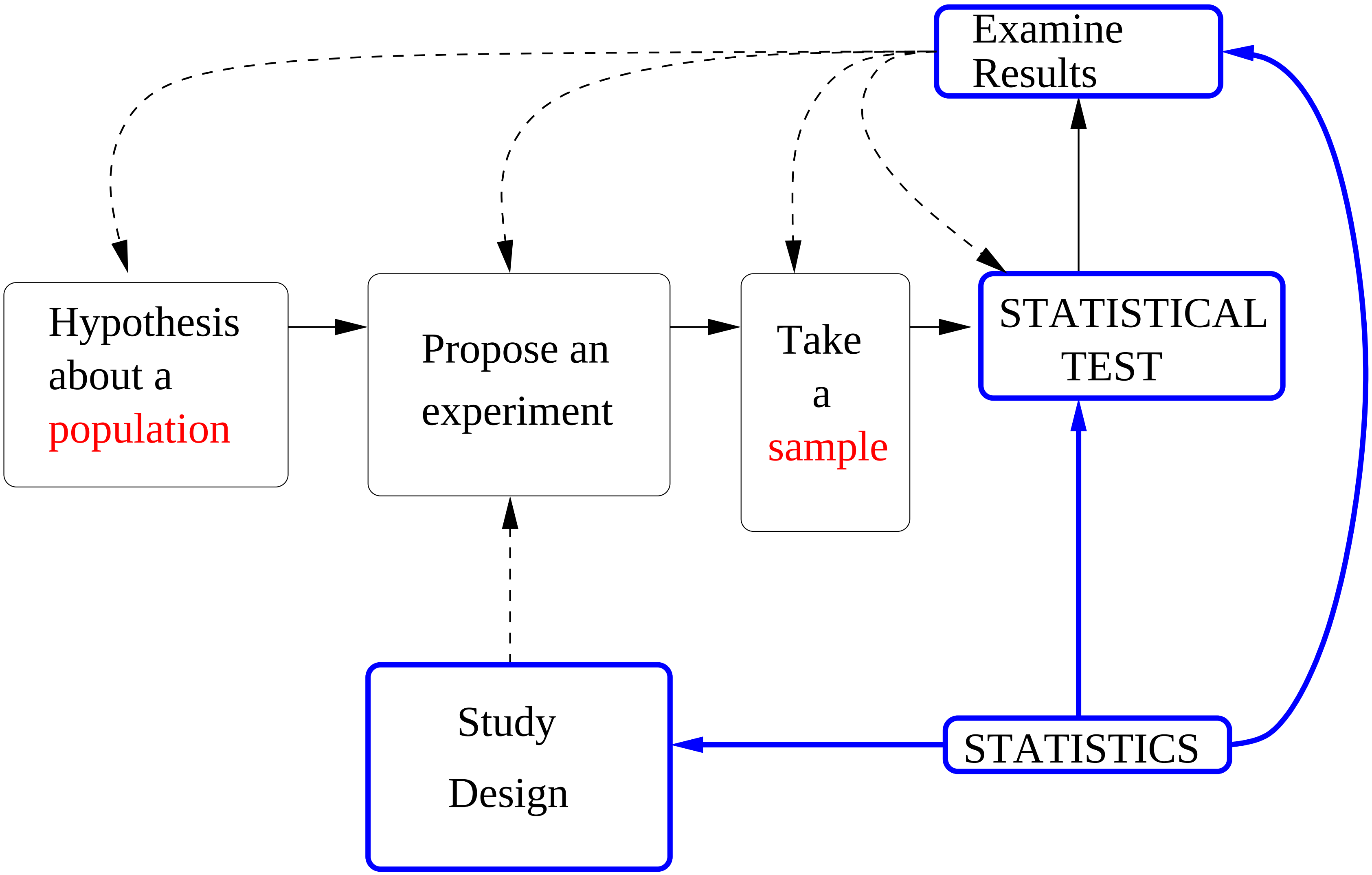
Lecture 7 : Hypothesis Tests

Jonathan Marchini

Hypothesis tests

All of these previous lectures have provided us with the basic tools we need to use statistics in practical situations.

In this lecture we consider the general framework used to test a specific hypothesis by examining some basic examples that utilize our knowledge of the Normal distribution.


Examine
Results
Hypothesis
about a
population
Propose an
experiment
Take
a
sample
STATISTICAL
TEST
Study
Design
STATISTICS

Single sample test for a population mean μ

Consider the following hypothetical situation:

From previous experience we know that the birth weights of babies in England are Normally distributed with a mean of 3000g and a standard deviation of 500g. We think that maybe babies in Australia have a mean birth weight greater than 3000g and we would like to test this hypothesis.

The Null and Alternative Hypothesis

The main hypothesis that we are most interested in is the **research hypothesis**, denoted H_1, that the mean birth weight of Australian babies is greater than 3000g.

The other hypothesis is the **null hypothesis**, denoted H_0, that the mean birth weight is equal to 3000g.

We can write this compactly as

$$H_0 : \mu = 3000\text{g}$$
$$H_1 : \mu > 3000\text{g}$$

The research hypothesis is often called the **alternative hypothesis**.

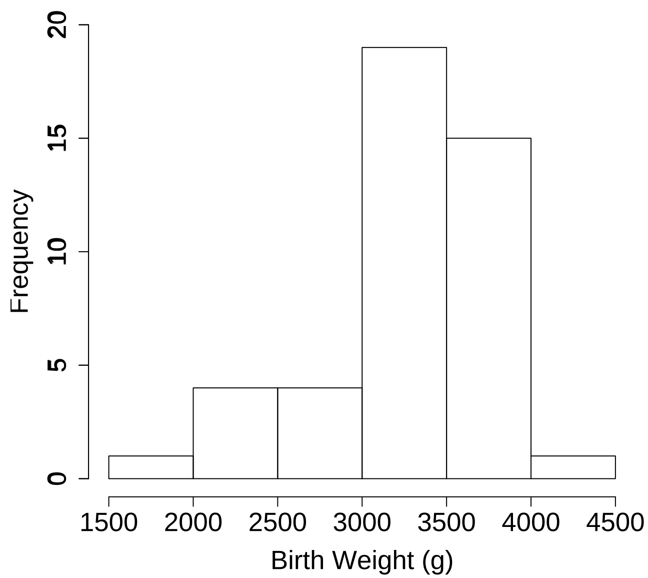
We start with the research hypothesis and “set up” the null hypothesis to be directly counter to what we hope to show.

We then try to show that, in the light of our collected data, that the null hypothesis is false.

We do this by calculating the probability of the data if the null hypothesis is true. If this probability is very small it suggests that the null hypothesis is false.

Collecting a dataset

Once we have set up our null and alternative hypothesis we can collect a sample of data. For example, the Babyboom dataset.



The sample mean of the dataset is

$$\bar{x} = 3275.955$$

We now want to calculate the probability of obtaining a sample with a mean as large as 3275.955 under the assumption of the null hypothesis H_0.

To do this we need to calculate the distribution of the mean of 44 values from a $N(3000, 500^2)$ distribution.

We know from Lecture 6 that if

$$X_1 \sim \mathrm{N}(\mu, \sigma^2) \qquad\qquad X_2 \sim \mathrm{N}(\mu, \sigma^2)$$

then

$$\overline{X} = \frac{1}{2}X_1 + \frac{1}{2}X_2 \sim \mathrm{N}\left(\frac{1}{2}\mu + \frac{1}{2}\mu, \left(\frac{1}{2}\right)^2\sigma^2 + \left(\frac{1}{2}\right)^2\sigma^2\right)$$

$$\Rightarrow \overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{2}\right)$$

In general,

If $X_1, X_2, \ldots, X_n$ are n independent and identically distributed random variables from a N(μ, σ^2) distribution then

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

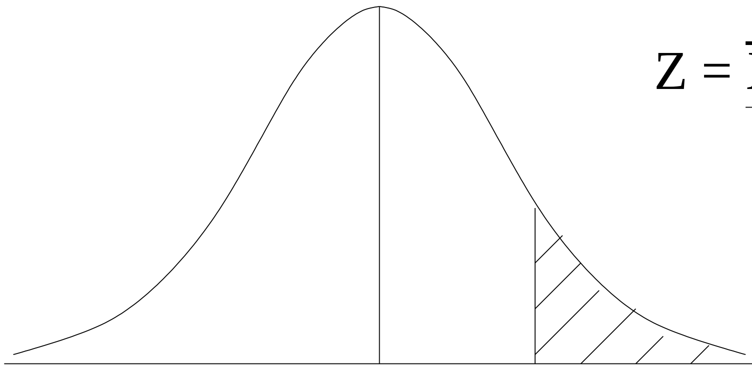
Thus, under the assumption of the null hypothesis the sample mean of 44 values from a N(3000, 500^2) distribution is

$$\overline{X} \sim \mathrm{N}\left(3000, \frac{500^2}{44}\right) = \mathrm{N}(3000, 5681.818)$$

Now we can calculate the probability of obtaining a sample with a mean as large as 3275.955 using standardization.

$\bar{X} \sim N(3000, 5681.818)$

$P(\bar{X} > 3275.955)$

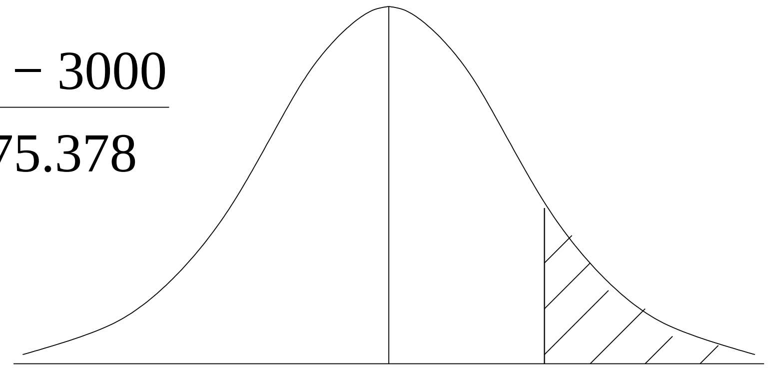


3000 3275.955

$$Z = \frac{\bar{X} - 3000}{75.378}$$

$Z \sim N(0, 1)$

$P(Z > 3.66)$



0

$$\frac{3275.955 - 3000}{75.378} = 3.66$$

$$
\begin{aligned}
P(\overline{X} > 3275.955) &= P\left(\frac{\overline{X} - 3000}{75.378} > \frac{3275.955 - 3000}{75.378}\right) \\
&= P(Z > 3.66) \\
&= 1 - 0.99985 \\
&= 0.00015
\end{aligned}
$$

This probability is called the **p-value** of the test. In this case the p-value is very low. This says that the probability of the data is very low if we assume the null hypothesis is true.

But how low does this probability have to be before we can conclude that the null hypothesis is false?

The convention within statistics is to choose a **level of significance** before the experiment that dictates how low the p-value should be before we reject the null hypothesis.

It is common to choose a significance level of 5% (or 1%).

We conclude that there is significant evidence against the null hypothesis if the p-value is less than or equal to 0.05 (0.01).

In our current example, the p-value is 0.00015 which is lower than 0.05.

In this case, we would conclude that

"there is signicant evidence against the null hypothesis at the 5% level"

Another way of saying this is that

"we reject the null hypothesis at the 5% level"

If the p-value for the test much larger, say 0.23, then we would conclude that

"the evidence against the null hypothesis is not significant at the 5% level"

Another way of saying this is that

"we cannot reject the null hypothesis at the 5% level"

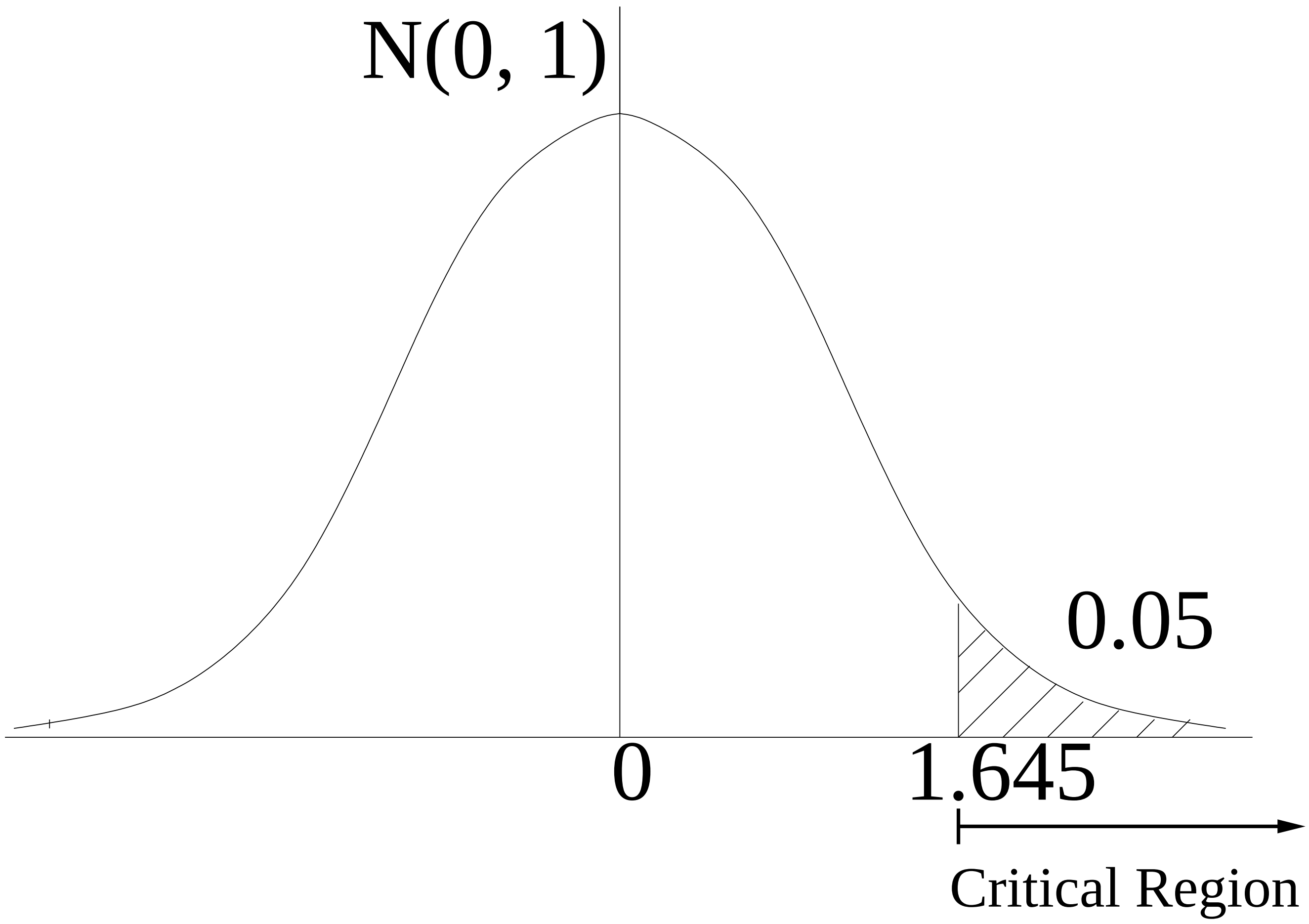
Test statistics and Critical Regions

Another way of thinking about this test is that we calculated a **test statistic**

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = 3.66$$

We can think of a **critical region** of values such that if the test statistic lies in this region then we will reject H_0. If the test statistic lies outside this region we will not reject H_0.

The Critical Region of the test



Overview of Hypothesis Testing

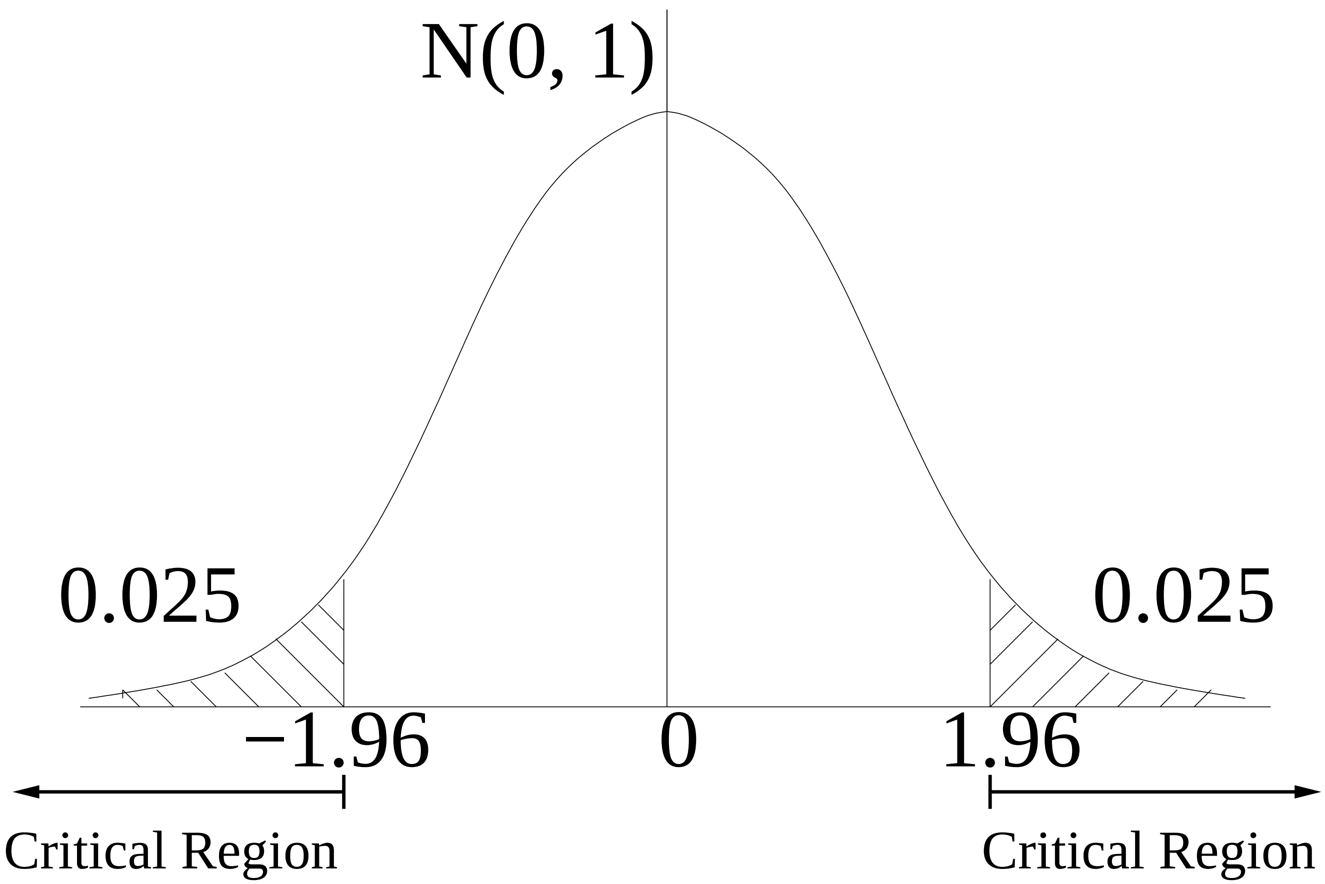
1. Begin with a **research (alternative) hypothesis** and decide upon a **level of significance** for the test.
2. Set up the **null hypothesis**.
3. Collect a sample of data.
4. Calculate a **test statistic** from the sample of data.
5. Calculate the **critical region** for the test.
6. Reject the null hypothesis if the test statistic lies in the **critical region**. Otherwise, retain the null hypothesis.

One and two-tailed tests

Our alternative hypothesis might have been

$$H_0 : \mu = 3000\text{g}$$
$$H_1 : \mu \neq 3000\text{g}$$

As before we would calculate our test statistic as 3.66. In this case we allow for the possibility that the mean value is less than 3000g by setting our critical region to be lowest 2.5% and highest 2.5% of the distribution.



This is an example of a **two-sided test** as opposed to the previous example which was a **one-sided test**.

Two sample test for a difference between two means

Suppose our research hypothesis is that the mean birth weight of boys is greater than mean birth weight of girls.

Suppose we know that the standard deviation of boys weights is 500g and the standard deviation of girls weights is 400g.

We want to test our research hypothesis using a significance level of 5%

Our null and alternative hypotheses are

$$
\begin{aligned}
\mathrm{H}_0 &: \mu_{boys} = \mu_{girls} \\
\mathrm{H}_1 &: \mu_{boys} > \mu_{girls}
\end{aligned}
$$

and we set our level of significance to be 5%. This dictates that we will carry out a one-tailed test.

In this example we will assume that we collected the Babyboom dataset. That is we have $n_{boys} = 26$ boys and $n_{girls} = 18$ girls.

Under the null hypothesis we know that

$$\bar{X}_{boys} \sim \mathrm{N}\Big(\mu, \frac{500^2}{26}\Big)$$

$$\bar{X}_{girls} \sim \mathrm{N}\Big(\mu, \frac{400^2}{18}\Big)$$

$$\Rightarrow \bar{X}_{boys} - \bar{X}_{girls} \sim \mathrm{N}\Big(0, \frac{500^2}{26} + \frac{400^2}{18}\Big)$$

So for our test statistic we use

$$Z = \frac{\bar{X}_{boys} - \bar{X}_{girls}}{\sqrt{\frac{500^2}{26} + \frac{400^2}{18}}} \sim \mathrm{N}(0, 1)$$

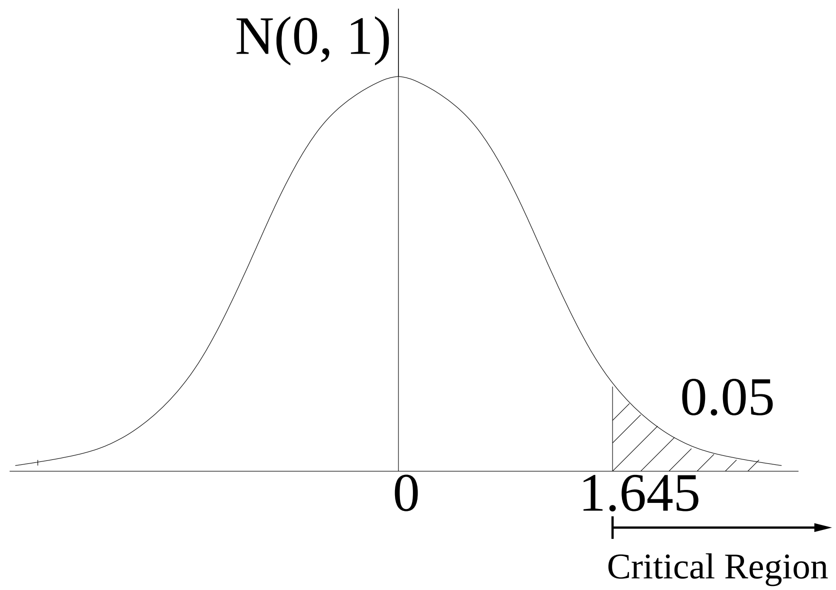
For our dataset we have $\bar{x}_{boys} = 3375.308$ and $\bar{x}_{girls} = 3132.444$ so we can calculate the test statistic as

$$Z = \frac{3375.308 - 3132.444}{\sqrt{\frac{500^2}{26} + \frac{400^2}{18}}} = 1.785$$

In general, to test for a difference between two means (with σ_1 and σ_2 known) from n_1 and n_2 observations from the two groups we use the test statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathbf{N}(0, 1)$$

The critical region of the test at the 5% level is $Z > 1.645$



The test statistic lies in the critical region so we conclude that is there is significant evidence against the null hypothesis at the 5% level of significance.

One sample test for a proportion p

Suppose that a university claims to admit equal numbers of state and public school students.

We have a research hypothesis that the university tends to admit more public school students so we interview 500 first year students and discover that 267 came from public schools.

We want to test our hypothesis at the 5% level.

First we write down our null and alternative hypotheses regarding the population proportion p of public school students

$$H_0 : p = 0.5$$
$$H_1 : p > 0.5$$

Using our sample of data we can obtain an estimate of p as

$$\hat{p} = \frac{267}{500} = 0.534$$

In this situation the test statistic used is

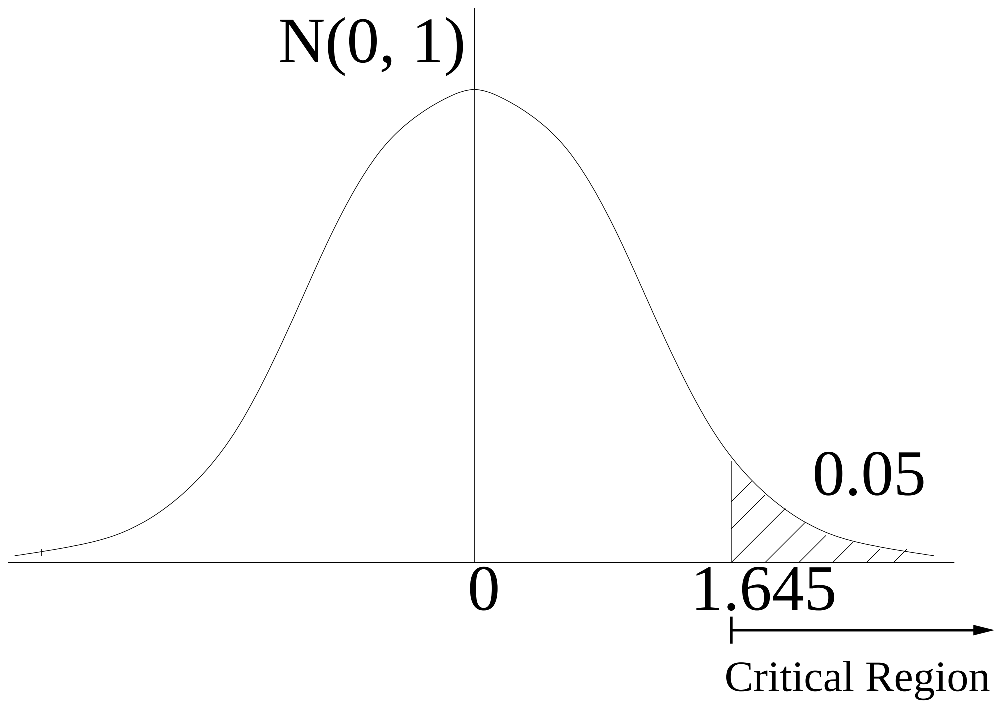
$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim \mathbf{N}(0, 1)$$

where p is the proportion dictated by the null hypothesis H_0 and n is the size of our sample.

In our example, the value of the test statistic is

$$Z = \frac{0.534 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{500}}} = 1.52$$

Again the critical region of the test is $Z > 1.645$



In this example the test statistic does not lie in the critical region so we conclude that the evidence against the null hypothesis is *not* significant at the 5% level.

Two sample test for a difference between proportions

Suppose that a newspaper carries out a poll in two cities A and B to ascertain the proportion of people who think the Prime Minister is doing a good job.

In city A 336 out of 600 people support the PM. In city B 656 out of 1000 people support the PM.

The newspaper would like to test the hypothesis that the proportion of people who think the PM is doing a good job differs between the cities.

First we write down our null and alternative hypotheses regarding the population proportions p_1 and p_2 of PM support in cities A and B

$$\mathrm{H}_0 : p_1 = p_2 (= p)$$
$$\mathrm{H}_1 : p_1 \neq p_2$$

Using our sample of data we can obtain an estimate of p_1 and p_2 as

$$\hat{p_1} = \frac{336}{600} = 0.56$$
$$\hat{p_2} = \frac{656}{1000} = 0.656$$

This is an example of a two-sided test.

In this situation the test statistic used is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathbf{N}(0, 1)$$

where n_1 and n_2 are the two sample sizes and p is either

(a) known from prior knowledge

(b) estimated from the data as $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

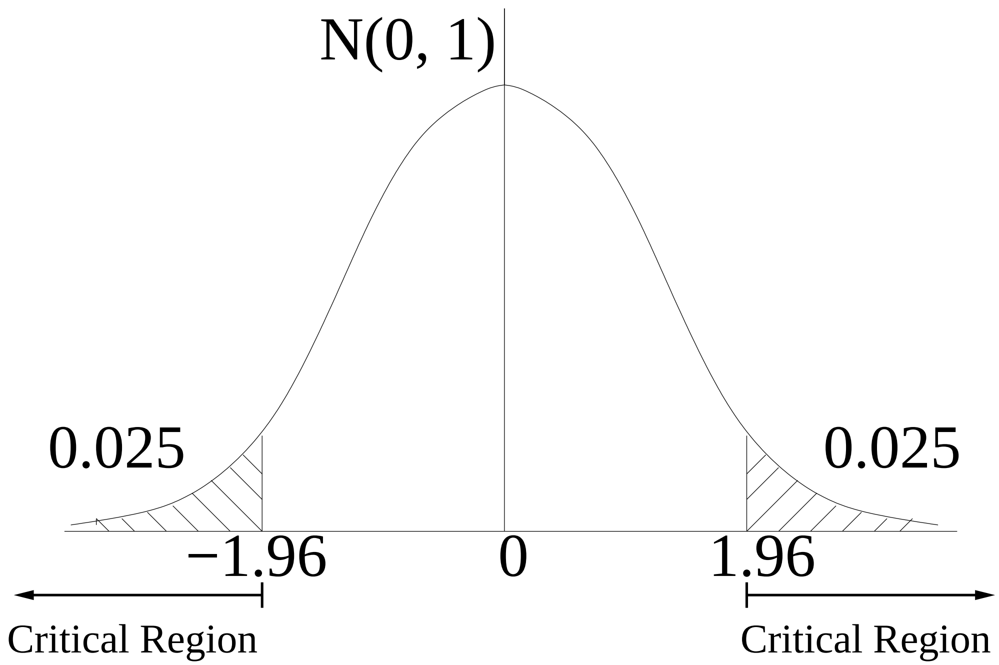
In our example, the value of p is estimated by

$$\hat{p} = \frac{600 \times 0.56 + 1000 \times 0.656}{600 + 1000} = 0.62$$

The test statistic is

$$Z = \frac{0.56 - 0.656}{\sqrt{0.62 \times (1 - 0.62)\left(\frac{1}{600} + \frac{1}{1000}\right)}} = -3.83$$

The critical region of the test is $Z < -1.96$ or $Z > 1.96$



In this example the test statistic lies in the critical region so we conclude that there is significant evidence against the null hypothesis at the 5% level.