Type I and II Errors and Power

6 Type I and II Errors and Power

In this section we will comment on some of the limitations of the NHBST. We will look at the fundamental, unavoidable errors that can be made by a test, followed by the attempt to quantify the probability of these errors.

6.1 Errors

Firstly we look at the framing of the Hypotheses.

6.1.1 The Alternative Hypothesis

The potential in a NHBST to make a mistake can be better studied if there is a clear alternative Hypothesis, that is, the alternative is not simply the opposite of the Null.

In many cases we have studied, for example a test on a parameter η, our Null Hypothesis was

$$H_0: \eta = \eta_0,$$

a given value, and then the alternative was

$$H_A: \eta \neq \eta_0.$$

A more robust and testable alternative Hypothesis would be

$$H_A: \eta = \eta_1,$$

an alternative value for the parameter. This is what is known as a Binary NHBST. When we define the different types of errors, this structure for a test allows us to calcualte probabilities.

6.2 Errors

Let H_0 be the Null Hypothesis; let α be a level of significance. This symbol will also be used to represent the statement that H_0 is true. With this in mind:

A type I error is rejecting the Null Hypothesis if it is true.

If the data and the statistic calculated from it is represented by D, then in a standard NHBST, the Null Hypothesis is rejected if $P[D|H_0] < \alpha$, where α is a low probability, the level of significance.

The event of rejecting the Null Hypothesis if it is true has the same probability as $P[D|H_0]$. Therefore the probability of a type I error is α, the level of significance.

A type II error is not rejecting the Null Hypothesis when it is false.

In a binary NHBST, this is equivalent to rejecting the Null Hypothesis when the alternative is true. The probability of a type II error is denoted β.

Here are the errors in a table:

	Do not Reject	Reject
NH True	Correct	Type I
NH False	Type II	Correct

6.3 Power

The power of a binary NHBST is defined as the probability of correctly rejecting the Null Hypothesis, in other words, the probability of not making a type II error.

This probability is sometimes referred to as π. Therefore it is

$$\pi = 1 - \beta = P[\text{ reject } H_0 \mid H_1].$$

6.3.1 Power in a test on a mean

Let a NHBST 'test on a mean' be carried out with the following Hypotheses:

Null Hypothesis $\mu = \mu_0$; this will be false.

Alternative Hypothesis $\mu = \mu_1$; this will be true.

We can derive an expression for the power of the test in terms of a true mean μ_1 and so state what this tells about the sample size.

Let X be a random variable with mean μ and let S be the sample standard deviation.

To do our test we would usually use the statistic $T_{0,n}$ given by

$$T_{0,n} = \frac{\bar{X} - \mu_0}{S / \sqrt{n}},$$

follows the t-distribution for small n and then becomes N(0,1) for large n. Now let α be a low probability and define a number $t_\alpha(n)$ such that $P[T_n > t_\alpha(n)] = \alpha$.

If H_0 is in fact false, then $T_{0,n}$ given above does not in fact follow the t distribution. We will now calculate the probability of events under H_1.

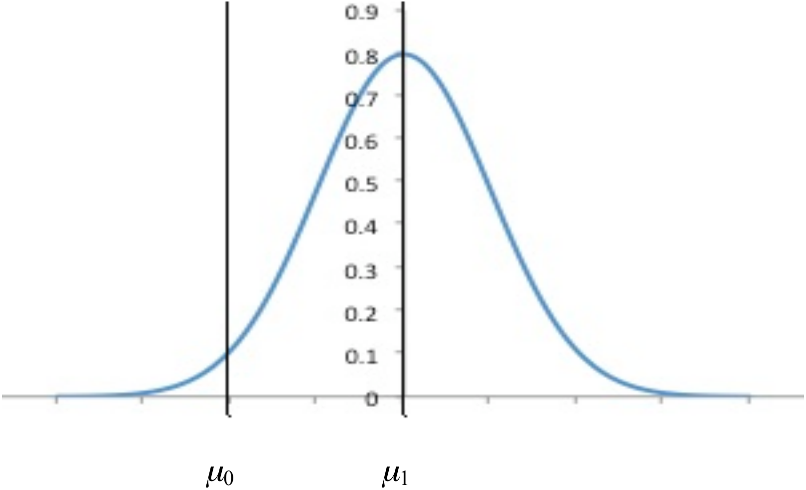
We wrongly fail to reject the Null Hypothesis in a two-tailed test on the occurrence of the event

$$-t_{\alpha/2}(n) < T_{0,n} < t_{\alpha/2}(n).$$

Given H_1 to be true, this is a type II error.

Consider first the case when $\mu_0 < \mu_1$.

Here is a diagram of the probability density function (blue line) for the sample mean $\overline{X}$, with the true mean and 'false' mean shown.



Keep in mind that as *n* increases, the values of $\bar{X}$ tend towards μ_1.

Looking at the type II error

$$P[-t_{\alpha/2}(n) < T_{0,n} < t_{\alpha/2}(n)],$$

We must rearrange this expression in terms of the quantity $T_{1,n}$ given by

$$T_{1,n} = \frac{\overline{X} - \mu_1}{S/\sqrt{n}},$$

since if H_1 is true, we know its distribution.

So if H_1 is true,

$$\beta = P[-t_{\alpha/2}(n) < T_{0,n} < t_{\alpha/2}(n)] =$$

$$P\left[-t_{\frac{1}{2}\alpha}(n) < \frac{\overline{X}-\mu_0}{S/\sqrt{n}} < t_{\frac{1}{2}\alpha}(n)\right].$$

Carrying out some algebra on this expression gives:

$$\beta = P\left[\mu_0 - t_{\frac{1}{2}\alpha}(n)\frac{s}{\sqrt{n}} < \overline{X} < \mu_0 + t_{\frac{1}{2}\alpha}(n)\frac{s}{\sqrt{n}}\right].$$

Now introduce μ_1 and the quantity $T_{1,n}$:

$$\beta = P\left[-t_{\frac{1}{2}\alpha}(n) - \frac{(\mu_1-\mu_0)\sqrt{n}}{S} < \frac{\overline{X}-\mu_1}{S/\sqrt{n}} < t_{\frac{1}{2}\alpha}(n) - \frac{(\mu_1-\mu_0)\sqrt{n}}{S}\right].$$

So

$$\beta = P\left[-t_{\frac{1}{2}\alpha}(n) - \frac{(\mu_1 - \mu_0)\sqrt{n}}{S} < T_{1,n} < t_{\frac{1}{2}\alpha}(n) - \frac{(\mu_1 - \mu_0)\sqrt{n}}{S}\right].$$

Now $T_{1,n}$ does follow the t distribution with $n - 1$ degrees of freedom and eventually becomes N(0,1). Break up this probability:

$$\beta = P\left[T_{1,n} < t_{\frac{1}{2}\alpha}(n) - \frac{(\mu_1 - \mu_0)\sqrt{n}}{S}\right] - P\left[T_{1,n} < -t_{\frac{1}{2}\alpha}(n) - \frac{(\mu_1 - \mu_0)\sqrt{n}}{S}\right].$$

As n becomes large, we know that S tends to σ.

Therefore the test values both decrease; the probability heads to 1.

How quickly depends on the value of

$$\frac{\mu_1 - \mu_0}{\sigma}.$$

For large n, we have

$$\beta = P\left[Z < z_{\frac{1}{2}\alpha} - \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma}\right] - P\left[Z < -z_{\frac{1}{2}\alpha} - \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma}\right].$$

The second probability falls away quickly, so looking at the power, which is the probability of correctly rejecting the Null;

$$\pi = P\left[Z > z_{\frac{1}{2}\alpha} - \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma}\right] + P\left[Z < -z_{\frac{1}{2}\alpha} - \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma}\right].$$

Therefore the power is

$$\pi = 1 - \Phi(z_{\alpha/2} - (\mu_1 - \mu_0)\sqrt{n}/S).$$

In our scenario of $\mu_1 > \mu_0$, the argument of Φ will decrease, S tends to σ and so the probability will tend to 1.

Due to the highly non-linear nature of the CDF for the standard normal distribution, the power calculation increase with $\sqrt{n}$ depends very much on the coefficient

$$\frac{\mu_1 - \mu_0}{\sigma}.$$

Once this number is above 2, the power has effectively converged to 1 before 10 for sample values.

The function

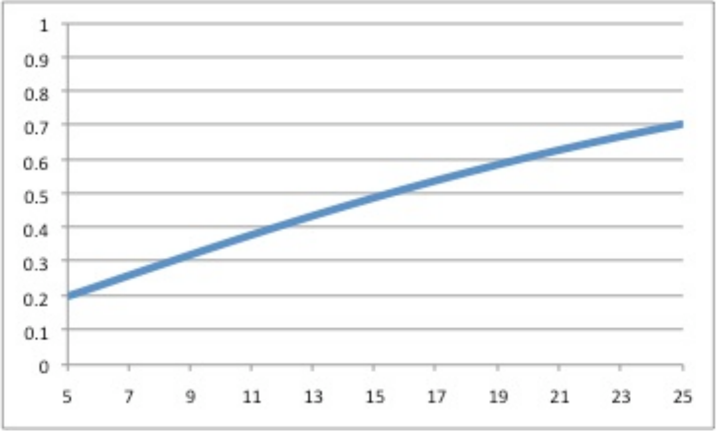
$$y = \Phi(x)$$

converges rapidly onto 1 once the argument x is above 2.

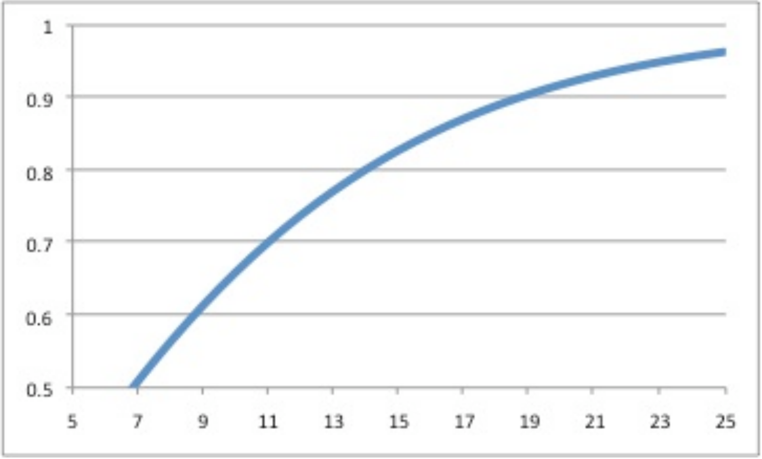
Here are graphs of the power for several values of

$$\frac{\mu_1 - \mu_0}{\sigma}.$$

$$\frac{\mu_1 - \mu_0}{\sigma} = 0.5:$$



$$\frac{\mu_1 - \mu_0}{\sigma} = 0.75:$$



$$\frac{\mu_1 \quad \mu_0}{\#} = 1:$$