Hypothesis Testing

Statistical Hypothesis Testing

- Null hypothesis procedures
    - Parametric
    - Permutational (non-parametric)
    - Rank-based permutational (non-parametric)
- Not Null hypothesis procedures
    - Bayesian and likelihood methods.

Null hypothesis and its motivation

trying to reject the null hypothesis

The null hypothesis (H_o) is often the statement that there is simply no effect or no difference between either two sample populations.

- This of’course is not what we expect when doing the experiment so we are hoping to reject the null hypothesis.
- H_A is the alternative hypothesis, the hypothesis we suspect.

EXAMPLE 6.1 Hansen et al. (2011) vaccinated 24 rhesus monkeys against a powerful form of simian immunodeficiency virus (SIV) (a simian cousin of HIV). The vaccine was unique in that it used a long-lived delivery vehicle, a herpesvirus called cytomegalovirus (CMV), to carry AIDS proteins that conferred immune responses. The researchers believed that the vaccine would provide additional protection from SIV infection. Thus, their null hypothesis was that the vaccine would provide no additional protection. The vaccine was found to protect half of the tested monkeys. This result would be highly unlikely if H_0 were true. Thus, the investigators rejected the null hypothesis of no effect, which provided implicit support for the efficacy of the new vaccine.

Significance testing

A statistical test is often performed to obtain a score. This statistical test score is then compared to the score that one would expect if the null hypothesis is true.

A p-value is the probability of calculating the test statistic score, or more extreme, than the statistic score if the null hypothesis is true.

P-value

EXAMPLE 6.2 To understand the concept of a P-value, consider a researcher who is interested in distinguishing the effect of two soil nutrient treatments, X and Y, on crop yield. The researcher stipulates the following null hypothesis:

H_0 : The true mean difference of X and Y is zero.

Data are gathered concerning the groups and a test statistic is calculated estimating the standardized (adjusted for variability) mean difference between X and Y.

To test the validity of H_0 , we define a random variable called the null distribution.

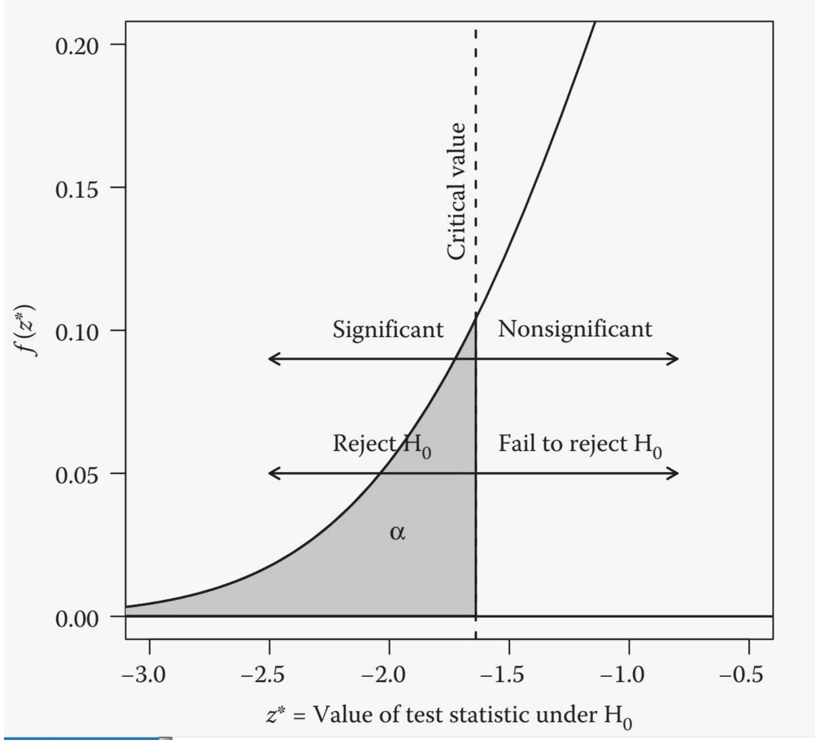
In the null distribution, the true difference of X and Y is assumed to be zero

Significance level and critical value

Significance level (α) is the cutoff where a p-value below this is considered significant enough to reject the null hypothesis.

Similarly, the critical value, is the value the test statistic must be greater than to be significant.

An α of 0.05 means that the test statistic will be greater than the critical value no more than 5% if the null hypothesis is true.



Interpretation of significance

If the p-value is low and the H_O is rejected, it doesn’t mean the H_O is wrong. It means that there is a small probability to get the test statistic score if H_O was true.

Upper, Lower, and Two-tailed tests

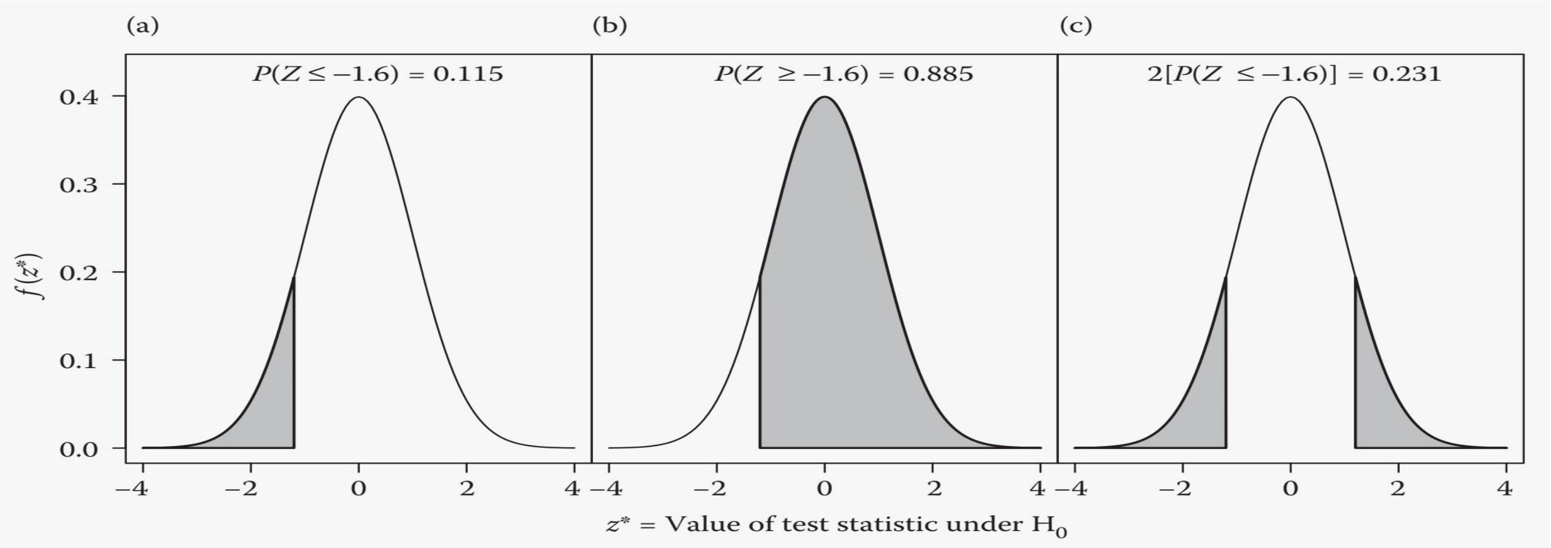
Depending on our question our test can for either extreme or both.

An upper-tail is when probability is calculated for statistic score to be greater than the score for the null hypothesis.

A lower tail is when probability is calculated for statistic score to be less than the score for the null hypothesis.

A two-tailed test is the probability where the absolute value of the statistic score is greater than the absolute value of the statistic score for null hypothesis.

Let the distribution under H0 be N(0, 1). Now, suppose that we have specified a lowertailed alternative hypothesis and calculated a test statistic of –1.2. Evidence concerning H0 comes from the tail of the distribution given in HA. Thus, the P-value is simply P(Z≤−1.2) = 0.11507 (Figure 6.3a). If we had specified an upper-tailed alternative, we would use the upper tail of the null distribution to calculate the P-value. In this case, the P-value is P(Z≥−1.2) = 0.88493 (Figure 6.3b). Finally, if we had required a two-tailed alternative, we would quantify evidence concerning the null distribution using both tails. In this case, since the null distribution is symmetric, the P-value is 2[P(Z≤|−1.2|)] = 0.23104 (Figure 6.3c). Note that, in contrast to the approach used here, only one type of test, lower-, upper-, or two-tailed, should be defined for an analysis, and this specification should be made a priori (prior to the start of the experiment).



One sample z-test

Port et al. (2000) found that for males 45–54 years of age, systolic * blood pressure was normally distributed with μ= 131 mm Hg (millimeters of mercury) and σ= 12 mm Hg. A medical administrator at a state university examines the records of 85 male faculty members in this age group and finds that x = 128 mm Hg. The administrator is concerned that the blood pressure of faculty members at his university may differ from that of the overall population of males 45–54 years of age.

$$z^* = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

One sample t-test

A one sample t-test is the same equation, however the degrees of freedom we consider is n-1 which will make the p-value slightly more conservative.

The statistical test score will be calculated using n

The probability will be determined using a distribution based on n-1

$$t^* = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Confidence intervals and Hypothesis testing

If the sample mean has a p-value of less than 0.05, than it will not be present within the range of the confidence intervals +/- $t_{(1-\alpha/2)}$, df=n-1

So for a score to be significant, it must be greater than 97.5% or less than 2.5%.

Inferences for two population means

Unpaired t-test - Samples are obtained from two independent populations. The null distribution is the difference in the means which we set to 0.

$$\bar{X} - \bar{Y} \sim N\left[\mu_X - \mu_Y, (\sigma_X^2 / n_X + \sigma_Y^2 / n_Y)\right].$$

Paired t-test - Samples are obtained from related populations. Since they are related we look at the distribution of their difference rather than difference of the means.

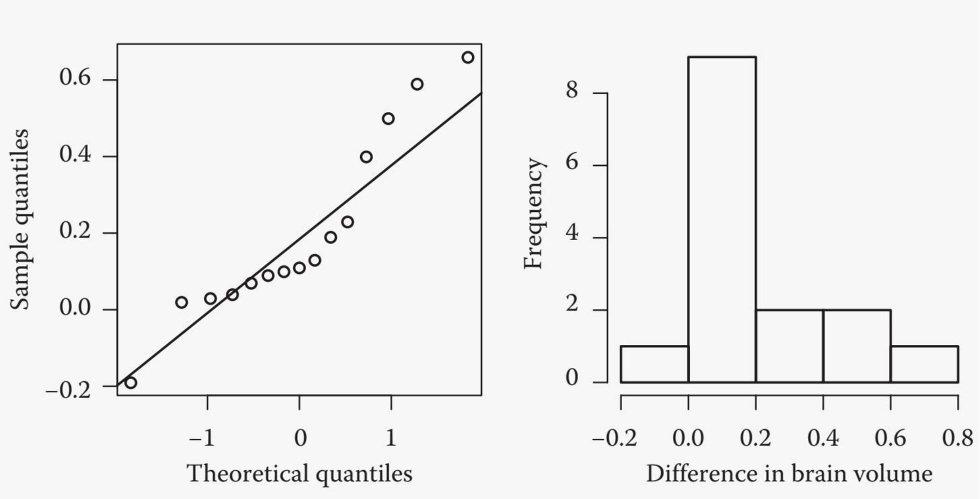
$$t* = \frac{\bar{X}_D - D_0}{S_D / \sqrt{n}}$$

Methods of determining normal distribution

Histogram - a quick visual representation can be helpful.

qqplot - theoretical quantile vs empirical quantile values. It should result in a straight diagonal line.

Shapiro- Wilk test - The null hypothesis is that the data is normally distributed.



Equal and not equal variance

Pooled Variance t-test - when Var of X and Y are equal.

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}} = \sqrt{\sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)} = \sigma\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}.$$

Welch’s Approximate t-test - when Variance of X and Y are not equal and X and Y are not paired.

$$\nu = \frac{\left((S_x^2/n_X) + (S_Y^2/n_Y)\right)^2}{\left((S_X^2/n_X)^2/(n_X - 1)\right) + \left((S_Y^2/n_Y)^2/(n_Y - 1)\right)}.$$

Type I and Type II Errors

Type I error is falsely rejecting the null hypothesis. We should expect this to occur at the rate of our alpha.

Type II error is failing to reject the null hypothesis when it is false. This occurs at the rate of probability of beta.

- Type I: reject but you should
- Type II: accept but you shouldn't