average shorter than women". The alternative hypothesis is "men are on average taller than women".

Using symbols, we would write:

Let μ_1 denote the mean height of men and let μ_2 denote the mean height of women. We are testing the hypotheses

$$H_0\colon \mu_1 \leqslant \mu_2 \text{ versus } H_1\colon \mu_1 > \mu_2$$

This is an example of a *one-sided test*. We would be testing to see if the data provides evidence to suggest that the *one mean is less than another*.

Choosing the hypotheses

Choosing the null and alternative hypotheses involves some nuance. Usually, the alternative hypothesis is the scenario you want to find evidence *for*.

- Is my new software faster than my old software?

- Does a new service offered by a company increase sales?

But it can vary depending on circumstances.

p-values

A p-value is the answer to the question,

If the null hypothesis is true, what is the probability of an outcome at least as extreme as this one?

If the outcome is deemed unlikely, we reject the null hypothesis.

This is done on the basis of a **significance level**, denoted by α.
Usually $\alpha = 0.05$, phrased as "a 5% level of significance".

Extremeness is measured by using a *test statistic*.

Test statistics

A **test statistic** is a quantity derived from a sample which used to test a hypothesis. For this week, we will consider:

- The z-test, which uses
$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$
Requires *known* σ.
- The t-test, which uses
$$t = \frac{\overline{x} - \mu_0}{\mathrm{SE}}$$
Requires either large sample size or normally distributed population.

The z-test

The z-test

For a z-test, the test statistic is

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}.$$

Here, $\overline{x}$ is the sample mean, μ_0 is the mean used in the null hypothesis, σ is the population standard deviation under the null hypothesis, and n is the sample size.

Necessary assumptions:

- Population is normally distributed OR large sample size
- Population standard deviation σ is known.

The central limit theorem implies that the test statistic has an $N(0,1)$ distribution. So the p-value for a two-sided test is given by

$$p = 2 \times P(Z \geqslant |z|), \text{ where } Z \sim N(0,1).$$

Why?

Example

A school tests 30 of their students on a maths test. The national average result on this test is 50 points, with a standard deviation of 3.

The students tested at this school resulted in an average of 47. Is this outcome statistically significant?

One-sided z-test

For a one-sided test, the test statistic is calculated the same way:

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}.$$

But the p-value is calculated differently.

If the hypothesis is

$$H_0\colon \mu \geqslant \mu_0 \text{ versus } H_1\colon \mu < \mu_0$$

then use $p = P(Z \leqslant z)$

If the hypothesis is

$$H_0\colon \mu \leqslant \mu_0 \text{ versus } H_1\colon \mu > \mu_0$$

then use $p = P(Z \geqslant z)$

Example

A school wants to determine whether their students perform at a higher standard than the national average on a certain maths test. The national average result on this test is 50 points, with a standard deviation of 3.

Formulate hypotheses:

The school then tested 40 students, resulting in an average of 52.

The t-test

The t-test

The t-test uses the test statistic

$$t = \frac{\overline{x} - \mu_0}{\mathrm{SE}}$$

Recall that $\mathrm{SE} = \frac{s}{\sqrt{n}}$, where s is the sample standard deviation.

Necessary assumptions:

- Population is normally distributed OR large sample size

The p-value is calculated in a similar manner to the z-test, but instead of $Z \sim N(0,1)$, it uses $T \sim t_{n-1}$, where t_{n-1} denotes the t-distribution.

For example,

- With a sample size of 30 and when testing $H_0\colon \mu = \mu_0$ versus $H_1\colon \mu \neq \mu_0$, use $2 \times P(T \geqslant |t|)$ with $T \sim t_{29}$.
- With a sample size of 55 and when testing $H_0\colon \mu \leqslant \mu_0$ versus $H_1\colon \mu > \mu_0$, use $P(T \geqslant t)$ with $T \sim t_{54}$.

The probabilities can be calculated in R using the `pt` function.

Example

The national average result on a maths test is 50 points, but the standard deviation is not known.

A school tests 30 of their students on this test. This resulted in an average of 47, with sample standard deviation 4. Is this outcome statistically significant?

Confidence intervals via t-distributions

Testing hypotheses by confidence intervals

The t-distribution can be used to calculate $\gamma \times 100\%$ confidence intervals:

$$\overline{x} \pm t_{n-1,(1+\gamma)/2} \times \text{SE}.$$

For a significance level of α, we have $\gamma = 1 - \alpha$.

The value of $t_{\nu,p}$ is calculated in R using the `qt` function.

After calculating a confidence interval, reject H_0 if μ_0 is not contained in the interval. Otherwise, fail to reject H_0.

Example

The national average result on a maths test is 50 points, but the standard deviation is not known.

A school tests 30 of their students on this test. This resulted in an average of 47, with sample standard deviation 4. Find a 90% confidence interval and a 95% confidence interval for the school's average maths test result. Is the outcome statistically significant at the 10% significance level?