

Power

February 16, 2021

Contents

- Definition of power
- Z-test example
- The 2x2 matrix of All Things That Can Happen
- $\Pr(\text{Type I error}) = \alpha$
- $\Pr(\text{Type II error}) = \beta$
- $\text{power} = 1 - \beta$
- Things that affect power
- Increasing effect size increases power
- Increasing sample size increases power
- Increasing α increases power
- Power goes down with two-tailed tests
- Calculating power for a t-test
- Example from the survey
- Power Curves
- Stats calculator
- Using R to calculate power for a t-test
- Questions
- Answers

Happy birthday to Jacquelyn Yang Shaw and Sophie Marie Frick!

Definition of power

The **power** of a hypothesis test is the probability of (correctly) rejecting the null hypothesis when it's false. In this tutorial we'll define what this means in the context of a specific example: a z-test on the mean of a sample of IQ scores. This first part should be a review for you.

Z-test example

Suppose you want to test the hypothesis that the mean IQ of students at UW have a higher IQ than the population that has a mean IQ of 100 points. We also know that the standard deviation of the IQ's in the population is 15.

You measure the IQ in 9 UW students and obtain a mean IQ of 106. First we'll determine if this mean is significantly greater than 100 using $\alpha = 0.05$.

To conduct this test we need to calculate the standard error of the mean and then convert our observed mean into a z-score:

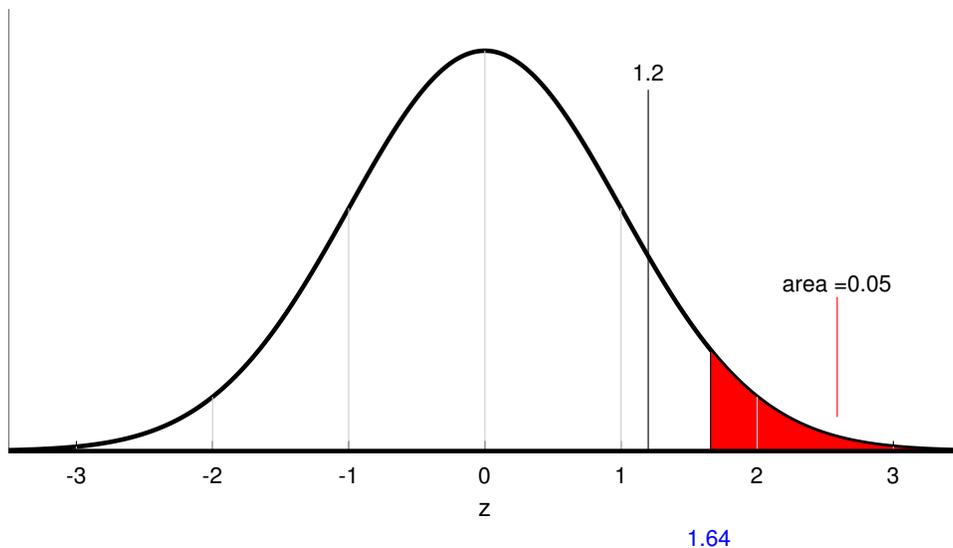
$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5$$

and

$$z = \frac{(\bar{x} - \mu_{hyp})}{\sigma_{\bar{x}}} = \frac{(106 - 100)}{5} = 1.2$$

With a one-tailed test and $\alpha = 0.05$, the critical value if z is $z_{crit} = 1.64$. Since our observed value of $z=1.2$ is less than the critical value of 1.64 we fail to reject H_0 and conclude that the mean IQ of UW students is not significantly greater than 100.

Here's our usual figure describing this outcome, with the rejection region shown in red:



We never know if we're making a correct decision because we never actually know whether H_0 is true or false (if we did know this, we wouldn't need to run the experiment). If we happen to reject H_0 it could be that H_0 is true and we just happened to grab a sample with a high mean IQ. This kind of mistake, when we accidentally reject H_0 is called a **Type I Error**.

If H_0 is true

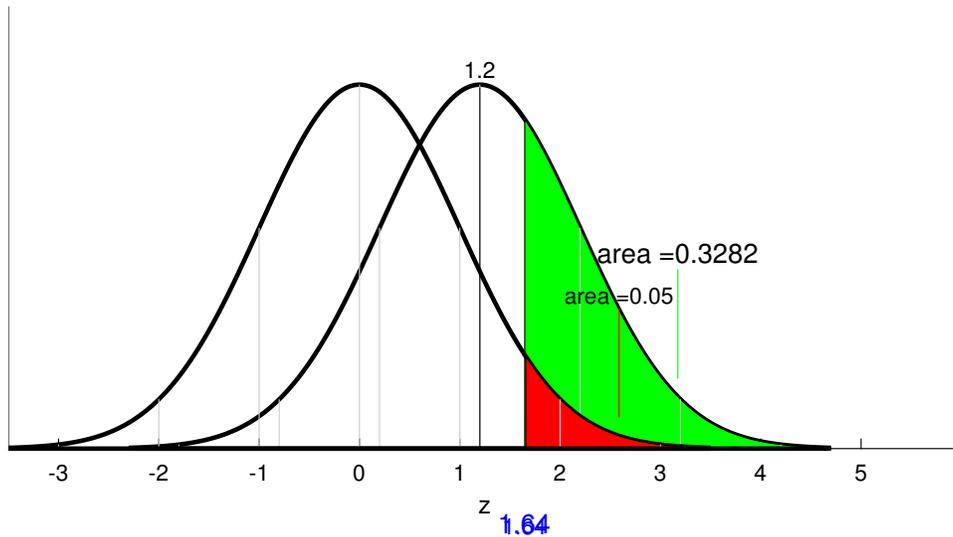
α is the probability of rejecting H_0 if it is true. So α is the probability of a Type I error, which is 0.05 in this example.

H_0 is true and we reject, then we've made a correct decision. Since we either reject or fail to reject H_0 , the probability of failing to reject H_0 is $1 - \alpha = 1 - 0.05 = 0.95$.

If H_0 is false

Remember, if H_0 is false then the probability of rejecting H_0 is the power of our test. How do we find power?

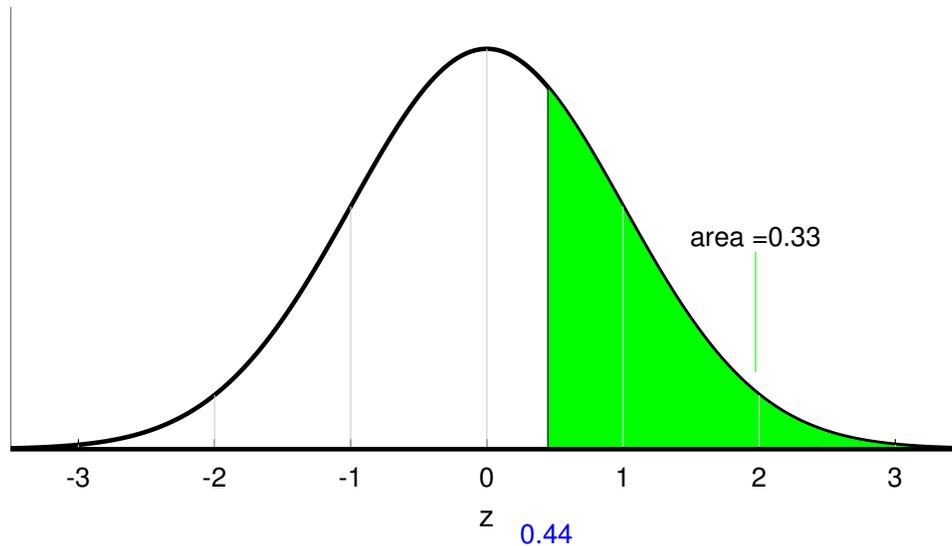
To calculate power we need to know the true mean of the population, which is different than the mean for H_0 . You probably appreciate that this is weird. If we knew the true mean then we wouldn't have to run an experiment in the first place.



Importantly, even though we're now drawing z-scores from the true distribution our decision rule remains the same: if our observed value of z is greater than the critical value ($z = 1.64$ in this example), then we reject H_0 . I've colored this region above $z = 1.64$ under the true distribution in green.

It should be clear that **this green area is the power of our hypothesis test**. It's the probability of, when drawing from the true distribution that we reject H_0 . You can see that the green area (power) is greater than the red area (α).

The green area is the area under the true distribution (which has a mean of $z_{true}=1.2$ and $\sigma=1$) above a value of $z_{crit} = 1.64$. If we subtract 1.2 from both the z_{true} and z_{crit} , power is equivalent to the area under the standard normal distribution ($\mu = 0, \sigma = 1$) above $z = 1.64 - 1.2 = 0.44$. If you look this up in the z-table (or z-calculator) you'll find that this area is 0.3282:



If the power of a test is 0.3282, it means that if the true mean of the population had an IQ of 106, then with our sample size of 9 and $\alpha=0.05$, the probability that we'd draw a mean that is significantly greater than 100 is 0.3282. You can think of the power as the probability of successfully detecting an effect that is different from the null hypothesis.

The 2x2 matrix of All Things That Can Happen

Out there in the world, the null hypothesis is either true or it is not true. To make a decision about the state of the world we run a hypothesis test and decide either to reject or to fail to reject the null hypothesis. Therefore, when we run a hypothesis test, there are four possible things that can happen. This can be described as a 2x2 matrix:

Of the four things, two are correct outcomes (green) and two are errors (red). Consider the case where the null hypothesis is H_0 is true. We've made an error if we reject H_0 . Rejecting H_0 when it's actually true is called a **Type I error**.

$$\Pr(\text{Type I error}) = \alpha$$

Type I errors happen when we accidentally reject H_0 . This happens when we happen to grab a mean that falls in the rejection region of the null hypothesis distribution. We know the probability of this happening because we've deliberately chosen it - it's α .

It follows that the probability of correctly rejecting H_0 is $1 - \alpha$.

$$\Pr(\text{Type II error}) = \beta$$

Type II errors happen when we accidentally fail to reject H_0 . This happens when we grab a mean from the 'true' distribution that happens to fall outside the rejection region of the

null hypothesis distribution. We have a special greek letter for this probability, β ('beta').

$$\text{power} = 1 - \beta$$

Remember, power is the probability of correctly rejecting H_0 . This is the opposite of a Type II error, which is when we accidentally fail to reject H_0 . Therefore, power = 1-Pr(Type II error) which is the same as power = $1 - \beta$.

We can summarize all of these probabilities in our 2x2 table of All Things that Can Happen:

True state of the world		
Decision	H_0 is true	H_0 is false
Fail to reject H_0	$1-\alpha$	Type II error: β
Reject H_0	Type I error: α	power: $1-\beta$

For our example, we can fill in the probabilities for all four types of events:

True state of the world		
Decision	H_0 is true	H_0 is false
Fail to reject H_0	0.95	0.67179
Reject H_0	0.05	0.32821

Things that affect power

Figures like the one above with overlapping null and true distributions show up in all textbook discussion of power. Google 'statistical power' and check out the associated images. It's an ubiquitous image because it's very useful for visualizing how power is affected by various factors in your experimental design and your data.

Looking at the figure, you should see that power will increase if we separate the two distributions. There are a variety of ways this can happen.

Increasing effect size increases power

The most obvious way is for us to choose a larger value for the mean of the true distribution. In other words, if we have a larger effect size, we'll get a larger value for power.

The effect size in our current example on IQ's is

$$\frac{|\mu_{true} - \mu_{H_0}|}{\sigma} = \frac{|106 - 100|}{15} = 0.4.$$

This is a medium effect size.

What if we think that the true distribution has a mean of $\mu_{true} = 112$ IQ points? This will increase the effect size to:

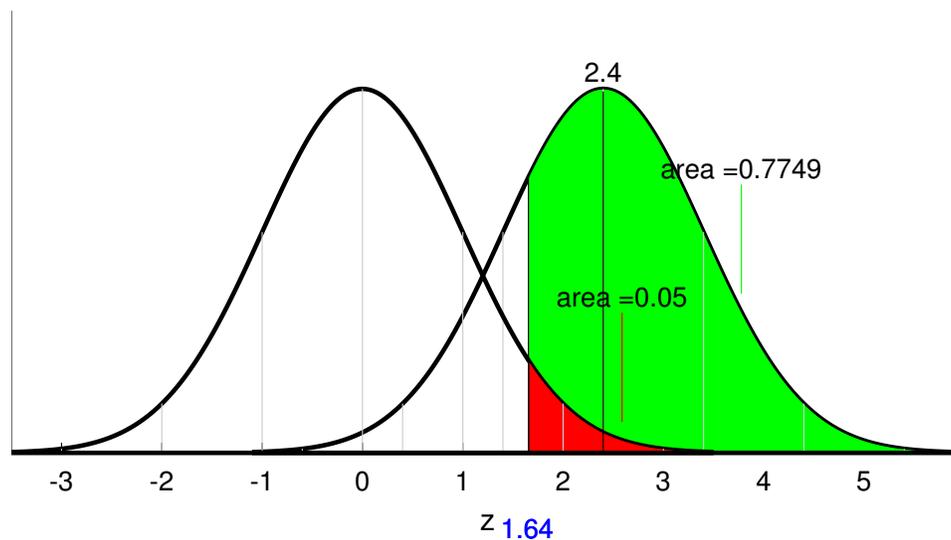
$$\frac{|\mu_{true} - \mu_{H_0}|}{\sigma} = \frac{|112 - 100|}{15} = 0.8.$$

which is a large effect size.

Means drawn from this new true distribution will have z-scores that are normally distributed with a mean of

$$z_{true} = \frac{(\mu_{true} - \mu_{H_0})}{\sigma_{\bar{x}}} = \frac{(112 - 100)}{5} = 2.4$$

Here's that standard diagram for this new true distribution:



Notice what has and has not changed. The true distribution is now shifted rightward to have a mean of $z_{true} = 2.4$. However, the critical region is unchanged - we still reject H_0 if our observed value of z is greater than 1.64.

Remember, power is the green area - the area under the true distribution above the critical value of z . You can now see how shifting the true distribution rightward increases power. For our numbers, increasing the effect size from 0.4 to 0.8 increased the power from 0.32821 to 0.7749.

Increasing sample size increases power

Another way to shift the true distribution away from the null distribution is to increase the sample size. For example, let's go back and assume that our true mean IQ is 106 points, but we increase our sample size from 9 to 36. The z-score for the mean of the true distribution is now:

$$z_{true} = \frac{(\mu_{true} - \mu_{H_0})}{\sigma_{\bar{x}}} = \frac{(\mu_{true} - \mu_{H_0})}{\frac{\sigma_x}{\sqrt{n}}} = \frac{(106 - 100)}{\frac{15}{\sqrt{36}}} = \frac{(106 - 100)}{2.5} = 2.4$$

This shift in z_{true} from 1.2 to 2.4 is the same as for the last example where μ_{true} was 112 IQ points and a sample size of 9. Thus, the power will also be 0.7764.

Note that increasing sample size increased the power without affecting the effect size (which doesn't depend on sample size:).

$$\text{effect size: } \frac{|\mu_{true} - \mu_{H_0}|}{\sigma} = \frac{|106 - 100|}{15} = 0.4.$$

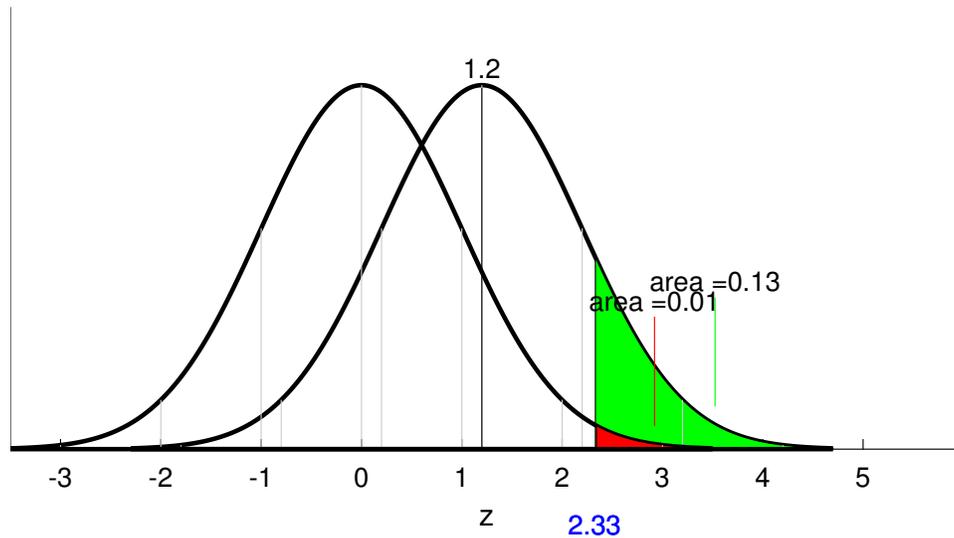
Increasing α increases power

A third thing that affects power is your choice of α . Recall that back in our original example of IQ scores with $\mu_{true} = 106$ points, the power for $\alpha = 0.05$ was 0.32821.

What if we keep the sample size at 9 so that effect size the same, but decrease our value of α to our second-favorite value, 0.01?

Decreasing α from 0.05 to 0.01 increases the critical value of z from 1.64 to 2.33.

Here's the new picture:



Notice that the true distribution still has a mean of $z_{true} = 1.2$. But shifting the critical value of z cuts into the green area, decreasing the power from 0.32821 to 0.13.

This illustrates a classic trade-off of Type I and Type II errors in decision making. Decreasing α by definition decreases the probability of making a Type I error. That is, decreasing α makes it harder to accidentally reject H_0 . But that comes with the cost decreasing α also makes it harder to correctly reject H_0 if it was true. That's the same as a decrease in power. and a corresponding increase the probability of a Type II error (β), since power = $1 - \beta$.

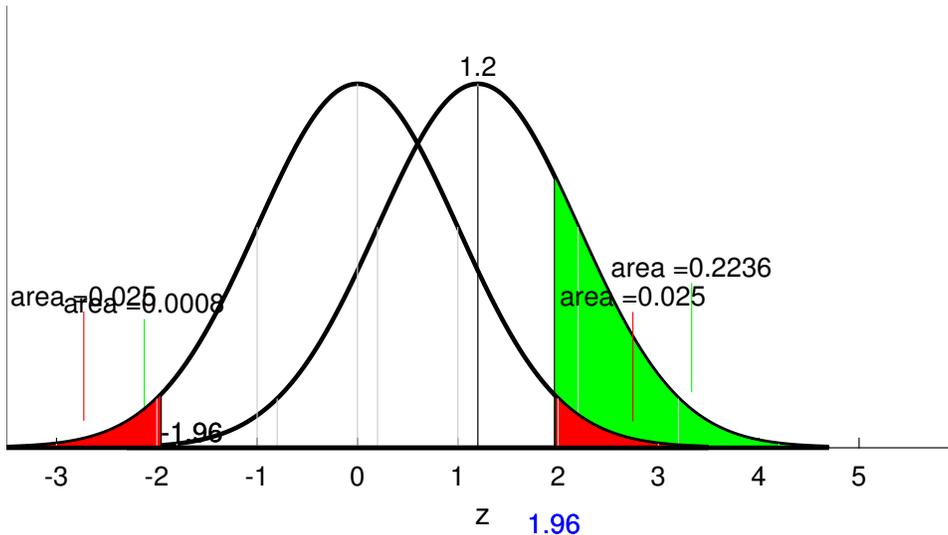
Get it? If you're still following things, when α goes down, z_{crit} goes up, power goes down and $\Pr(\text{Type II error})$ goes up. ***whew***

Power goes down with two-tailed tests

Let's look at our power calculation as we shift our original example from a one-tailed to a two-tailed test. Power calculations with two-tailed tests are a little more complicated because we have two rejection regions, but the concept is the same. We'll keep the true mean at 106 IQ points, the sample size $n = 9$ and $\alpha = 0.05$.

Recall that the critical value of z for a two-tailed test increases because we need to split the area in rejection region in half. You can therefore think of shifting to a two-tailed test as sort of like decreasing α , which as we know decreases power.

Here's the new picture:



We have two rejection regions now, one for $z < -1.64$ and one for $z > 1.64$. Power is, as always, the area under the true distribution that falls in the rejection region. For a two-tailed test we need to add up two areas. For this example, the area in the negative region is so far below the true distribution that the area below $z = -1.96$ is effectively zero. This leaves the green area above $z = 1.96$. Our power is therefore 0.22. This is lower than the power for the one-tailed test calculated above (0.32821).

Here's a summary of how things affect power:

Things that affect power			
thing	Pr(Type I error)	effect size	power
increasing effect size	same	increases	increases
increasing sample Size	same	same	increases
increasing alpha	increases	same	increases
two-tailed test	same	same	decreases

Go through the table and be sure to understand not only what happens with each thing, but also **why**.

Calculating power for a t-test

This last example with IQ's using the z-distribution is a bit artificial because we know the the standard deviation of the population. T-tests, where we use the sample standard deviation as an estimate for the population are much more common.

T-distributions look a lot like z-distributions, so calculating power for a t-test, by drawing bell-shaped curves and filling in areas, is very simliar at for what we just did for a z-test. But instead of working though a t-test example by hand we'll work with an example from the survey to show three other ways of calculating power: (1) with 'power curves', (2) using the Excel spreadsheet power calculator, and (3) in R.

Example from the survey

Suppose you want to test the hypothesis that the heights of women at the University of Washington differs from the average height of women in the US, which is 64 inches. We'll use the women in our class as a sample from the UW population. According to the survey that you filled out, the average height of the 122 women in our class is 64.7 inches with a standard deviation of 2.8246. Using an α value of 0.01, is this height significantly different than the US populaton?

To answer this we need to conduct a t-test for a single mean, since we don't know the standard deviation of the population. We first calculate the standard error of the mean:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{2.8246}{\sqrt{122}} = 0.2557$$

and use it to calculate our observed value of t:

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{64.7 - 64}{0.2557} = 2.76$$

This will be a two tailed test. Using table D we find that the critical value of t is:

$$t_{crit} = \pm 2.62 (using df = 120)$$

We reject H_0 . The height of UW women (M = 64.7, SD = 2.82) is significantly different than 64 , $t(121) = 2.76$, $p = 0.0067$.

To calculate power we need to calculate the effect size:

$$d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|64.7 - 64|}{2.8246} = 0.2496$$

This is a small effect size.

Let's figure out the 'observed power' of this test. That is, if we assume that our observed mean is the actual true population mean, let's calculate the probability that we'd correctly detect this difference between the observed mean (64.7 inches) and the null hypothesis mean (64 inches).

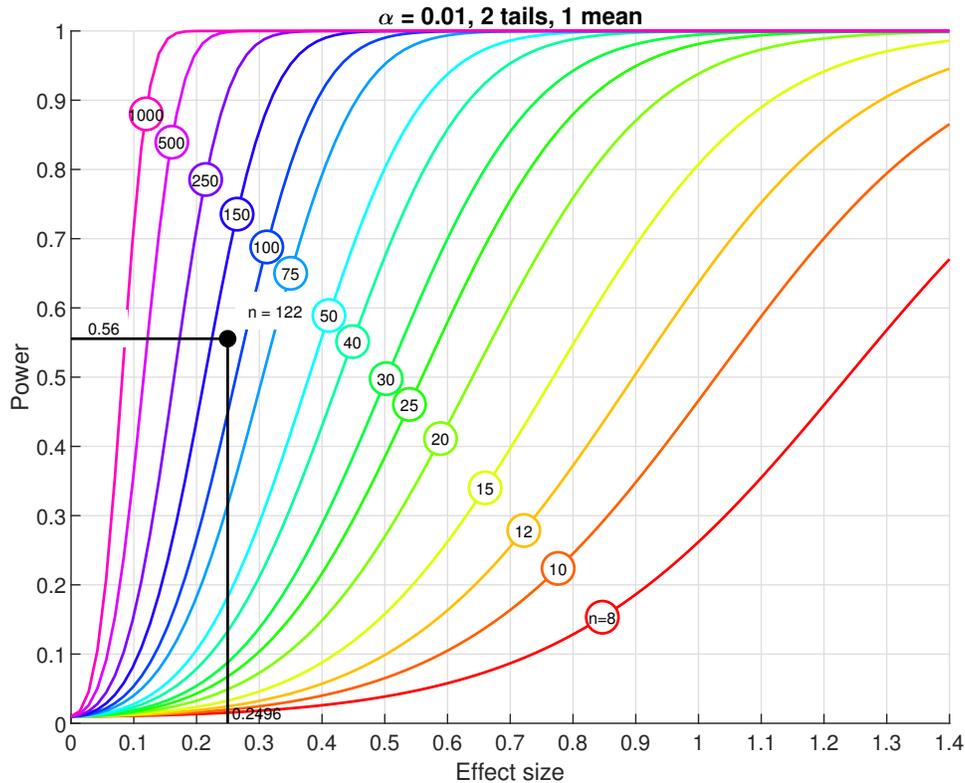
To calculate power for this two-sided t-test we'll use our measurements of effect size ($d = 0.2496$), sample size ($n = 122$), alpha ($\alpha = 0.01$). We'll start with 'Power Curves'

Power Curves

If your computer explodes on the way to work (or if you're not allowed to use a computer for some reason) I've provided 'power curves' that allow you to estimate power based on effect size, the sample size, α and the number of tails.

You can download the pdf file containing the power curves at:
<http://courses.washington.edu/psy315/pdf/PowerCurves.pdf>

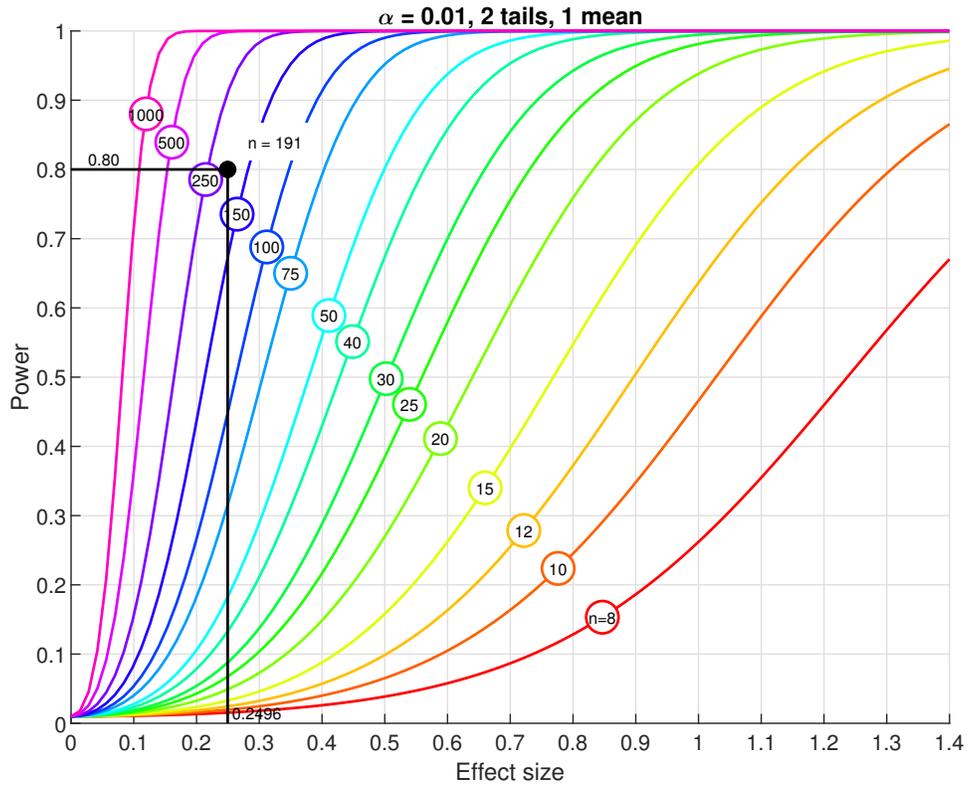
For our two-tailed test t-test with $\alpha = 0.01$, the curves look like this:



The x-axis is effect size, and each of the lines corresponds to a different sample size. I've drawn lines to indicate how to use these curves to estimate the observed power in this example. The line starts on the x-axis at the effect size of $d = 0.2496$, travels up to intersect with the line about where the sample size of $n = 122$ should be, and then moves left to the y-axis to estimate the observed power.

Most of the things that affect power can be seen in the power curves. What happens when effect size increases? What happens when the sample size increases? If you open up the power curve document you can also see what happens when alpha and the number of tails change by flipping through the pages in the PowerCurves.pdf document.

Finally, if we were to determine the number of subjects needed to obtain a power of 0.8, we can find use the power curve again and estimate the sample size:



So it looks like for an effect size of 0.2496 and a two-tailed test with $\alpha = 0.01$, we'd need to have about 190 students to get a power of 0.8.

Stats calculator

A second way to compute power for a t-test is to use the power calculator in the Excel spreadsheet.

The final tab 'Power' lets you plug in effect size, sample size and alpha to obtain the power for a one-tailed or two-tailed test for either one mean or two means.

If you plug in 0.2496 for the effect size, our sample size of $n = 122$ and $\alpha = 0.01$ into the power calculator, you'll get a power of 0.6548 for a two-tailed test with one mean:

effect size (d)	n	α
0.2496	122	0.01

One tailed test one mean			
t_{crit}	$t_{crit} - t_{obs}$	area	power
2.3576	-0.3994	0.6548	0.6548

Two tailed test one mean			
t_{crit}	$t_{crit} - t_{obs}$	area	power
-2.6171	-5.374	0	0.5555
2.6171	-0.1398	0.5555	

Using R to calculate power for a t-test

A third way for calculating power for a t-test is using R's 'power.t.test' function. The following script calculates the observed power and the desired sample size for a power of 0.8 from the example on heights above.

The R commands shown below can be found here: [Power.R](#)

```
# Power for a t-test is calculated with the conveniently named
# function 'power.t.test'
#
# We'll go through the example in the power tutorial comparing the heights of the
# women in our class to 64 inches.

# Load in the survey data:
rm(list = ls())
survey <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W21survey.csv")
# Get the heights of the women:
x <- survey$height[survey$gender=="Female"]
# Get rid of any NA's
x = x[!is.na(x)]
# Calculate the effect size (d), and sample size (n), comparing the mean to 64 inches.
d <- (mean(x)-64)/sd(x)
n <- length(x)
# set alpha
alpha <- .01
# We're ready to calculate power. Power depends on:
# 1) sample size (n)
# 2) effect size(d)
# 3) alpha

# Plus
# 4) one or two sided
# 5) one or two means (we've only worked with one mean so far)

# If you know #s 1 2 and 3, then you can find power. Or, if you know #2 and #3
# and power, then you can find #1.

# power.t.test works by letting you set 3 of these four things to numbers and
# the fourth to 'NULL'. Whatever you set to 'NULL' will be the thing that is
# calculated.
```

```
# Example 1: use power.t.test to determine the observed power based on the
# sample size, effect size, and alpha. We'll set power= NULL so it'll calculate
# the power:
out1 <- power.t.test(n = n,
                    d = d,
                    power = NULL,
                    sig.level = alpha,
                    alternative = "two.sided",
                    type = "one.sample" )
# This is our observed power:
out1$power
[1] 0.556817

# Example 2: use power.t.test to find the sample size given our effect size,
# alpha and a desired power of 0.8. We'll set n = NULL so it'll calculate n:
out2 <- power.t.test(n = NULL,
                    d = d,
                    power = .8,
                    sig.level = 0.01,
                    alternative = "two.sided",
                    type = "one.sample" )
# This is the desired sample size:
out2$n
[1] 190.8423
```

Questions

Here are 10 practice questions followed by their answers.

1) Your advisor asks you to sample the volume of 40 ubiquitous musical groups from a population and obtain a mean volume of 18.38 and a standard deviation of 1.5877 . Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected volume of 19?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

2) We decide to sample the IQ of 91 psych 315 students from a population and obtain a mean IQ of 85.38 and a standard deviation of 5.6851 .

Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected IQ of 86?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

3) You sample the damage of 54 comfortable baby names from a population and obtain a mean damage of 36.32 and a standard deviation of 5.4503 .

Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected damage of 38?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

4) Let's pretend that you sample the determination of 37 politicians from a population and obtain a mean determination of 53.46 and a standard deviation of 3.8657 .

Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected determination of 55?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

5) You are walking down the street and sample the laughter of 56 first ice dancers from a population and obtain a mean laughter of 6 and a standard deviation of 8.5164 .

Using an alpha value of $\alpha = 0.01$, is this observed mean significantly greater than an expected laughter of 5?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

6) Let's pretend that you sample the morality of 17 elbows from a population and obtain a mean morality of 45.94 and a standard deviation of 6.1763 .
Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected morality of 46?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

7) For some reason you sample the anger of 39 defective oceans from a population and obtain a mean anger of 16.63 and a standard deviation of 5.3131 .
Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected anger of 19?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

8) You are walking down the street and sample the grief of 102 UW undergraduates from a population and obtain a mean grief of 44.09 and a standard deviation of 5.3098 .
Using an alpha value of $\alpha = 0.01$, is this observed mean significantly different than an expected grief of 45?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

9) I sample the homework of 51 exams from a population and obtain a mean homework of 5.44 and a standard deviation of 1.6811 .
Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected homework of 6?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

10) Let's sample the baggage of 100 dollars from a population and obtain a mean baggage of 87.24 and a standard deviation of 5.7193 .
Using an alpha value of $\alpha = 0.01$, is this observed mean significantly greater than an expected baggage of 87?

What is the effect size?

Is the effect size small, medium or large?

What is the observed power?

What sample size would you need to obtain a power of 0.8?

Answers

1) volume of musical groups

$$\bar{x} = 18.38, s_x = 1.5877, n = 40$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{1.5877}{\sqrt{40}} = 0.251$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{18.38 - 19}{0.251} = -2.46$$

$$df = (n-1) = (40-1) = 39$$

$$t_{crit} = -2.43$$

We reject H_0 .

The volume of ubiquitous musical groups ($M = 18.38$, $SD = 1.59$) is significantly less than 19, $t(39) = -2.46$, $p = 0.0092$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|18.38 - 19|}{1.5877} = 0.3894$$

This is a medium effect size.

The observed power for one tailed test with an effect size of $d = 0.3894$, $n = 40$ and $\alpha = 0.01$ is 0.5146.

You'd need a sample of 69 musical groups to obtain a power of 0.8.

```
m <- 18.38
s <- 1.5877
n <- 40
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-19)/sem
t
[1] -2.469751
# effect size
d <- abs((m-19)/s)
d
[1] 0.390502
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.5229871
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")  
out$n  
[1] 68.54562
```

2) IQ of psych 315 students

$$\bar{x} = 85.38, s_x = 5.6851, n = 91$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{5.6851}{\sqrt{91}} = 0.596$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{85.38 - 86}{0.596} = -1.05$$

$$df = (n-1) = (91-1) = 90$$

$$t_{crit} = -1.66$$

We fail to reject H_0 .

The IQ of psych 315 students ($M = 85.38$, $SD = 5.69$) is not significantly less than 86, $t(90) = -1.05$, $p = 0.1492$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|85.38 - 86|}{5.6851} = 0.1097$$

This is a small effect size.

The observed power for one tailed test with an effect size of $d = 0.1097$, $n = 91$ and $\alpha = 0.05$ is 0.2699.

You'd need a sample of 515 psych 315 students to obtain a power of 0.8.

```
m <- 85.38
s <- 5.6851
n <- 91
alpha <- 0.05
sem <- s/sqrt(n)
t <- (m-86)/sem
t
[1] -1.040338
# effect size
d <- abs((m-86)/s)
d
[1] 0.109057
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.2701592
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")
out$n
[1] 521.185
```

3) damage of baby names

$$\bar{x} = 36.32, s_x = 5.4503, n = 54$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{5.4503}{\sqrt{54}} = 0.7417$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{36.32 - 38}{0.7417} = -2.26$$

$$df = (n-1) = (54-1) = 53$$

$$t_{crit} = \pm 2.67$$

We fail to reject H_0 .

The damage of comfortable baby names ($M = 36.32$, $SD = 5.45$) is not significantly different than 38, $t(53) = -2.26$, $p = 0.0279$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|36.32 - 38|}{5.4503} = 0.3076$$

This is a small effect size.

The observed power for two tailed test with an effect size of $d = 0.3076$, $n = 54$ and $\alpha = 0.01$ is 0.3412.

You'd need a sample of 127 baby names to obtain a power of 0.8.

```
m <- 36.32
s <- 5.4503
n <- 54
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-38)/sem
t
[1] -2.265092
# effect size
d <- abs((m-38)/s)
d
[1] 0.3082399
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "two.sided",
  type = "one.sample")
out$power
[1] 0.3514689
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "two.sided",
```

```
      type = "one.sample")
out$n
[1] 126.2516
```

4) determination of politicians

$$\bar{x} = 53.46, s_x = 3.8657, n = 37$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{3.8657}{\sqrt{37}} = 0.6355$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{53.46 - 55}{0.6355} = -2.42$$

$$df = (n-1) = (37-1) = 36$$

$$t_{crit} = -2.43$$

We fail to reject H_0 .

The determination of politicians ($M = 53.46$, $SD = 3.87$) is not significantly less than 55, $t(36) = -2.42$, $p = 0.0104$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|53.46 - 55|}{3.8657} = 0.3975$$

This is a medium effect size.

The observed power for one tailed test with an effect size of $d = 0.3975$, $n = 37$ and $\alpha = 0.01$ is 0.4934.

You'd need a sample of 66 politicians to obtain a power of 0.8.

```
m <- 53.46
s <- 3.8657
n <- 37
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-55)/sem
t
[1] -2.423223
# effect size
d <- abs((m-55)/s)
d
[1] 0.3983755
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.502304
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")
out$n
[1] 65.97088
```

5) laughter of ice dancers

$$\bar{x} = 6, s_x = 8.5164, n = 56$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{8.5164}{\sqrt{56}} = 1.138$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{6-5}{1.138} = 0.88$$

$$df = (n-1) = (56-1) = 55$$

$$t_{crit} = 2.40$$

We fail to reject H_0 .

The laughter of first ice dancers ($M = 6$, $SD = 8.52$) is not significantly greater than 5 , $t(55) = 0.88$, $p = 0.192$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|6-5|}{8.5164} = 0.1173$$

This is a small effect size.

The observed power for one tailed test with an effect size of $d = 0.1173$, $n = 56$ and $\alpha = 0.01$ is 0.0673.

You'd need a sample of 732 ice dancers to obtain a power of 0.8.

```
m <- 6
s <- 8.5164
n <- 56
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-5)/sem
t
[1] 0.8786946
# effect size
d <- abs((m-5)/s)
d
[1] 0.1174205
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.07090933
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")
out$n
[1] 730.6122
```

6) morality of elbows

$$\bar{x} = 45.94, s_x = 6.1763, n = 17$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{6.1763}{\sqrt{17}} = 1.498$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{45.94 - 46}{1.498} = -0.04$$

$$df = (n-1) = (17-1) = 16$$

$$t_{crit} = -1.75$$

We fail to reject H_0 .

The morality of elbows ($M = 45.94$, $SD = 6.18$) is not significantly less than 46, $t(16) = -0.04$, $p = 0.4836$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|45.94 - 46|}{6.1763} = 0.0102$$

This is a small effect size.

The observed power for one tailed test with an effect size of $d = 0.0102$, $n = 17$ and $\alpha = 0.05$ is 0.0539.

You'd need a sample of 10000 elbows to obtain a power of 0.8.

```
m <- 45.94
s <- 6.1763
n <- 17
alpha <- 0.05
sem <- s/sqrt(n)
t <- (m-46)/sem
t
[1] -0.04005413
# effect size
d <- abs((m-46)/s)
d
[1] 0.009714554
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.05408665
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")  
out$n  
[1] 65513.59
```

7) anger of oceans

$$\bar{x} = 16.63, s_x = 5.3131, n = 39$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{5.3131}{\sqrt{39}} = 0.8508$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{16.63 - 19}{0.8508} = -2.78$$

$$df = (n-1) = (39-1) = 38$$

$$t_{crit} = -2.43$$

We reject H_0 .

The anger of defective oceans (M = 16.63, SD = 5.31) is significantly less than 19, $t(38) = -2.78$, $p = 0.0042$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|16.63 - 19|}{5.3131} = 0.4451$$

This is a medium effect size.

The observed power for one tailed test with an effect size of $d = 0.4451$, $n = 39$ and $\alpha = 0.01$ is 0.6363.

You'd need a sample of 53 oceans to obtain a power of 0.8.

```
m <- 16.63
s <- 5.3131
n <- 39
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-19)/sem
t
[1] -2.785689
# effect size
d <- abs((m-19)/s)
d
[1] 0.4460673
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.6404909
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")
out$n
[1] 53.17783
```

8) grief of UW undergraduates

$$\bar{x} = 44.09, s_x = 5.3098, n = 102$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{5.3098}{\sqrt{102}} = 0.5257$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{44.09 - 45}{0.5257} = -1.74$$

$$df = (n-1) = (102-1) = 101$$

$$t_{crit} = \pm 2.63$$

We fail to reject H_0 .

The grief of UW undergraduates (M = 44.09, SD = 5.31) is not significantly different than 45, $t(101) = -1.74$, $p = 0.0853$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|44.09 - 45|}{5.3098} = 0.1721$$

This is a small effect size.

The observed power for two tailed test with an effect size of $d = 0.1721$, $n = 102$ and $\alpha = 0.01$ is 0.1885.

You'd need a sample of 397 UW undergraduates to obtain a power of 0.8.

```
m <- 44.09
s <- 5.3098
n <- 102
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-45)/sem
t
[1] -1.730865
# effect size
d <- abs((m-45)/s)
d
[1] 0.1713812
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "two.sided",
  type = "one.sample")
out$power
[1] 0.1912564
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "two.sided",
```

```
      type = "one.sample")
out$n
[1] 400.9506
```

9) homework of exams

$$\bar{x} = 5.44, s_x = 1.6811, n = 51$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{1.6811}{\sqrt{51}} = 0.2354$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{5.44 - 6}{0.2354} = -2.4$$

$$df = (n-1) = (51-1) = 50$$

$$t_{crit} = -2.40$$

We fail to reject H_0 .

The homework of exams ($M = 5.44$, $SD = 1.68$) is not significantly less than 6, $t(50) = -2.4$, $p = 0.0101$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|5.44 - 6|}{1.6811} = 0.3357$$

This is a small effect size.

The observed power for one tailed test with an effect size of $d = 0.3357$, $n = 51$ and $\alpha = 0.01$ is 0.4977.

You'd need a sample of 92 exams to obtain a power of 0.8.

```
m <- 5.44
s <- 1.6811
n <- 51
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-6)/sem
t
[1] -2.378919
# effect size
d <- abs((m-6)/s)
d
[1] 0.3331152
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.4952873
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")
out$n
[1] 93.16785
```

10) baggage of dollars

$$\bar{x} = 87.24, s_x = 5.7193, n = 100$$

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \frac{5.7193}{\sqrt{100}} = 0.5719$$

$$t = \frac{\bar{x} - \mu_{hyp}}{s_{\bar{x}}} = \frac{87.24 - 87}{0.5719} = 0.42$$

$$df = (n-1) = (100-1) = 99$$

$$t_{crit} = 2.37$$

We fail to reject H_0 .

The baggage of dollars ($M = 87.24$, $SD = 5.72$) is not significantly greater than 87, $t(99) = 0.42$, $p = 0.3374$.

$$\text{Effect size: } d = \frac{|\bar{x} - \mu_{hyp}|}{s_x} = \frac{|87.24 - 87|}{5.7193} = 0.0421$$

This is a small effect size.

The observed power for one tailed test with an effect size of $d = 0.0421$, $n = 100$ and $\alpha = 0.01$ is 0.0274.

You'd need a sample of 5665 dollars to obtain a power of 0.8.

```
m <- 87.24
s <- 5.7193
n <- 100
alpha <- 0.01
sem <- s/sqrt(n)
t <- (m-87)/sem
t
[1] 0.4196318
# effect size
d <- abs((m-87)/s)
d
[1] 0.04196318
# observed power
out <- power.t.test(n = n, d = d, power = NULL, sig.level = alpha,
  alternative = "one.sided",
  type = "one.sample")
out$power
[1] 0.02791186
# desired n
out <- power.t.test(n = NULL, d = d, power = 0.8, sig.level = alpha,
  alternative = "one.sided",
```

```
      type = "one.sample")  
out$n  
[1] 5702.054
```