

# Making decisions about populations using observed data.

Linda Angulo Lopez

09/12/2020

Necessary Background: Statistical methods for dealing with large & small datasets.

- An important concept in hypothesis testing is the **NULL hypothesis**, usually denoted as  $H_0$ : This is the hypothesis that **represents the status\_quo** that is which is assumed to be true. It's a baseline against which you're testing alternative hypotheses, usually denoted by  $H_a$ .
- Statisticalevidence is required to reject  $H_0$  in favor of the research or alternative hypothesis.

For example, if a **respiratory disturbance index (RDI)** of more than 30 events / hour is considered evidence of **severe sleep disordered breathing (SDB)**. Suppose that in a sample of 100 overweight subjects with other risk factors for SDB at a sleep clinic, the mean RDI ( $\bar{X}$ ) was 32 events / hour with a standard deviation (s) of 10 events / hour.

Then we can test the null hypothesis  $H_0$  that  $\mu = 30$ . Our alternative hypothesis  $H_a$  is  $\mu > 30$ . Here  $\mu$  represents the hypothesized population mean RDI.

We have two competing hypotheses,  $H_0$  and  $H_a$ , of which we'll have to pick one, using statistical evidence. That means we have **four possible outcomes** determined by what really is - the truth - and which hypothesis we accept based on our data. Two of the outcomes are correct and two are errors. It's correct to accept a true hypothesis or reject a false one & rejecting a true hypothesis or accepting a false one, is an error.

**A Type I error is possible REJECTS a TRUE null hypothesis  $H_0$  (An innocent person is convicted)**

**or a Type II error is possible ACCEPTS a FALSE null hypothesis  $H_0$ . (A guilty person is not convicted)**

**or we can evaluate the TRUTH correctly against  $H_a$  or against  $H_0$ .**

Since there's some element of uncertainty in questions concerning populations, we deal with probabilities. In our hypothesis testing we'll set the probability of making errors small. The probabilities of making these two kinds of errors are related. If you decrease the probability of making a Type I error (rejecting a true hypothesis), you increase the probability of making a Type II error (accepting a false one) and vice versa.

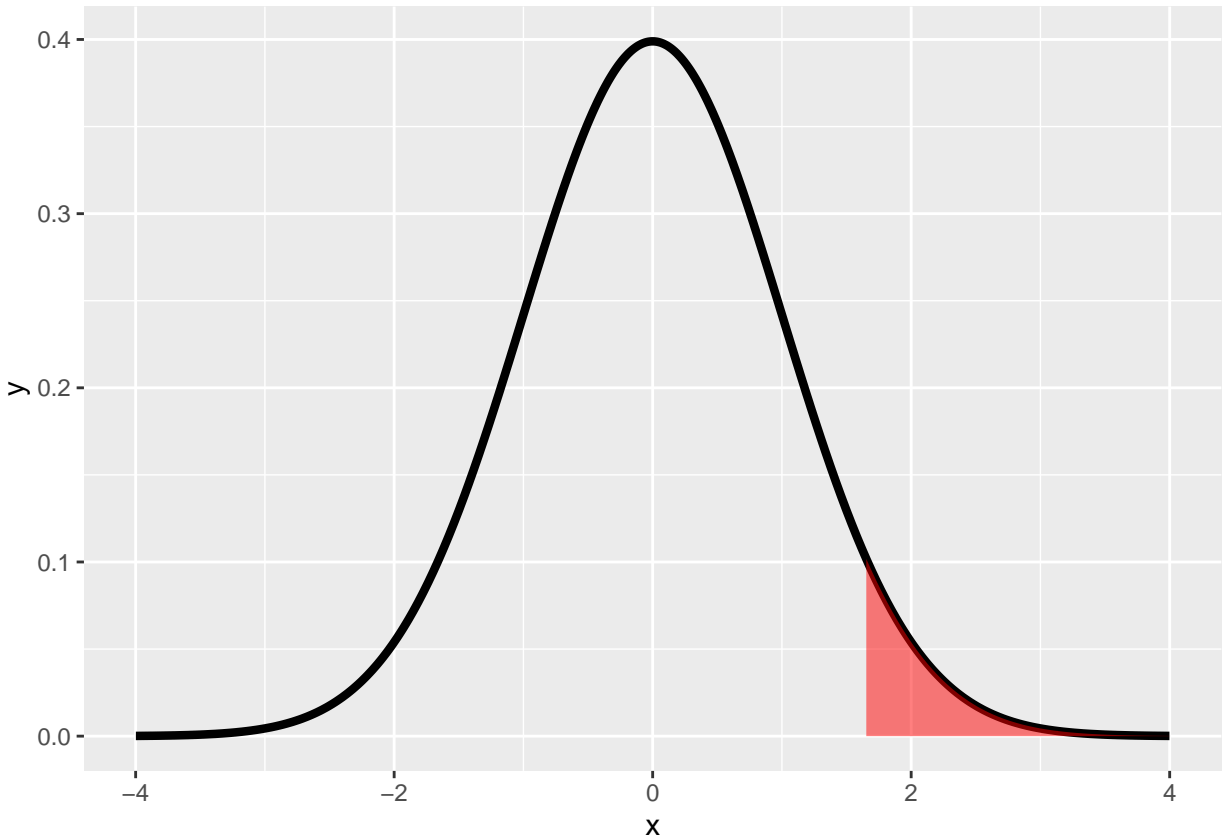
**compute a constant C, alpha:**

A reasonable strategy would reject the null hypothesis if our sample mean  $\bar{X}$  was larger than some constant C. 'We choose C so that the probability of a Type I error, alpha, is .05, or some other favorite constant. Many scientific papers use .05 as a standard level of rejection.

- If alpha was too low we would never reject the null hypothesis even if it's false.

For example the standard error of the mean, given by the formula  $s/\sqrt{n}$  of the [sleep example] (<https://rpubs.com/lindangulopez/702246>) where we had a sample of 100 subjects, our mean RDI ( $X'$ ) was 32 events / hour with a standard deviation ( $s$ ) of 10 events / hour.  $H_0$ ,  $X'$  is normally distributed with mean  $\mu=30$  and variance 1. We want to choose the constant  $C$  so that the probability that  $X$  is greater than  $C$  given  $H_0$  is 5%, that is,  $P(X > C | H_0)$  is 5%.

Here's a plot to show what we mean. The shaded portion represents 5% of the area under the curve and those  $X$  values in it are those for which the probability that  $X > C$  is 5%



The shaded portion represents 5% of the area under this normal density curve. To find out the smallest value  $X$  for which the area is shaded, we can use the R function `qnorm()` to find its standard normal, that is distance by standard deviations from the mean.

```
qnorm(.95)
```

```
## [1] 1.644854
```

So we see that the 95th percentile of a standard normal distribution is 1.645 standard deviations from the mean, so in our example we have to set  $C$  to be 1.645 standard deviations MORE than our hypothesized mean of 30, that is,  $C = 30 + 1.645 * 1 = 31.645$ . Recall that the variance and standard deviation equalled 1, as this is a normal distribution.

This means that if our OBSERVED sample mean  $X' \geq C$ , then it's only a 5% chance that a random draw from this  $N(30,1)$  distribution is larger than  $C$ . Recall that our observed mean  $X'$  is 32 which is greater than  $C=31.645$ , so it falls in that 5% region. - So we reject  $H_0$ .

## compute a Z score

Instead of computing a constant C as a cutpoint for accepting or rejecting the null hypothesis, we can simply compute a Z score, the number of standard deviations the sample mean is from the hypothesized mean. We can then compare it to quantile determined by alpha.

Recall that the Z score is  $(\bar{X} - \mu) / (s / \sqrt{n})$ . In this example  $\bar{X}=32$ ,  $\mu=30$ , and the standard error is  $10/\sqrt{100}=1$ .

The Z score is 2, from  $(32-30)/1$  and the quantile is 1.645, so since  $2 > 1.645$  that is the Z score is greater than the quantile of interest, we reject  $H_0$ :

**The general rule for rejection is if  $(\bar{X} - \mu) / (s / \sqrt{n}) > Z_{1-\alpha}$**

where: - our test statistic is  $(\bar{X} - \mu) / (s / \sqrt{n})$  which is standard normal - our test statistic has what mean of 0 and standard deviation of 1

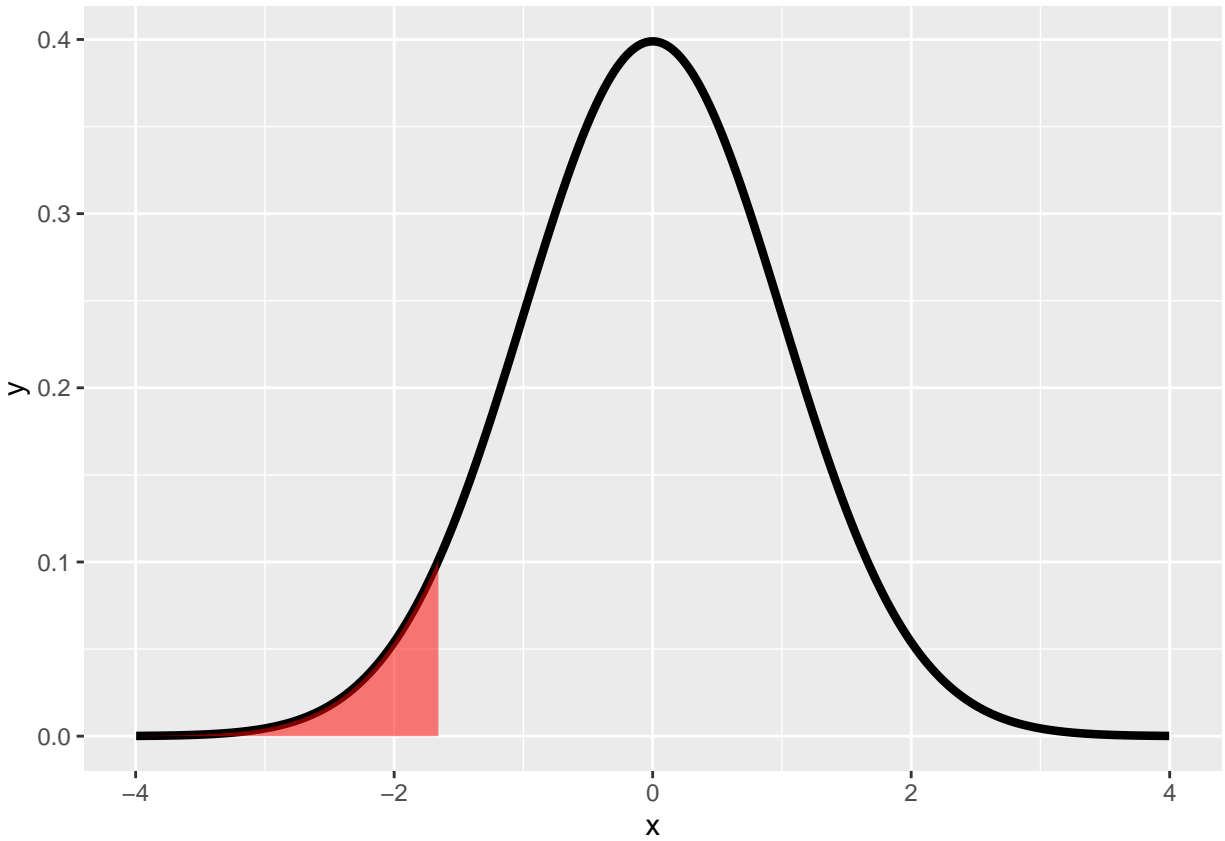
Our null hypothesis is that the population mean  $\mu$  equals the value  $\mu_0$  and  $\alpha=.05$ . Then this is the probability that we reject  $H_0$  if this is true.

## alternative hypotheses:

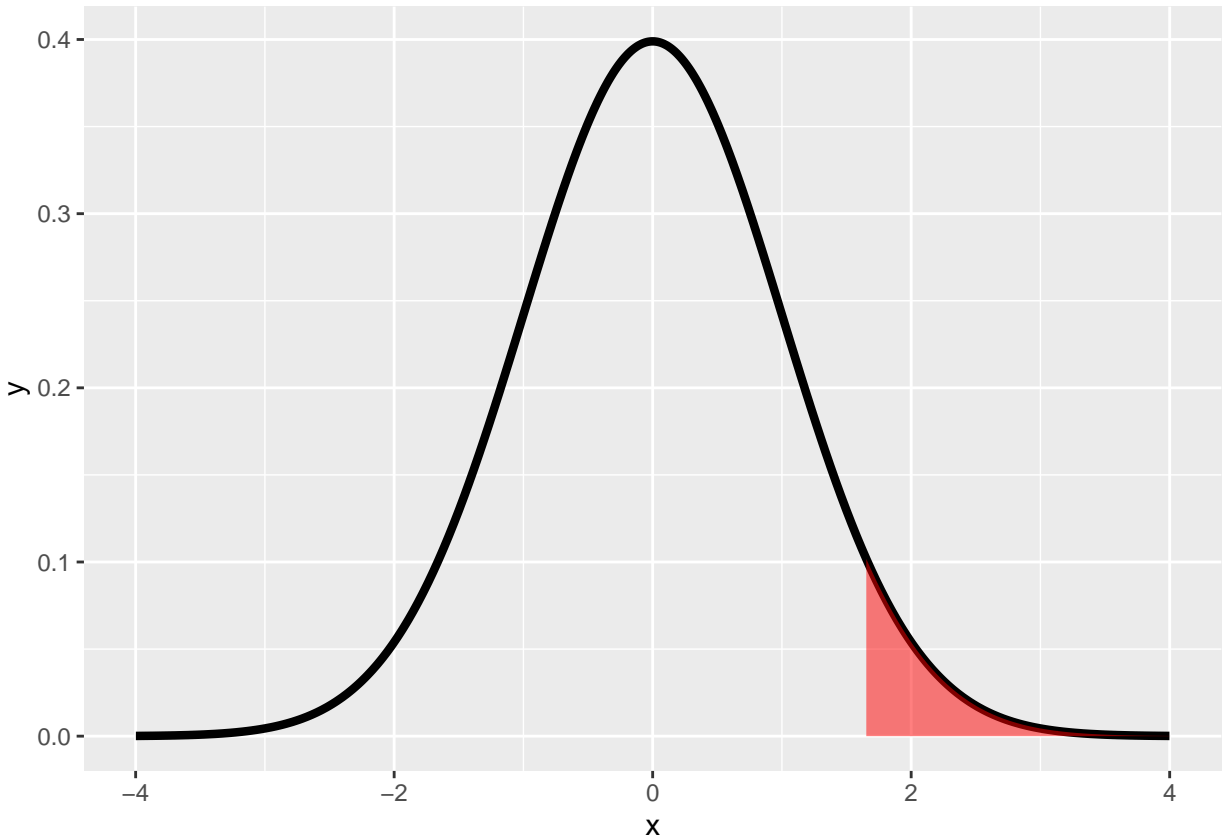
Suppose our first alternative,  $H_a$ , is that  $\mu < \mu_0$ . We would reject  $H_0$  and accept  $H_a$  when our observed sample mean is significantly less than  $\mu_0$ .

That is, our test statistic  $(\bar{X} - \mu) / (s / \sqrt{n})$  is less than  $Z_{\alpha}$ . Specifically, it is more than 1.64 standard deviations to the left of the mean  $\mu_0$ .

In this plot the shaded portion represents 5% of the area under the curve and those X values in it are those which are at least 1.64 standard deviations less than the mean. The probability of this is 5%. This means that if our sample mean fell in this area, we would reject a true null hypothesis,  $\mu=\mu_0$ , with probability 5%.



As the alternative hypothesis  $H_a$  is that  $\mu > \mu_0$ , if we accept  $H_a$ , that the true  $\mu$  is greater than the  $H_0$  value  $\mu_0$  we would want our sample mean to be significantly greater than  $\mu_0$ . This means that our test statistic  $(\bar{X} - \mu) / (s / \sqrt{n})$  is at least 1.64 std dev greater than  $\mu_0$ .

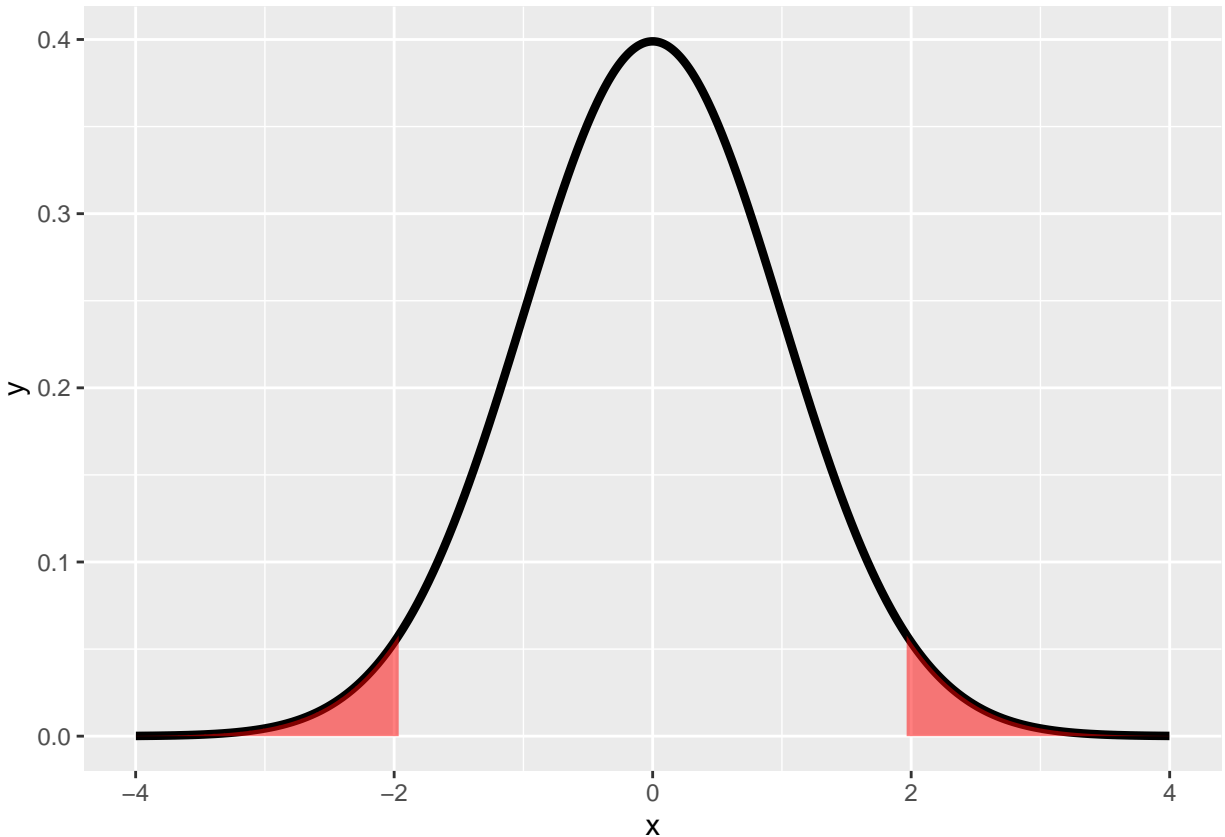


In the plot we see a shaded portion to the right it represents 5% of the area under the curve and those X values in it are those which are at least 1.64 standard deviations greater than the mean. The probability of this is 5%. This means that if our observed mean fell in this area we would reject a true null hypothesis, that  $\mu = \mu_0$ , with probability 5%.

Let's consider the alternative hypothesis  $H_a$  that  $\mu$  is simply not equal to  $\mu_0$ , the mean hypothesized by the null hypothesis  $H_0$ . We would reject  $H_0$  (and accept  $H_a$ ) when our sample mean is significantly different than  $\mu_0$ , that is, either less than OR greater than  $\mu_0$ .

If we want to stick with a 5% rejection rate, we divide it in half and consider values at **both tails, at the .025 and the .975 percentiles**. This means that our test statistic  $(\bar{X} - \mu) / (s / \sqrt{n})$  is less than .025,  $Z_{(\alpha/2)}$ , or greater than .975,  $Z_{(1-\alpha/2)}$ .

In the following plot, the shaded portion represents the 5% of the area composing the region of rejection. This time, though, it's composed of two equal pieces, each containing 2.5% of the area under the curve. The X values in the shaded portions are values which are at least 1.96 standard deviations away from the hypothesized mean.



Notice that if we reject  $H_0$ , either it was FALSE (and hence our model is wrong and we are correct to reject it) OR  $H_0$  is TRUE and we have made an error (Type I). The probability of this is 5%.

Since our tests were based on alpha, the probability of a Type I error, we say that we "fail to reject  $H_0$ " rather than we "accept  $H_0$ ". If we fail to reject  $H_0$ , then  $H_0$  could be true OR we just might not have enough data to reject it.

### fixing beta, the probability of a type II error:

We have not fixed the probability of a type II error (accepting  $H_0$  when it is false), which we call it beta. The term POWER refers to the quantity  $(1 - \text{beta})$  and it represents the probability of rejecting  $H_0$  when it's false. This is used to determine appropriate sample sizes in experiments.

### using the t distribution:

So far we've been talking about NORMAL distributions and implicitly relying on the CENTRAL LIMIT THEOREM (CLT), where the sample size is large. If we don't have a large sample size, we can use the t distribution which conveniently uses the same test statistic  $(\bar{X} - \mu) / (s / \sqrt{n})$  we used above. That means that all the examples we just went through would work exactly the same EXCEPT instead of using NORMAL quantiles, we would use t quantiles and  $n-1$  degrees of freedom.

For example, in the sleep disorder example, suppose our sample size=16, instead of 100, as before, the sample mean is  $\bar{X}=32$ , the standard deviation is  $s=10$ .  $H_0$  says the true mean  $\mu=30$ , and  $H_a$  is that  $\mu > 30$ . With this smaller sample size we use the t test, but our test statistic is computed the same way, namely  $(\bar{X} - \mu) / (s / \sqrt{n})$  computing in R:

```
t_sleep = (32-30)/(10/4)
t
```

```
## function (x)
## UseMethod("t")
## <bytecode: 0x000000000b7e1918>
## <environment: namespace:base>
```

```
# [1] 0.8
```

```
df = 16-1
df
```

```
## [1] 15
```

```
# [1] 15
```

Under  $H_0$ , the probability that the test statistic is larger than the 95th percentile of the t distribution is 5%. Use the R function qt with the arguments .95 and the correct number of degrees of freedom to find the quantile.

```
qt(.95, 15)
```

```
## [1] 1.75305
```

```
# [1] 1.75305
```

So the test statistic (.8) is less than 1.75, the 95th percentile of the t distribution with 15 degrees of freedom. This means that our sample mean (32) does not fall within the region of rejection since  $H_a$  was that  $\mu > 30$ , so: - we fail to reject  $H_0$ .

## two-sided test:

Now let's consider a two-sided test. Suppose that we would reject the null hypothesis if in fact the sample mean was too large or too small. That is, we want to test the alternative  $H_a$  that  $\mu$  is not equal to 30. We will - reject if the test statistic, 0.8, is either too large or too small.

As we discussed, we want the probability of rejecting under the null to be 5%, so we split the tails equally as 2.5% in the upper tail and 2.5% in the lower tail. Thus we reject if our test statistic is larger than the result of qt(.975, 15) or smaller than the result of qt(.025, 15).

You would expect qt(.975,15) to be bigger than qt(.95,15) and since we found that the test statistic was smaller than qt(.95,15) will it be smaller than qt(.975,15).

Generally, if you fail to reject the one sided test, you know that you will fail to reject the two sided. The left tail is less than 0, negative.

So the test statistic .8 failed both sides of the test, and - we fail to reject  $H_0$ .

## The R function `t.test`:

The test statistic tells us how many standard deviations the sample mean is from the hypothesized one.

- $t = (X' - \mu) / (s / \sqrt{n})$

With a csv file with the father\_son height data from John Verzani's UsingR website, read into a data structure `fs` see if fathers and sons have similar heights (our null hypothesis).

```
library(UsingR); data(father.son)

## Warning: package 'UsingR' was built under R version 4.0.3

## Loading required package: MASS

## Loading required package: HistData

## Warning: package 'HistData' was built under R version 4.0.3

## Loading required package: Hmisc

## Warning: package 'Hmisc' was built under R version 4.0.3

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 4.0.3

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

##
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':
##
##     cancer
```

```
t.test(father.son$sheight - father.son$fheight)
```



```
##
## One Sample t-test
##
## data: father.son$sheight - father.son$fheight
## t = 11.789, df = 1077, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 0.8310296 1.1629160
## sample estimates:
## mean of x
## 0.9969728
```

```
# One Sample t-test

#data: fs$sheight - fs$fheight
#t = 11.789, df = 1077, p-value < 2.2e-16
#alternative hypothesis: true mean is not equal to 0
#95 percent confidence interval:
# 0.8310296 1.1629160
#sample estimates:
#mean of x
#0.9969728
```

So the test statistic is 11.79 which is quite large it says there are almost 12 std errors between the sample and hypothesized means, so we REJECT the null hypothesis that the true mean of the difference was 0. We can test this by multiplying the t statistic,  $t=11.7885$ , by the standard deviation of the data divided by the square root of the sample size.

```
11.7885 * sd(father.son$sheight-father.son$fheight)/sqrt(1078)
```

```
## [1] 0.9969686
```

```
#[1] 0.9969686
```

Note the 95% confidence interval, 0.8310296 1.1629160, returned by t.test. It does not contain the hypothesized population mean 0 so we're pretty confident we can safely reject the hypothesis. This tells us that either our hypothesis is wrong or we're making a mistake (Type 1 error) in rejecting it.

If you set alpha to some value (say .05) and ran many tests checking alternative hypotheses against  $H_0$ , that  $\mu=\mu_0$ , the set of all possible values for which you fail to reject  $H_0$  forms the  $(1-\alpha)\%$  (that is 95%) confidence interval for  $\mu_0$ .

Similarly, if a  $(1-\alpha)\%$  interval contains  $\mu_0$ , then we fail to reject  $H_0$ .

## binomial distributions:

- Example of a family with 8 children, 7 of whom are girls and none are twins.

Let the null hypothesis be that either gender is equally likely, like an iid coin flip. So our  $H_0$  is that  $p=.5$ , where  $p$  is the probability of a girl. We want to see if we should reject  $H_0$  based on this sample of size 8. Our  $H_a$  is that  $p>.5$ , so we'll do a one-sided test, i.e., look at only the right tail of the distribution.

Let's set alpha, the level of our test, to .05 and find the probabilities associated with different rejection regions, where a rejection region  $i$  has at least  $i-1$  girls out of a possible 8.

```
mybin <- c(1.00000000, 0.99609375, 0.96484375, 0.85546875, 0.63671875, 0.36328125, 0.14453125, 0.03515625)
mybin
```

```
## [1] 1.00000000 0.99609375 0.96484375 0.85546875 0.63671875 0.36328125 0.14453125
## [8] 0.03515625 0.00390625
```

```
#[1] 1.00000000 0.99609375 0.96484375 0.85546875 0.63671875 0.36328125 0.14453125
#[8] 0.03515625 0.00390625
```

So  $\text{mybin}[1]=1.0$ , meaning that with probability 1 there are at least 0 girls, and  $\text{mybin}[2]=.996$  is the probability that there's at least 1 girl out of the 8, and so forth. The probabilities decrease as  $i$  increases. As  $\text{mybin}[7]=.144$  and  $\text{mybin}[8]=.035$  the least value of  $i$  for which the probability is less than .05 is  $i = 8$ . Here less than .05 so our sample falls in this region of rejection, so -we reject  $H_0$

So we note that in a 2-sided test our alternative hypothesis is that  $p$  is not equal to .5, and it's not obvious how to do this with a binomial distribution.

For discrete distributions such as binomial and Poisson, which don't rely on the CLT, we have to calculate by inverting 2-sided tests in R.

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

