# Economics 120A

# 8. Hypothesis Tests

Graham Elliott

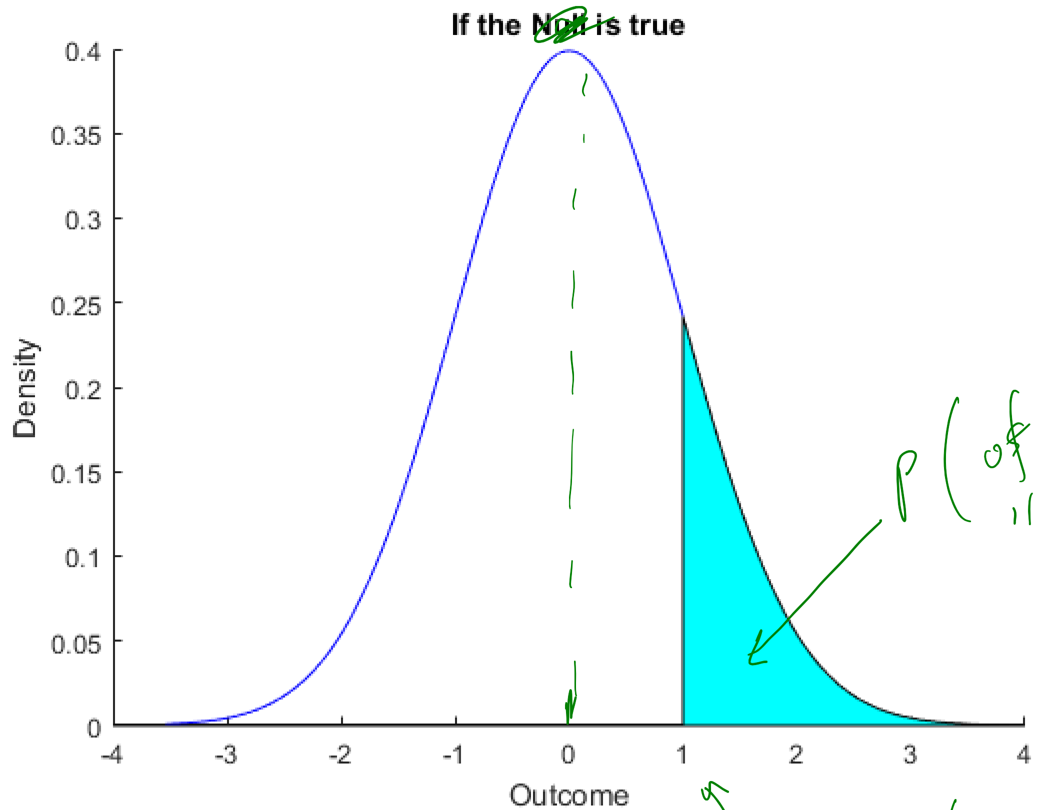
# Hypothesis Testing

The basic idea is to come up with a decision rule, where taking a function of the data that, depending on the value we get, we decide if our theory is correct or not.

For the sample mean and a theory for $\mu$ this will come down to the question of whether or not the sample mean is close to μ or not.

We will use our understanding of the distributions for the sample mean (or any other statistic) to understand what 'close' means.

# Hypothesis Testing – The approach in pictures

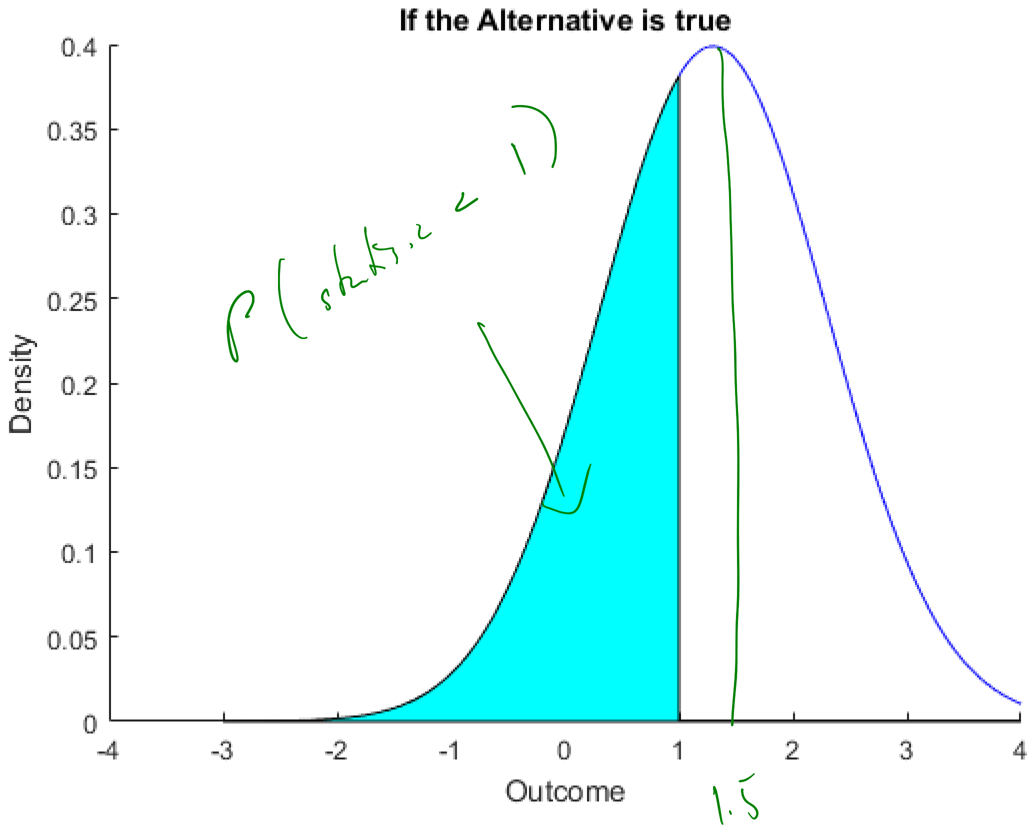



P (of seeing a more "extreme" value when my theory is true)

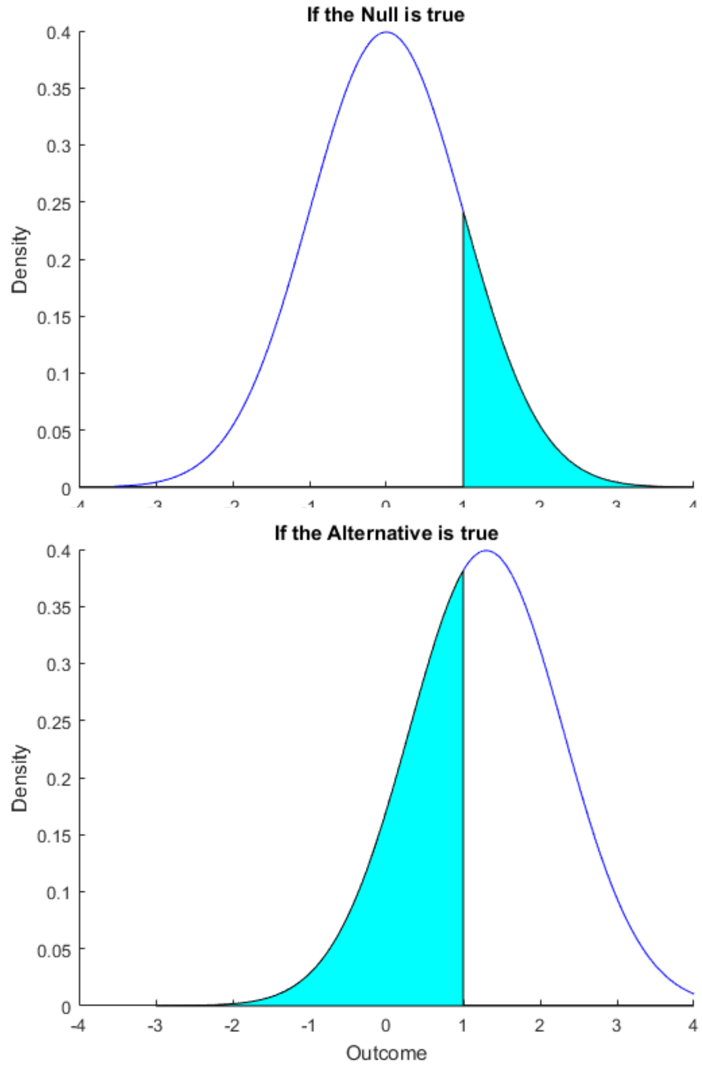
$\bar{x}$ - data

Theory says 0

# Hypothesis Testing – The approach in pictures



# Hypothesis Testing – The approach in pictures



# Examples

Example 1: Normal Example.

A manufacturer has a cereal box filler machine that is supposed to put 16 ounces of cereal in each box. The manufacturer knows the machine is not perfect and that the actual distribution of the filled boxes is normal with a standard deviation of .4

The producer takes a random sample of 25 boxes and finds that the boxes have an average weight of 16.2 ounces. The question the manufacturer may want to ask is, if the machine is overfilling and should be adjusted.

Example 2: Efficient markets hypothesis
This says that average returns $E[X_i] = 0$ (in the absence of risk). The hypothesis says nothing about the size of the variance, but assume $X_i \sim (0, \sigma^2)$.

Our sample estimate for average returns, is 0.02 and our sample estimate for the variance is $s^2$=0.05 We have n = 61. We would like to know if x = 0.02, which is not zero, refutes the hypothesis or not.

# Examples

Example 3: Proportions

This is based on a CNN poll on cloning. The question asked if respondents approved of human cloning for medical research purposes. The sample size was 507 and the response rates were

| | |
|---|---|
| Approve | 54% |
| Disapprove/No Idea | 46% |

We would be interested in the possibility that less than (or equal to) 50% of the people in the country approve.

# Classical Approach

Frequentist

The first step in the classical approach is to work out which values for μ suggest that our theory is correct and which values suggest that our theory is incorrect.

We have

| | | |
|---|---|---|
| Null Hypothesis | $H_0$ | what our theory says is true |
| Alternative Hypothesis | $H_1$ | what suggest our theory is not true |

$H_0$ : values for $\mu$ where theory is true

$H_1$ : values for $\mu$ where theory is not true

We have to be careful of these because the choices change the way we do the test.

# Examples

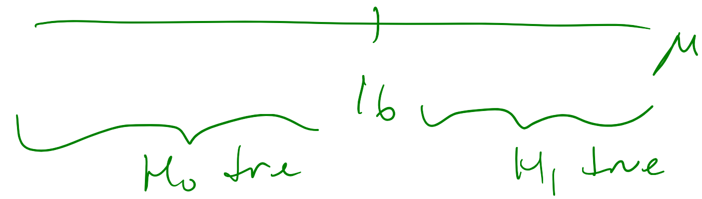
Example 1:

What values for μ correspond to the null and alternative?

$H_0: \mu \leq 16$

$H_1: \mu > 16$

one sided alternative

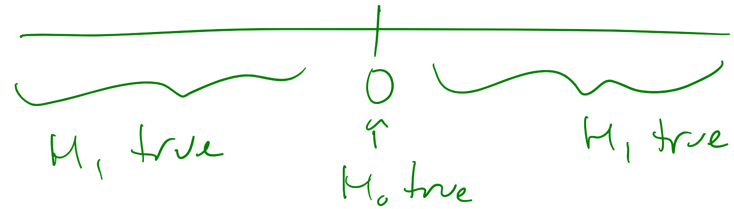



Example 2:

$H_0: \mu = 0$

$H_1: \mu \neq 0$

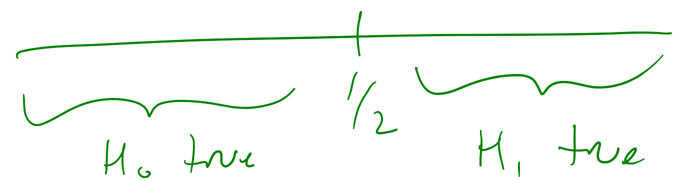
2 sided alternative




Example 3:

$H_0: \pi \leq \frac{1}{2}$

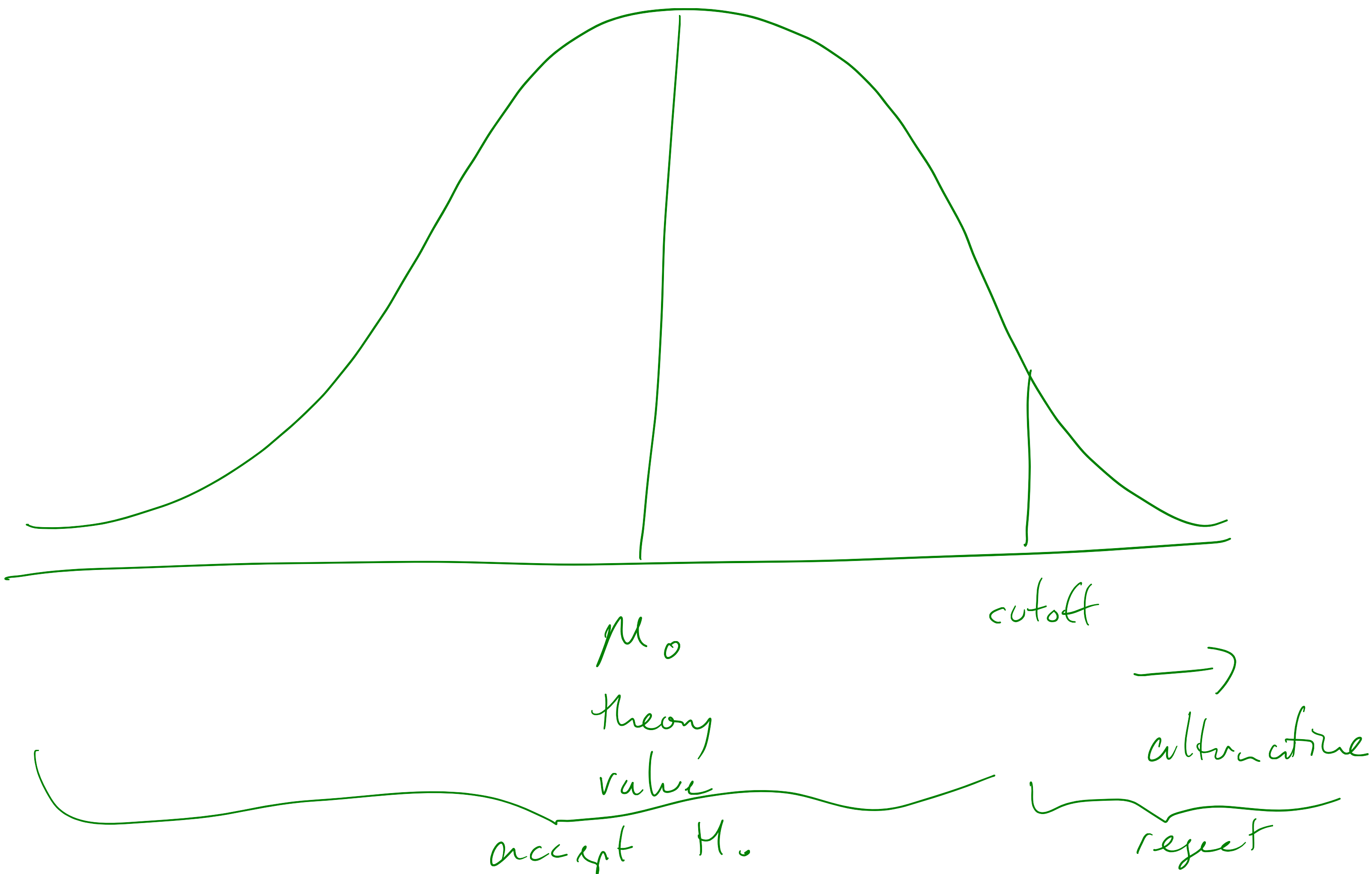
$H_1: \pi > \frac{1}{2}$



# Any Choice of a cutoff value results in errors

1 sided alternative

Picture in Example 1.



# Classification of the errors.

We can write the errors in a Table.

| | Choose Null | Choose Alternative |
|---|---|---|
| Null True | Correct | Type I error |
| Alternative True | Type II error | Correct |

P[Type I error] "tolerable"

The classical solution is to set the chance of rejecting a true hypothesis to be at some small 'tolerable' level, and accept whatever error we get for falsely choosing the null when the alternative is true.

The standard choice is 5%, but 10% or 1% are occasionally considered.

# Any Choice of a cutoff value results in errors

Picture in Example 1.

# Example 1

We determined that the null and alternative hypotheses were

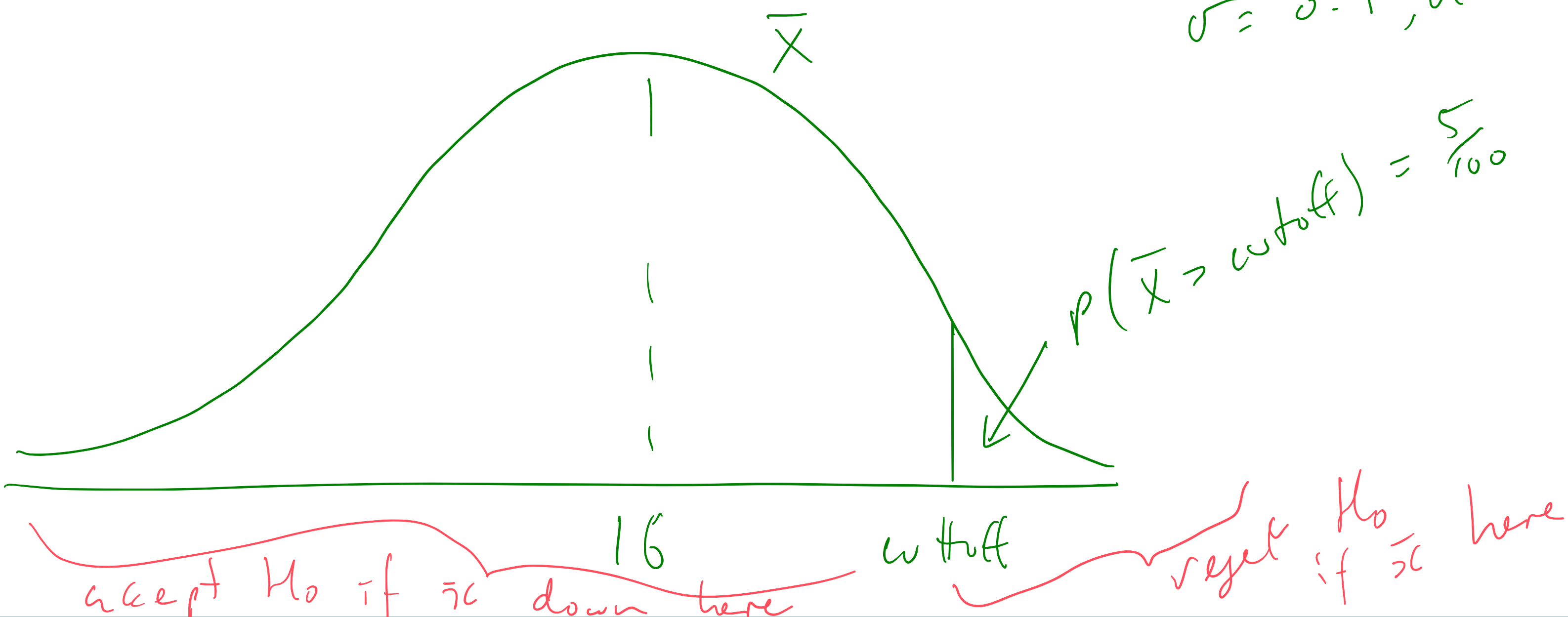
| | |
|---|---|
| $H_0$ | $\mu=16$ |
| $H_1$ | $\mu>16$ |

$(\mu \leq 16)$

"hypothesize $\mu=16$"

$\sigma = 0.4$, $n=25$

Draw it: is our estimate of 16.2 close to 16 or too far above?



# Example 1

We can easily compute the chance that if our null were true, we would get a number of 16.2 or more.

We have $P[\bar{X} > 16.2] =$

$$P\left[\frac{\bar{X}-16}{0.4/\sqrt{25}} > \frac{16.2-16}{0.4/\sqrt{25}}\right]$$

$$= P[Z > 2.5]$$

$$= 0.006 \quad \left(\frac{6}{10}\text{'s of } 1\%\right)$$

P[Z.2.5]=0.006

# An alternative way of doing the same thing

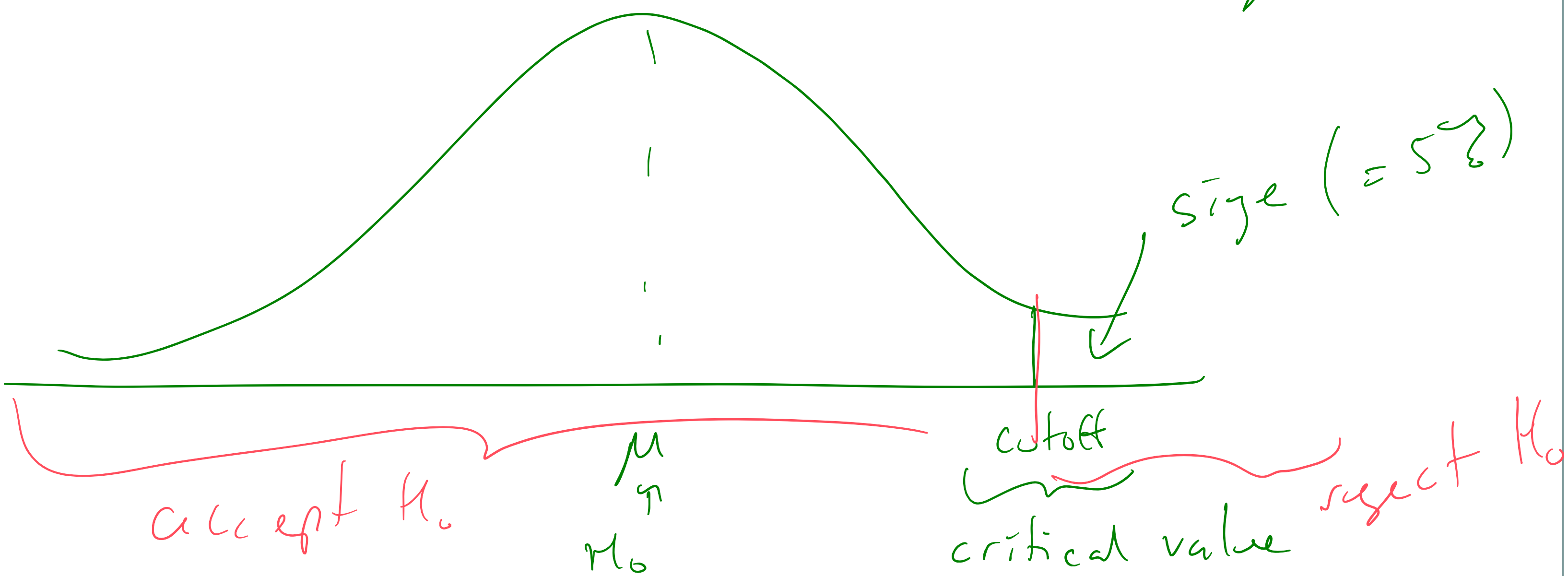
size

Step 1: Set up the problem.

Work out the null and alternative, choose which values for μ reject our null.
Choose a probability (we will call this size) of a Type 1 error.
This defines a 'critical' value which defines whether we reject or accept for a standard normal distribution.

1 sided test, reject for large values

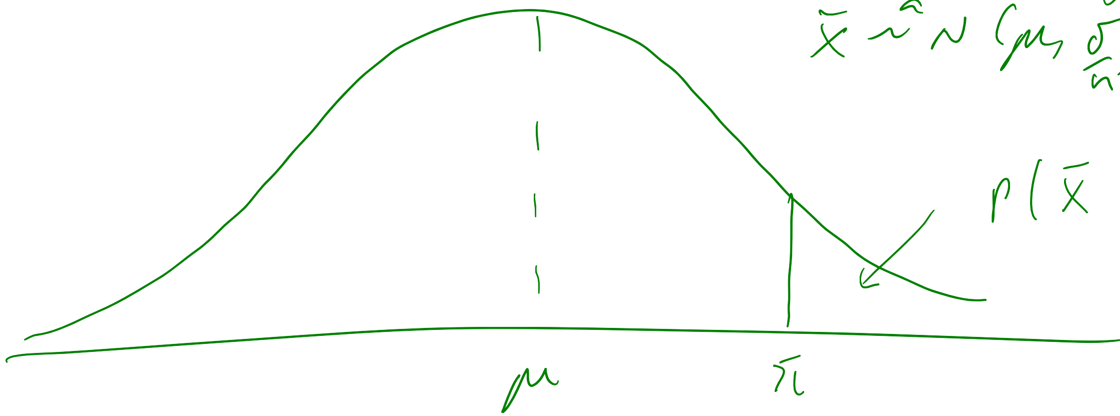
In pictures:



How we have been doing things

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

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

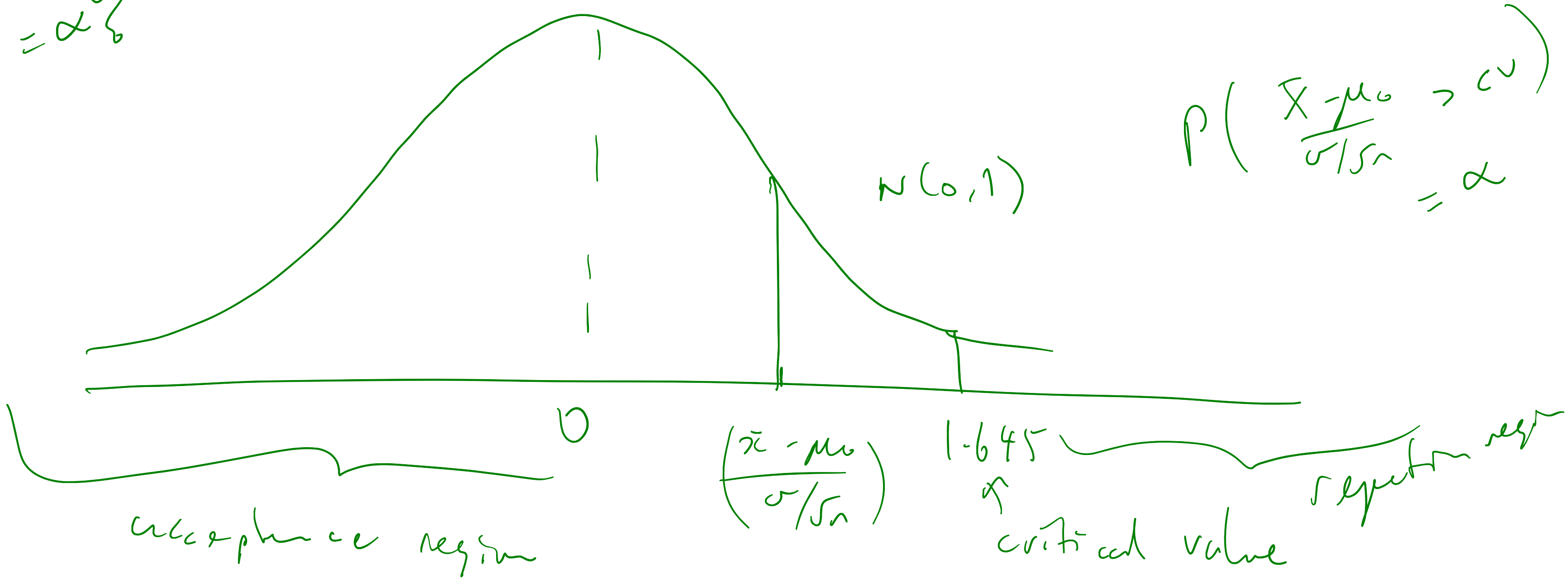



$$t = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

size $= \alpha$?

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > cv\right) = \alpha$$

$N(0,1)$



# An alternative way of doing the same thing

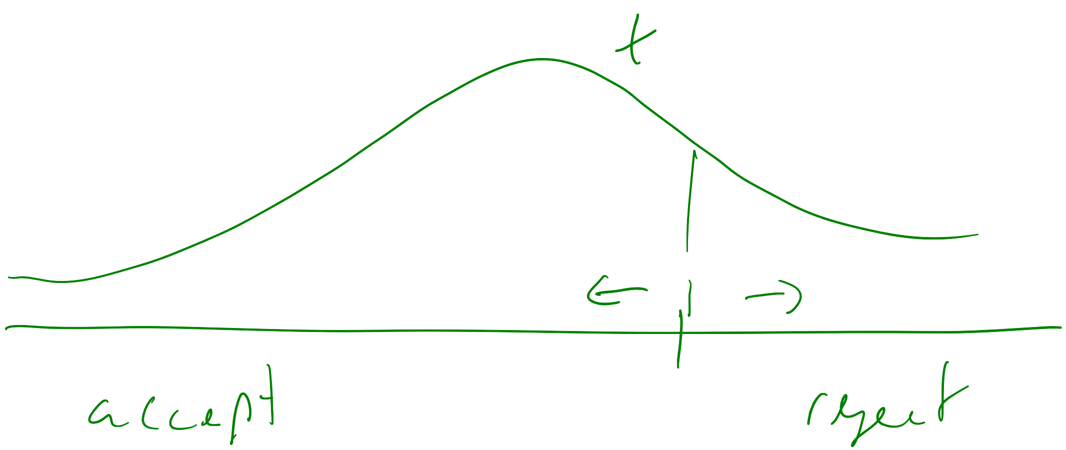
Step 2: Construct a t-statistic

$\bar{x} \quad - \quad \mu_0$

$$t = \frac{\textit{estimated value} - \textit{null value}}{\textit{standard error of estimated value}} \qquad \sigma/\sqrt{n}$$

For $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ this is going to be $t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

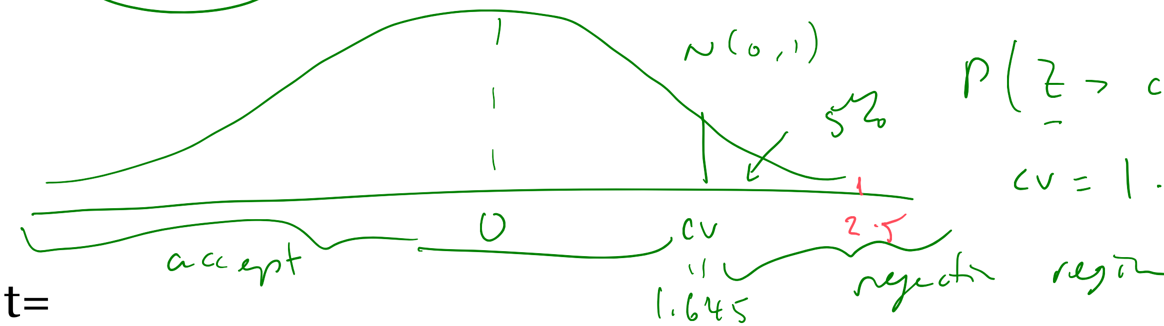
Step 3: Make the decision.
If the t value falls in the accept region, we accept the null (we think the theory is still true), if not we reject the theory.



# Example 1

$H_0: \mu \le 16 \qquad H_a: \mu > 16$

Step 1: choose size to be 5%. We already worked out the null and alternative hypotheses.




$P(Z > cv) = 5\%$

$cv = 1.645$

Step 2: we have t=

$\bar{x} = 16.2 \qquad \sigma = 0.4 \qquad n = 25$

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{16.2 - 16}{0.4/\sqrt{25}} = 2.5$$

(i.e reject $H_0$)

Step 3: our decision is that the machines are overfilling boxes

# When the variance and/or distribution of X is unknown

The sample mean of normal random variables with known variances is normal, otherwise we do not have this result.

If we knew the variance but $(X_1, X_2, \ldots, X_n)$ are not normally distributed, we still have the result from the CLT (For a VSRS) that $\bar{X}_n \sim^a N(\mu, \sigma^2/n)$.

So we can go through the same steps as before, but now our test is only approximately of size α instead of exactly.

This is to say that now

$$t = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim^a Z = N(0,1)$$

All of the steps are the same.

# Variance unknown

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Most likely we do not know the variance or the distribution.

In this case we could consider replacing the unknown $\sigma^2$ with our estimate $s^2$.

Now the second step (centering and standardizing) only changes with this replacement. But does the approximate normality still hold?

The answer is that so long as $s^2$ is a consistent estimator of $\sigma^2$, then yes, we still have approximate normality in large samples.

Consider:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right) \cdot \left(\frac{\sigma}{s}\right)$$

$Z$ by CLT

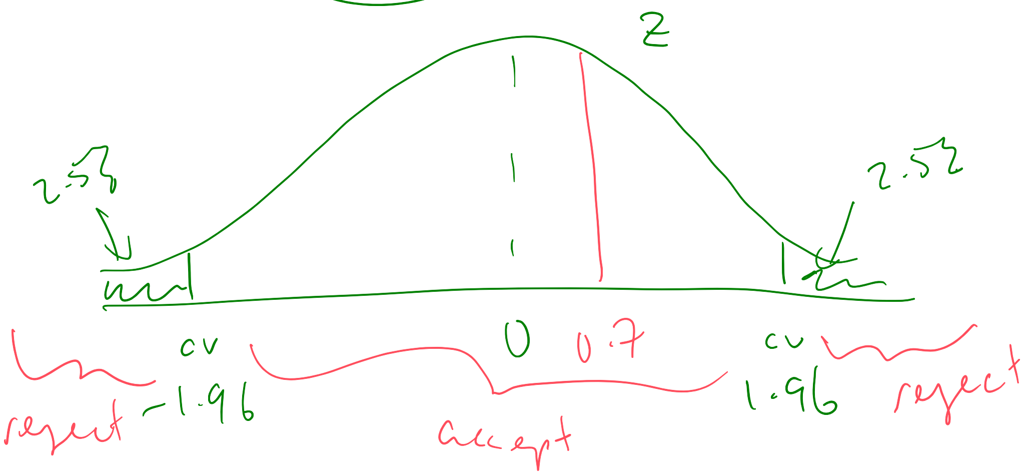
$s^2 \xrightarrow{p} \sigma^2$

so $\frac{\sigma}{s} \xrightarrow{p} 1$

# Example 2.

$H_0: \mu = 0$ vs $H_a: \mu \neq 0$

Step 1. Let size=5%, we have worked out the null and alternative, it is two sided.




$P(Z > cv) = 2.5\%$

$cv = 1.96$

Step 2. We have t= $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{0.02 - 0}{\sqrt{0.05}/\sqrt{61}} = 0.7$

$\bar{x} = 0.02$

$s^2 = 0.05$

$n = 61$

Step 3: Our decision is (t=0.7) that the efficient markets hypothesis holds true.

# Proportions

$X_i \sim \text{Bernoulli}(\pi)$

For proportions, our random variables are Bernoulli (π).

$X_i = \begin{cases} 0 & 1-\pi \\ 1 & \pi \end{cases}$

This means that

$\mu = \pi$

$\sigma^2 = \pi(1-\pi)$

So this is just a special case of the previous section where we will use the CLT and estimate of the variance to construct the t statistic.

Otherwise all the steps are the same.

$$\frac{\bar{X} - \pi}{\sqrt{\pi(1-\pi)/n}}$$

for our sample

# Proportions

$(n, \bar{x})$

$H_0: \pi = \pi_0$ vs $H_1: \pi \neq \pi_0$

There is an issue with the variance though – we have two ways to do it.

Method 1: Use our null hypothesis value for π, i.e. use $\pi_0(1-\pi_0)$.

$$t = \frac{\bar{x} - \pi_0}{\sqrt{\pi_0(1-\pi_0)}/\sqrt{n}}$$

Method 2: Use our estimated proportion for π, i.e. use $\bar{x}(1-\bar{x})$.

$x_i = (0, 1)$

Notice that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2 + \bar{x}^2 - 2\bar{x}x_i\right)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i + \bar{x}^2 - 2\bar{x}x_i\right)$$

$$= \frac{1}{n-1}\left(n\bar{x} + n\bar{x}^2 - 2\bar{x}(n\bar{x})\right)$$

$$= \frac{n}{n-1}\left(\bar{x} - \bar{x}^2\right) = \frac{n}{n-1}\bar{x}(1-\bar{x})$$

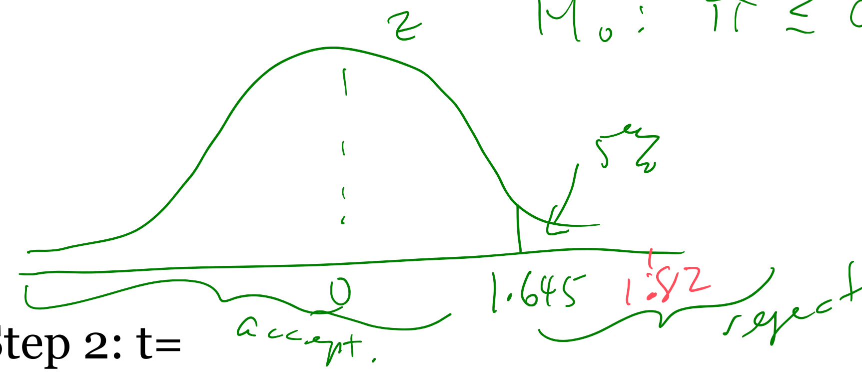
0.022

# Example 3:

$n = 507 \quad \pi = 0.54$

Step 1: set size = 5%, we already worked out the null and alternative.

$H_0: \pi \le 0.5$ vs $H_1: \pi > 0.5$




Step 2: t=

$$t = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi(1-\pi)/n}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.54 \times 0.46}{507}}} = 1.82$$

Step 3: Our decision is (t=1.82)

to reject that ½ or less are in favour, instead more than ½ are in favour

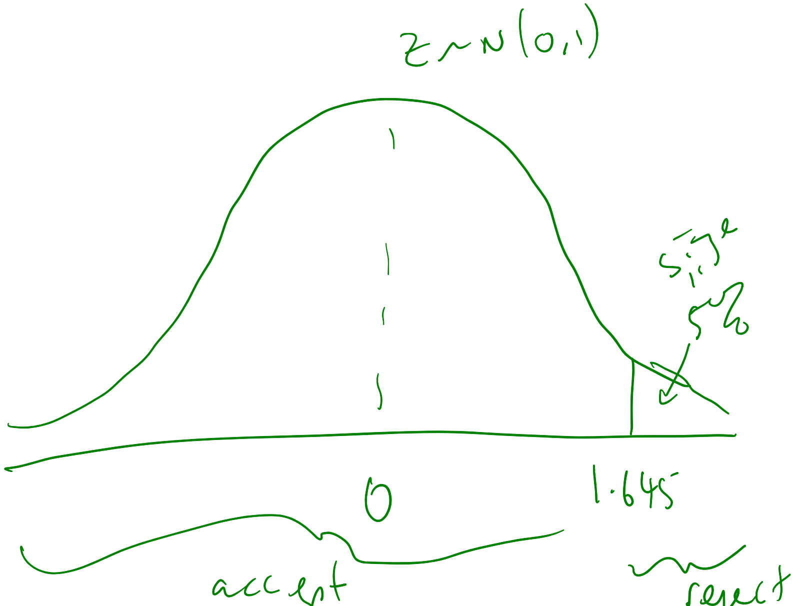
# p-values

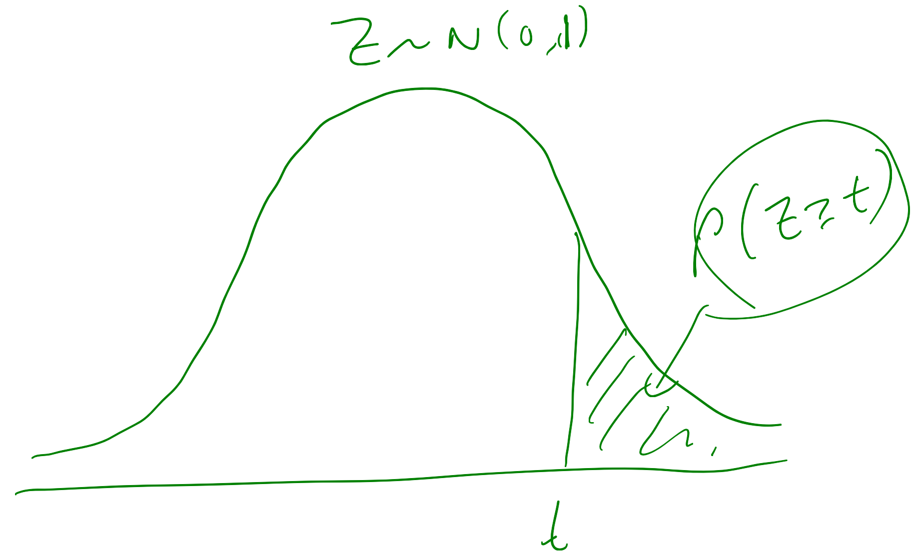
Suppose we have already computed the t statistic, which is either exactly or approximately a standard normal random variable.

$H_0: \mu \le \mu_0 \quad H_1: \mu > \mu_0$

For an upper tail one sided test, we have for a 5% test a critical value of 1.645.

We could also, as we have before, compute P[Z>t].




There is a clear relationship between the two of them.

# p-values

The second of these is the p-value approach.

For an upper tail one sided test we compute P[Z>t]. $H_1: \mu > \mu_0$

For a lower tail one sided test we compute P[Z<t]. $H_1: \mu < \mu_0$

These can be compared directly against size.

If the p-value is smaller than size, we know we have a t statistic that is more extreme than the critical value so we reject.

We interpret it as the chance of getting that value or a more extreme one.

# p-values

But what do we do for a two tail test.

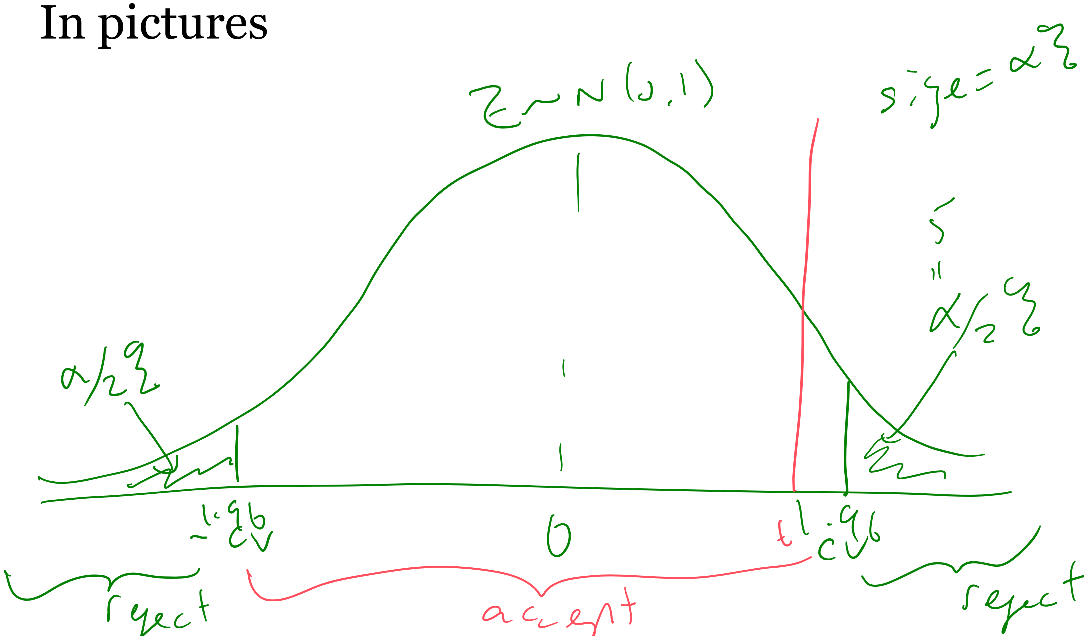
$H_1: \mu \neq \mu_0$

We can compute the probability that we have a more extreme value than the one we observed, i.e.

P[Z>t] if t>0
P[Z<t] if t<0.

$p(z>t) = 4\%$

In pictures



# p-values for two tail tests

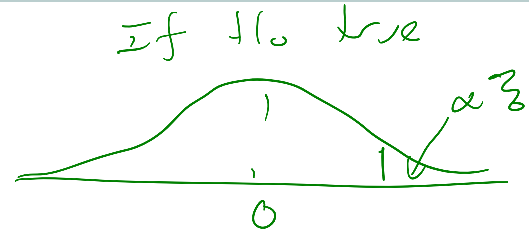
The problem is that we are comparing to the critical value for size divided by 2. So we could compare to this, however the convention is to simply double the size of the probability calculation for two tail tests.

So we compute the p-value as

2*P[Z>t] if t>0
2*P[Z<t] if t<0.

This way we can compare it directly to size as before.

# Power




In section 8.2 we saw the classical strategy.

By choosing size=probability of making a Type I error we define the test, which means that Type II error is fixed.

Recall that Type II error is the error where we fail to reject the null hypothesis when the alternative is true.

It is easier to think about this the other way around, i.e. what is the probability that we reject the null hypothesis when the alternative is true. This is what power gives us.

Probability we reject when $H_1$ is true

The power of a test is 1-P[Type II error].

Since the probability of this error depends on the value of the alternative, this is really a function of possible alternatives.

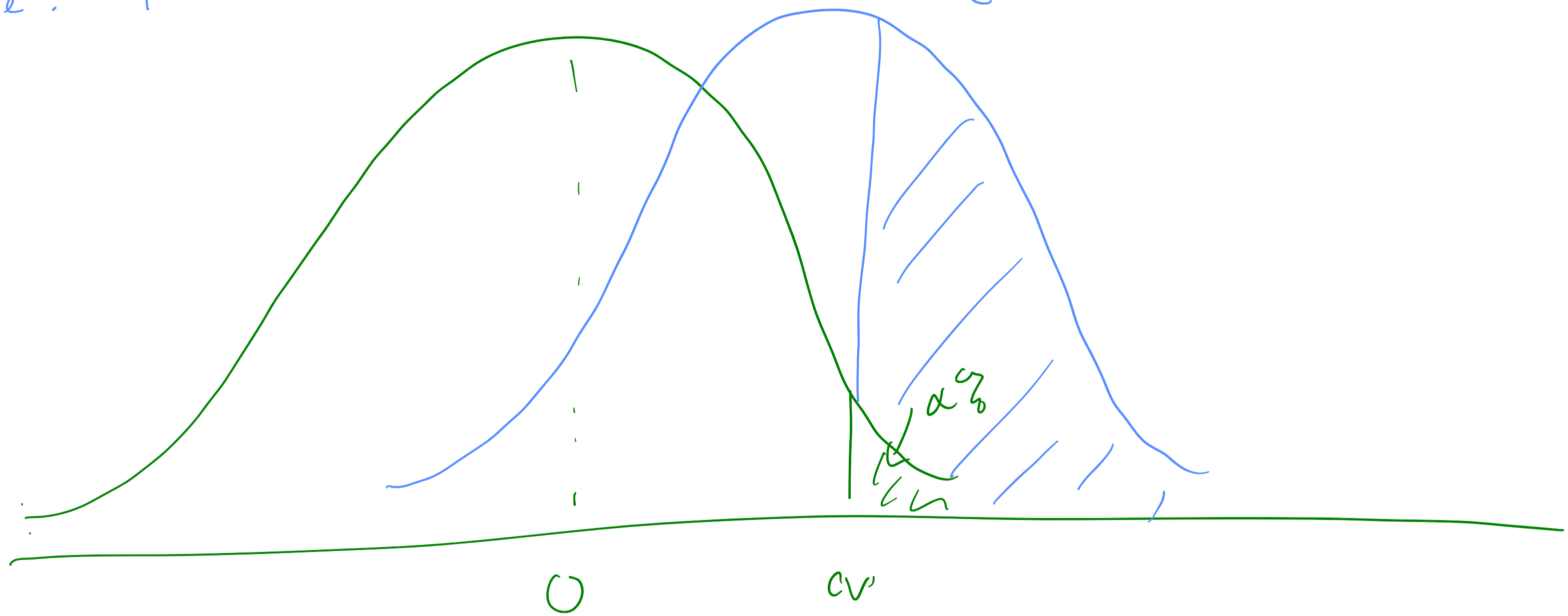
$H_0: \mu \le \mu_0$ $\quad H_1: \mu > \mu_0$

$H_0: \mu \le \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0$

Green: $H_0$ true

Blue: $H_1$ true

$Z \sim N(0,1)$ $\qquad$ $Z \sim N(?,1)$




Reject prob = $P(t \geq cv)$

# Example.

Suppose there is a drug that cures some problem 60% of the time. A new drug comes along, we can test

$$H_0 : \pi = 0.6$$
$$H_1 : \pi > 0.6.$$

We could do this test with a 5% size, which means that we would incorrectly think the new drug is better (supposing it is not) 5% of the time when we see an $\bar{x}$ (so t) that is too large.

But suppose instead that the new drug really is better, and has a cure rate of 70%. Would we reject very often?

Let the sample size be 100 (we need to know to compute the standard error).

$n = 100$

$P(t > cv)$ if $\pi = 0.7$

# Example

To compute the t-test for the null that $\pi=0.6$, we say that

$$t= \frac{\bar{\pi} - 0.6}{\sqrt{\bar{\pi}(1-\bar{\pi})/100}}$$

and compare to the standard normal.

Here we interpret this as ‘this is correct IF the true mean is 0.6”.

For power calculations, we do the computation under the assumption that the alternative is true, for example assuming that the true mean is now 0.7. But we still have t computed as though 0.6 was the center.

So now, instead of centering on 0.6 we center on 0.7.

# Example

We want to compute

5%

CV

$$P\left[\frac{\bar{x}-0.6}{\sqrt{\bar{x}(1-\bar{x})/n}} > 1.645\right] = P\left[\frac{\bar{x}-0.7+0.7-0.6}{\sqrt{\bar{x}(1-\bar{x})/100}} > 1.645\right]$$

$$= P\left[\underbrace{\frac{\bar{x}-0.7}{\sqrt{\bar{x}(1-\bar{x})/100}}}_{N(0,1)} + \frac{0.7-0.6}{\sqrt{\bar{x}(1-\bar{x})/100}} > 1.645\right]$$

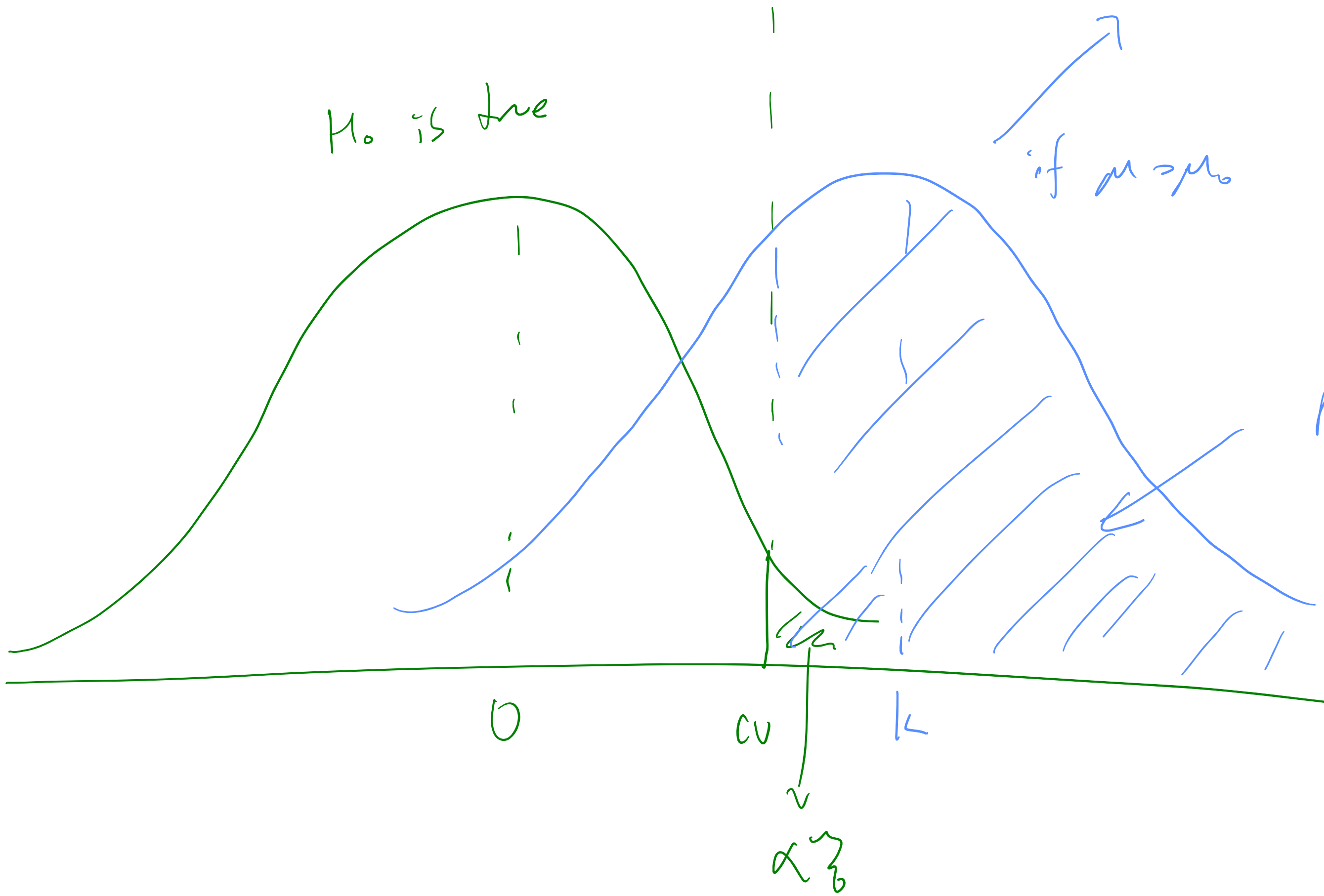
$$\simeq P\left[Z > 1.645 - \frac{0.7-0.6}{\sqrt{\bar{x}(1-\bar{x})/100}}\right]$$

P(Z>-0.4236)=66.41%

recall $\quad t = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ "what if" $\mu_0$ is true

$$= \frac{\bar{x} - \mu + \mu - \mu_0}{\sigma/\sqrt{n}} = \underbrace{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}_{Z \sim N(0,1)} + \underbrace{\frac{\mu - \mu_0}{\sigma/\sqrt{n}}}_{k}$$

$(\mu > \mu_0)$




$p\left[t > cv \text{ if } \mu > \mu_0\right]$

$=$ power at $\mu$.

# How big is power?

For many problems, power is very low meaning that although we do not reject in a hypothesis test, it is entirely possible that there are economically or scientifically meaningful models that could be true but would be found with low probability.

This turns out to be a very real problem with statistical studies.

# Conclusions

Hypothesis testing was simple a foundation of our question "is a sample mean of x likely given the model we think is true?' that we have been doing all along. At the same time, this is central to the scientific method of proposing models and examining them with data.

$$t = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Summary table for t tests

| | |
|---|---|
| \|t\|<1 | weak evidence against our theory, high chance of observing this randomly |
| 1<\|t\|<2 | stronger but not great evidence against our theory |
| 2<\|t\|<3 | evidence that our theory is unlikely |
| \|3\|<t | strong evidence against our theory |