Chapter 6

Hypothesis testing

In hypothesis testing we assume two possibilities, called hypotheses, about the state of nature. They must be mutually exclusive and collectively exhaustive events. Practically, the true state of nature is never known, but we can use statistical evidence to make judgement and reject one of the hypothesis tested.

One hypothesis is called a null hypothesis and denoted by H_0 and the other is called an alternative hypothesis and denoted by H_1. Suppose we want to test where a population parameters θ is equal to a specific value θ_0 against the opposite hypothesis that θ is different than θ_0. Formally we define null hypothesis and an alternative hypothesis as follows:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

and we call this hypothesis testing as a two-sided alternative. Alternatively, we could test whether population parameters θ is larger or equal a specific value θ_0 against the opposite hypothesis that θ is lower than θ_0. In this case, we define null hypothesis and an alternative hypothesis as follows:

$$H_0 : \theta \geq \theta_0$$
$$H_1 : \theta < \theta_0$$

and we call this hypothesis testing as a one-sided alternative.

6.1. Test statistics

To determine whether H_0 can be rejected or not we construct a test statistics T, which has a known sampling distribution assuming that the null hypothesis H_0 were true. Given the sampling distribution, we can determine what is the probability of observing a certain outcome of T. If this probability is smaller than a predetermined value, α, called significance level, then we can reject the null hypothesis.

Significance level. The significance level α of a test is also the probability of making what is called *type I error*, i.e. the error of falsely rejecting the null hypothesis when it is true. To minimize the error, the significance level is therefore chosen to be small, usually to larger than 5%.

For a given significance level α, we can find the value, cv, such that, if the null hypothesis were true, obtaining a T value larger than cv has probability α, i.e.

$$P(T > \text{cv}_\alpha | \text{H}_0) = \alpha$$

cv_α are called *critical value* for the hypothesis test. What if the value of our test statistic, T, is larger than the value cv_α? This means there is, at most, an $\alpha 100\%$ chance that the null hypothesis is true therefore we reject the null hypothesis.

Test power. Suppose now that the alternative hypothesis H_1 is true. The power of a test is the probability of rejecting H_0 given that H_1 is true (i.e. the probability of correctly rejecting the null). The larger is the power of a test, the lower are the chances of committing what is called *type II error*, i.e. the error of failing to reject the null hypothesis H_0 when the alternative is true.

We can write the power of a test as

$$P(T > \text{cv}_\alpha | \text{H}_1)$$

for a one-sided test and

$$P(|T| > \text{cv}_{\frac{\alpha}{2}} | \text{H}_1)$$

for a two-sided test. What determines the power of a test? Typically the power increases as 1) the estimator takes values that are different than the value under the null hypothesis and 2) as the sample size n increases. A test is said to be consistent if its power approaches one as n increases.

P-values The p-value or probability value is the probability of obtaining test results at least as extreme as the results actually observed during the test, assuming that the null hypothesis is correct. Let $\hat{T}$ be the observed value for our test statistics T. Then the p-value is equal to i.e.

$$\text{p-value} = P(T > \hat{T} | \text{H}_0)$$

If $\hat{T} = \text{cv}_\alpha$, it must be that p-value $= \alpha$. If $\hat{T} > \text{cv}_\alpha$ it must be that p-value $< \alpha$

6.2. Applications

Two-sided hypothesis test of population mean Consider a normally distributed random variable X, with population mean μ and standard deviation σ. We have shown that the distribution of the sample mean of an independent $x_1, ..., x_n$ is given by

$$\bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

and we know that

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

where s is an estimator of σ. We want to test the claim that the population mean μ is equal to a certain value μ_0, i.e.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

The test statistics under the null hypothesis is:

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Remember that under H_0, $\mu = \mu_0$, and therefore T follows a t_{n-1} distribution. This implies that the critical value is given by

$$\text{cv}_\alpha = t_{\frac{\alpha}{2}, n-1}$$

We reject the null hypothesis, H_0, if $|T| > \text{cv}_\alpha$.

One-sided hypothesis test of population mean It is also possible to compute a one-sided hypothesis test, where we test the alternative hypothesis that μ is either greater than or less than some specified value, i.e. we use either of the following set of hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

or

$$H_0 : \mu = \mu_0$$
$$H_2 : \mu < \mu_0$$

In both cases the test statistics under the null hypothesis is equal to the one computed above. Instead the rejection rules are different:

- For H_0 versus H_1, we reject if $T > \text{cv}_\alpha$
- For H_0 versus H_1, we reject if $T < -\text{cv}_\alpha$

Notice finally that the one-sided tests require use of one-sided critical values, $t_{\alpha,n-1}$, instead of $t_{\frac{\alpha}{2},n-1}$

Hypothesis Test of a difference in population means. Suppose that there exist two independent normal random variables X and Y, with a population expected value and standard deviation equal to μ_x and σ_x, μ_y and σ_y, respectively.

An independent random sample of size n_x is drawn from X and of size n_y is drawn from Y. The sample means and sample standard deviations for each sample are given by

- $\bar{x}$ and s_x for X
- $\bar{y}$ and s_y for Y

It follows that $\bar{x} \sim N(\mu_x, \frac{\sigma_x^2}{n_x})$ and $\bar{y} \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$. Since X and Y are independent, then it must be that

$$\bar{x} - \bar{y} \sim \mathcal{N}\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)$$

Suppose now we test the null hypothesis

$$H_0 : \mu_x - \mu_y = \mu_0$$
$$H_1 : \mu_x - \mu_y \neq \mu_0$$

Under the null hypothesis. The test statistics

$$T = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

and since $\mu_x - \mu_y$ follows a normal distribution, T follows a t-distribution with v degree of freedom, equal to

$$v = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{1}{n_x-1}\left(\frac{s_x^2}{n_x}\right)^2 + \frac{1}{n_y-1}\left(\frac{s_y^2}{n_y}\right)^2}$$

An alternative approach is to use the conservative degrees of freedom setting that v is equal to the smaller of $n_x - 1$ or $n_y - 1$.

6.3. Exercises

Exercise 1 A telemarketing group claim that, after training, employees will earn an average of 1500GBP in their firs month of work. A random sample of 150 employees is drawn. Sample mean earnings for the first month of work were 1262GBP and the sample standard deviation was 432GBP. Test at the 5% significance level the null hypothesis that the population mean is 1500GBP, against the alternative that it is less than 1,500.

Exercise 2 Explain what the following hypothesis testing terms mean:

- Type I and Type II errors
- Test Power

Exercise 3 A manufacturer states that the mean number of matches in a matchbox is 50. The number of matches in a box is normally distributed. In a random sample of 20 boxes, we find that the sample mean number of matches is 49.3 and the sample standard deviation is 1.25. Test the claim that the population mean is equal to 50, against the alternative that it is not equal to 50, at a 5% level of significance.

Exercise 4 Consider a different manufacturer of matchboxes who also claims that the number of matches in a box has a population mean of 50. The number of matches in a box is normally distributed. In a random sample of 20 boxes, we find that the sample mean number of matches is 50.4 and the sample standard deviation is 1.25. Test the claim that the population mean is equal to 50, against the alternative that it is greater than 50, at a 5% level of significance.

Exercise 5 Now consider a third manufacturer of matchboxes who claims that the number of matches in a box has a population mean of 55. The number of matches in a box is normally distributed. In a random sample of 16 boxes, we find that the sample mean number of matches is 52 and the sample standard deviation is 1.5. Test the claim that the population mean is equal to 55, against the alternative that it is less than 55, at a 5% level of significance.

Exercise 6 A car company owns two factories. The manager of the company wants to know if there is any difference in the mean number of cars produced by each factory per day. She considers a random sample of 30 days. The sample mean output for Factory 1 is 420 and for Factory 2 is 408. The sample standard deviation for Factory 1 is 12.5 and for Factory 2 is 18. Test the hypothesis that Factory 1 and 2 have the same population mean, against the alternative that they have different population means, at a 5% level of significance.