Exploration librairie swirl : Power

> swirl()

Welcome to swirl! Please sign in. If you've been here before, use the same name as you did then. If you are new, call yourself something unique.

What shall I call you? ***jlbellier***

Please choose a course, or type 0 to exit swirl.

1: Statistical Inference
2: Take me to the swirl course repository!

Selection: ***1***

Please choose a lesson, or type 0 to return to course menu.

1: Introduction	*2: Probability1*	*3: Probability2*
4: ConditionalProbability	*5: Expectations*	*6: Variance*
7: CommonDistros	*8: Asymptotics*	*9: T Confidence Intervals*
10: Hypothesis Testing	*11: P Values*	*12: Power*
13: Multiple Testing	*14: Resampling*	

Selection: ***12***

Attempting to load lesson dependencies...
Package 'reshape2' loaded correctly!
Package 'ggplot2' loaded correctly!

0%

Power. (Slides for this and other Data Science courses may be found at github https://github.com/DataScienceSpecialization/courses/. If you care to use them, they must be downloaded as a zip file and viewed locally. This lesson corresponds to 06_Statistical_Inference/11_Power.)

= 1%

In this lesson, as the name suggests, we'll discuss POWER, which is the probability of rejecting the null hypothesis when it is false, which is good and proper.

== 2%

Hence you want more POWER.

=== 3%

Power comes into play when you're designing an experiment, and in particular, if you're trying to determine if a null result (failing to reject a null hypothesis) is meaningful. For instance, you might have to determine if your sample size was big enough to yield a meaningful, rather than random, result.

==== 4%

Power gives you the opportunity to detect if your ALTERNATIVE hypothesis is true.

===== ***5%***

Do you recall the definition of a Type II error? Remember, errors are bad.

1: Rejecting a true null hypothesis
2: Misspelling the word hypothesis
3: Miscalculating a t score
4: Accepting a false null hypothesis

Selection: ***1***

Nice try, but that's not exactly what I was hoping for. Try again.
Remember the courtroom example? Letting a guilty person walk, accepting the null hypothesis of innocence, is a Type II error.

1: Miscalculating a t score
2: Rejecting a true null hypothesis
3: Accepting a false null hypothesis
4: Misspelling the word hypothesis

Selection: ***3***

That's the answer I was looking for.

====== ***7%***

Beta is the probability of a Type II error, accepting a false null hypothesis; the complement of this is obviously (1 - beta) which represents the probability of rejecting a false null ypothesis. This is good and this is POWER!

======= ***8%***

Recall our previous example involving the Respiratory Distress Index and sleep disturbances. Our null hypothesis H_0 was that mu = 30 and our alternative hypothesis H_a was that mu > 30.

======== ***9%***

Which of the following expressions represents our test statistic under this null hypothesis? Here X' represents the sample mean, s is the sample std deviation, and n is the sample size. Assume X' follows a t distribution.

1: (X'-30)/(s^2/n)
2: 30/(s/sqrt(n))
3: X'/(s^2/n)
4: (X'-30)/(s/sqrt(n))

Selection: ***4***

Nice work!

========= ***10%***

In the expression for the test statistic (X'-30)/(s/sqrt(n)) what does (s/sqrt(n)) represent?

1: a standard variance
2: a standard error
3: a standard sample
4: a standard measure
5: a standard bearer

Selection: ***2***

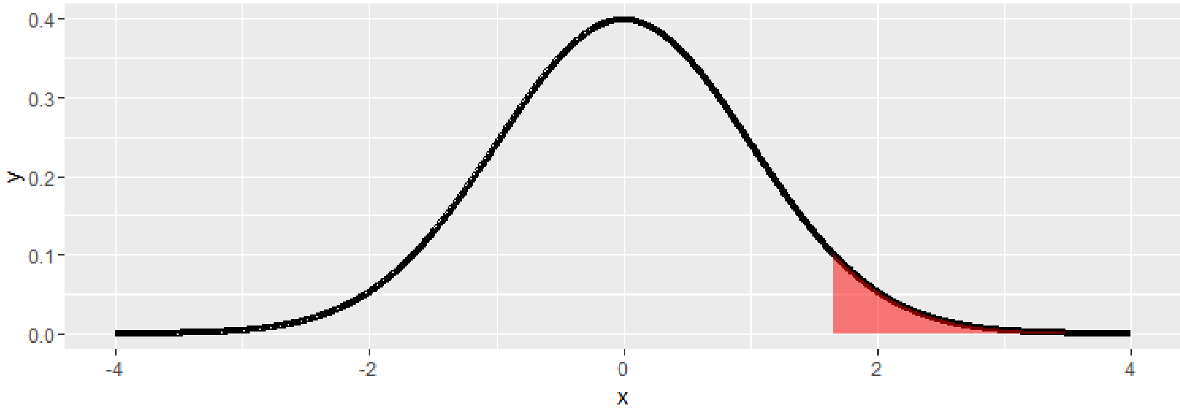
Your dedication is inspiring!

========== ***11%***

Suppose we're testing a null hypothesis H_0 with an alpha level of .05. Since H_a proposes that mu > 30 (the mean hypothesized by H_0), power is the probability that the true mean mu is greater than the (1-alpha) quantile or qnorm(.95). For simplicity, assume we're working with normal distributions of which we know the variances.

=========== ***12%***

Here's the picture we've used a lot in these lessons. As you know, the shaded portion represents 5% of the area under the curve. If a test statistic fell in this shaded portion we would reject H_0 because the sample mean is too far from the mean (center) of the distribution hypothesized by H_0. Instead we would favor H_a, that mu > 30. This happens with probability .05.



============ ***13%***

You might well ask, "What does this have to do with POWER?" Good question. We'll look at some pictures to show you.

============= ***14%***

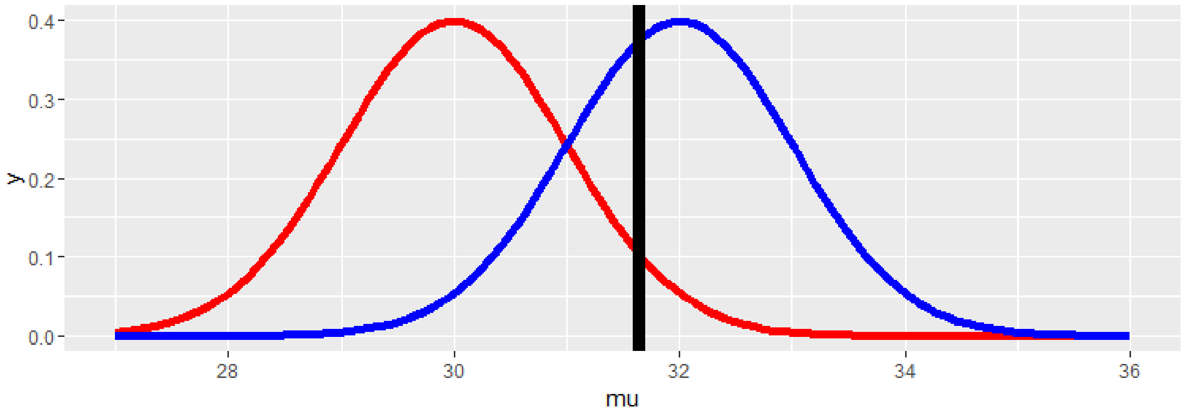
First we have to emphasize a key point. The two hypotheses, H_0 and H_a, actually represent two distributions since they're talking about means or centers of distributions. H_0 says that the mean is mu_0 (30 in our example) and H_a says that the mean is mu_a.

============== ***15%***

We're assuming normality and equal variance, say sigma^2/n, for both hypotheses, so under H_0, X'~ N(mu_0, sigma^2/n) and under H_a, X'~ N(mu_a, sigma^2/n).

=============== ***16%***

Here's a picture with the two distributions. We've drawn a vertical line at our favorite spot, at the 95th percentile of the red distribution. To the right of the line lies 5% of the red distribution.



================ **17%**

Quick quiz! Which distribution represents H_0?

1: the red
2: the blue

Selection: ***1***

You are quite good my friend!

================= **18%**

Which distribution represents H_a?

1: the red
2: the blue

Selection: ***2***

Excellent work!

================== **20%**

From the picture, what is the mean proposed by H_a?

1: 36
2: 28
3: 32
4: 30

Selection: ***3***

Keep up the great work!

=================== **21%**

See how much of the blue distribution lies to the right of that big vertical line?

==================== **22%**

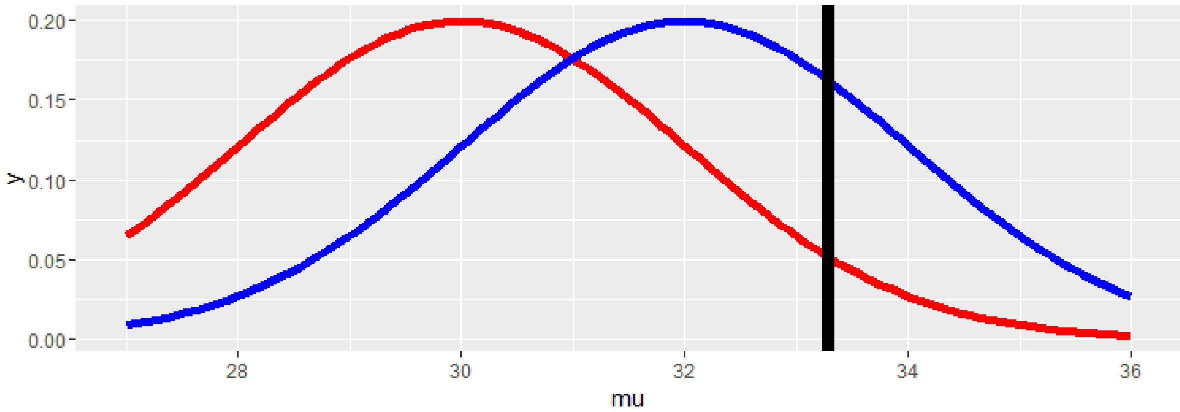
That, my friend, is POWER!

===================== **23%**

It's the area under the blue curve (H_a) to the right of the vertical line.

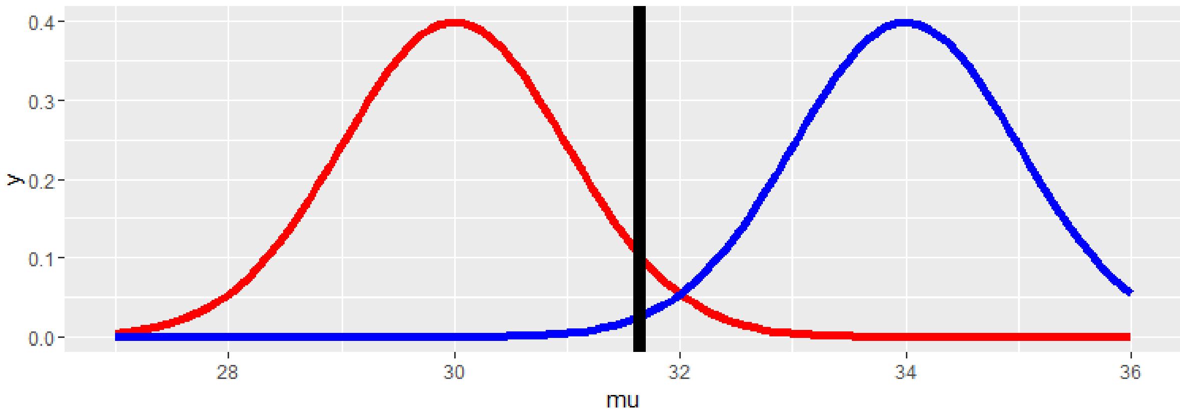
====================== ***24%***

Note that the placement of the vertical line depends on the null distribution. Here's another picture with fatter distributions. The vertical line is still at the 95th percentile of the null (red) distribution and 5% of the distribution still lies to its right. The line is calibrated to mu_0 and the variance.



======================= ***25%***

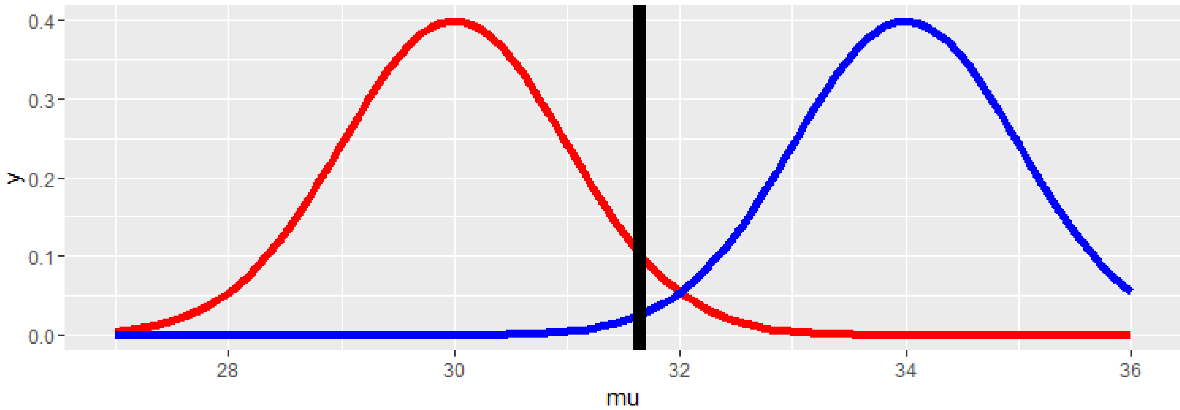
Back to our original picture.



======================== ***26%***

We've shamelessly stolen plotting code from the slides so you can see H_a in action. Let's look at pictures before we delve into numbers. We've fixed mu_0 at 30, sigma (standard deviation) at 4 and n (sample size) at 16. The function myplot just needs an alternative mean, mu_a, as argument. Run myplot now with an argument of 34 to see what it does.

> ***myplot(34)***



That's a job well done!

========================= **27%**

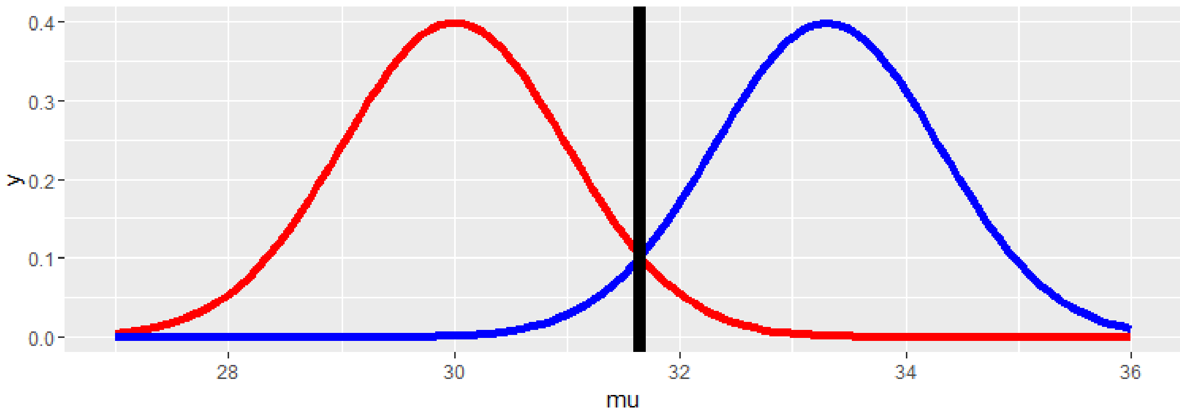
The distribution represented by H_a moved to the right, so almost all (100%) of the blue curve is to the right of the vertical line, indicating that with mu_a=34, the test is more powerful, i.e., there's a higher probability that it's correct to reject the null hypothesis since it appears false. Now try myplot with an argument of 33.3.

> ***myplot(33.3)***

That's a job well done!

========================== **28%**

This isn't as powerful as the test with mu_a=34 but it makes a pretty picture. Now try myplot with an argument of 30.

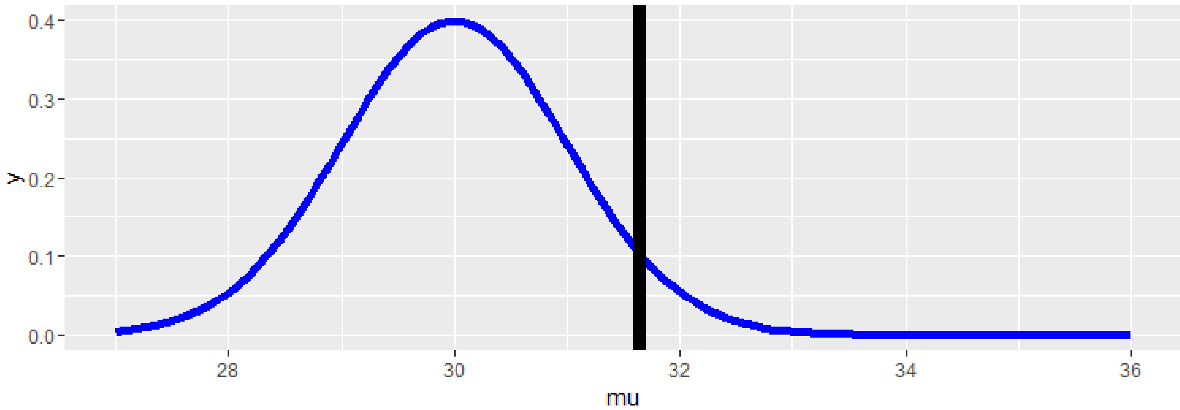


> ***myplot(30)***

All that practice is paying off!

=========================== **29%**

Uh Oh! Did the red curve disappear? No. it's just under the blue curve. The power now, the area under the blue curve to the right of the line, is exactly 5% or alpha!



=========================== **30%**

So what did we learn?

============================= **32%**

First, power is a function that depends on a specific value of an alternative mean, mu_a, which is any value greater than mu_0, the mean hypothesized by H_0. (Recall that H_a specified mu>30.)

============================== **33%**

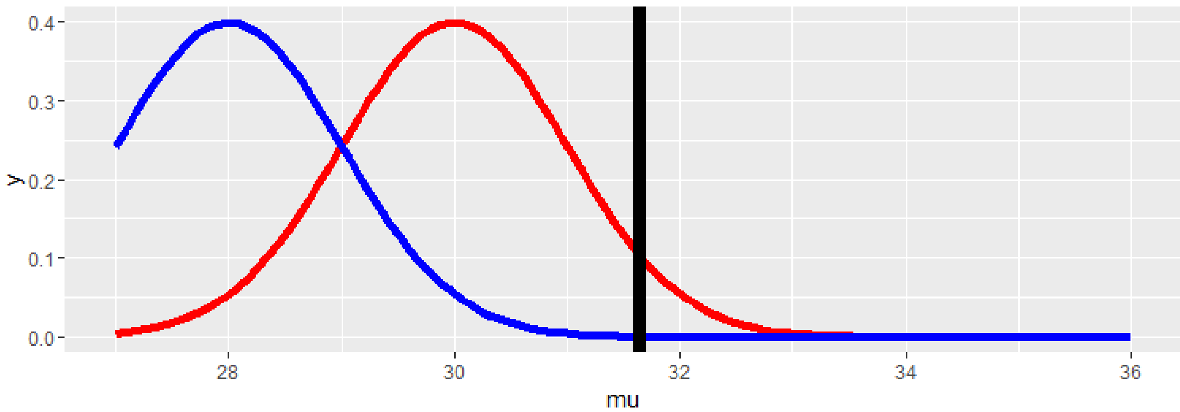
Second, if mu_a is much bigger than mu_0=30 then the power (probability) is bigger than if mu_a is close to 30. As mu_a approaches 30, the mean under H_0, the power approaches alpha.

=============================== **34%**

Just for fun try myplot with an argument of 28.

> **myplot(28)**

You nailed it! Good job!

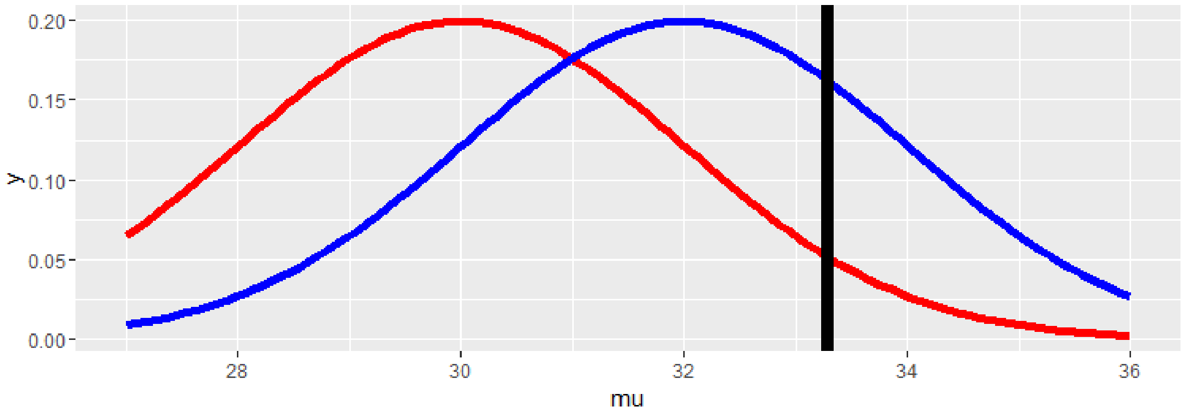


================================ **35%**

We see that the blue curve has moved to the left of the red, so the area under it, to the right of the line, is less than the 5% under the red curve. This then is even less powerful and contradicts H_a so it's not worth looking at.

================================= **36%**

Here's a picture of the power curves for different sample sizes. Again, this uses code "borrowed" from the slides. The alternative means, the mu_a's, are plotted along the horizontal axis and power along the vertical.



==================================== **37%**

What does the graph show us about mu_a?

1: as it gets bigger, it gets less powerful
2: power is independent of mu_a
3: as it gets bigger, it gets more powerful

Selection: ***3***

All that hard work is paying off!

===================================== **38%**

What does the graph show us about sample size?

1: as it gets bigger, it gets more powerful
2: power is independent of sample size
3: as it gets bigger, it gets less powerful

Selection: ***1***

Perseverance, that's the answer.

====================================== **39%**

*Now back to numbers. Our test for determining rejection of H_0 involved comparing a test statistic, namely Z=(X'-30)/(sigma/sqrt(n)), against some quantile, say Z_95, which depended on our level size alpha (.05 in this case). H_a proposed that mu > mu_0, so we tested if Z>Z_95. This is equivalent to X' > Z_95 * (sigma/sqrt(n)) + 30, right?*

======================================= **40%**

Recall that nifty R function pnorm, which gives us the probability that a value drawn from a normal distribution is greater or less than/equal to a specified quantile argument depending on the flag lower.tail. The function also takes a mean and standard deviation as arguments.

== **41%**

*Suppose we call pnorm with the quantile 30 + Z_95 * (sigma/sqrt(n)) and specify mu_a as our mean argument. This would return a probability which we can interpret as POWER. Why?*

=== **42%**

*Recall our picture of two distributions. 30 + Z_95 * (sigma/sqrt(n)) represents the point at which our vertical line falls. It's the point on the null distribution at the (1-alpha) level.*

==	**43%**

*Study this picture. Calling pnorm with 30 + Z_95 * (sigma/sqrt(n)) as the quantile and mu_a, say 32, as the mean and lower.tail=FALSE does what?*

1: returns the area under the red curve to the right of the line
2: returns the area under the blue curve to the right of the line
3: returns the area under the red curve to the left of the line
4: returns the area under the blue curve to the left of the line

*Selection: **2***

All that hard work is paying off!

===	**45%**

Let's try some examples now. Before we do, what do we know pnorm will return if we specify a quantile less than the mean?

1: an answer dependent on alpha
2: an answer less than .50
3: an answer dependent on beta
4: an answer greater than 1

*Selection: **2***

Nice work!

==	**46%**

First, define a variable z as qnorm(.95)

*> **z <- qnorm(0.95)***

Keep working like that and you'll get there!

===	**47%**

Run pnorm now with the quantile 30+z, mean=30, and lower.tail=FALSE. We've specified sigma and n so that the standard deviation of the sample mean is 1.

*> **pnorm(30+z, mean=30, lower.tail = FALSE)***
[1] 0.05

You are really on a roll!

==	**48%**

That's not surprising, is it? With the mean set to mu_0 the two distributions, null and alternative, are the same and power=alpha. Now run pnorm now with the quantile 30+z, mean=32, and lower.tail=FALSE.

*> **pnorm(30+z, mean=32, lower.tail = FALSE)***
[1] 0.63876

Your dedication is inspiring!

== **49%**

See how this is much more powerful? 64% as opposed to 5%. When the sample mean is quite different from (many standard errors greater than) the mean hypothesized by the null hypothesis, the probability of rejecting H_0 when it is false is much higher. That is power!

=== **50%**

Let's look again at the portly distributions.

== **51%**

With this standard deviation=2 (fatter distribution) will power be greater or less than with the standard deviation=1?

1: less than
2: greater
3: the same

Selection: ***1***

Perseverance, that's the answer.

=== **52%**

To see this, run pnorm now with the quantile 30+z, mean=32 and sd=1. Don't forget to set lower.tail=FALSE so you get the right tail.

> ***pnorm(30+z, mean=32, sd=1,lower.tail = FALSE)***
[1] 0.63876

That's correct!

== **53%**

*Now run pnorm now with the quantile 30+z*2, mean=32 and sd=2. Don't forget to set lower.tail=FALSE so you get the right tail.*

> ***pnorm(30+z, mean=32, sd=2,lower.tail = FALSE)***
[1] 0.5704709

Not quite, but you're learning! Try again. Or, type info() for more options.

*Type pnorm(30+z*2,mean=32,sd=2,lower.tail=FALSE) at the command prompt.*

>
> ***pnorm(30+z*2, mean=32, sd=2,lower.tail = FALSE)***
[1] 0.259511

You got it right!

=== **54%**

See the power drain from 64% to 26% ? Let's review some basic facts about power. We saw before in our pictures that the power of the test depends on mu_a. When H_a specifies that mu > mu_0, then as mu_a grows and exceeds mu_0 increasingly, what happens to power?

1: it decreases
2: it increases
3: it doesn't change

Selection: ***2***

All that hard work is paying off!

=== ***55%***

Here's another question. Recall our power curves from before.

== ***57%***

As the sample size increases, what happens to power?

1: it increases
2: it decreases
3: it doesn't change

Selection: ***1***

That's a job well done!

=== ***58%***

Here's another one. More power curves.

== ***59%***

As variance increases, what happens to power?

1: it doesn't change
2: it decreases
3: it increases

Selection: ***2***

All that practice is paying off!

=== ***60%***

Here's another one. And even more power curves.

== ***61%***

As alpha increases, what happens to power?

1: it increases
2: it decreases
3: it doesn't change

Selection: ***1***

That's the answer I was looking for.

=== ***62%***

If H_a proposed that mu != mu_0 we would calculate the one sided power using alpha / 2 in the direction of mu_a (either less than or greater than mu_0). (This is only approximately *right, it excludes the probability of getting a large test statistic in the opposite direction of the truth.*

=== 63%

Since power goes up as alpha gets larger would the power of a one-sided test be greater or less than the power of the associated two sided test?

1: less than
2: greater
3: they're the same

*Selection: **2***

Nice work!

=== 64%

Finally, if H_a specified that mu < mu_0 could we still do the same kind of power calculations?

1: No
2: Yes

*Selection: **1***

Not quite, but you're learning! Try again.
We just have to look at the right tail and flip all our reasoning.

1: No
2: Yes

*Selection: **2***

All that hard work is paying off!

=== 65%

*Suppose H_a says that mu > mu_0. Then power = 1 - beta = Prob (X' > mu_0 + z_(1-alpha) * sigma/sqrt(n)) assuming that X'~ N(mu_a,sigma^2/n). Which quantities do we know in this statement, given the context of the problem? Let's work through this.*

=== 66%

What does the null hypothesis H_0 tell us that the population mean equals?

1: mu_a
2: mu_0
3: alpha
4: beta

*Selection: **2***

Keep up the great work!

=== 67%

After the null mean mu_0 is proposed what does the designer of the hypothesis test specify in order to reject or fail-to-reject H_0? In other words, what is the level size of the test?

1: mu_0
2: alpha
3: beta
4: mu_a

Selection: ***2***

Keep working like that and you'll get there!

=== 68%

*So we know that the quantities mu_0 and alpha are specified by the test designer. In the statement 1 - beta = Prob(X' > mu_0 + z_(1-alpha) * sigma/sqrt(n)) given mu_a > mu_0, mu_0 and alpha are specified, and X' depends on the data. The other four quantities, (beta, sigma, n, and mu_a), are all unknown.*

=== 70%

It should be obvious that specifying any three of these unknowns will allow us to solve for the missing fourth. Usually, you only try to solve for power (1-beta) or the sample size n.

== 71%

An interesting point is that power doesn't need mu_a, sigma and n individually. Instead only sqrt(n)(mu_a - mu_0) /sigma is needed. The quantity (mu_a - mu_0) / sigma is called the EFFECT SIZE. This is the difference in the means in standard deviation units. It is unit free so it can be interpreted in different settings.*

=== 72%

We'll work through some examples of this now. However, instead of assuming that we're working with normal distributions let's work with t distributions. Remember, they're pretty close to normal with large enough sample sizes.

== 73%

Power is still a probability, namely P((X' - mu_0)/(S /sqrt(n)) > t_(1-alpha, n-1) given H_a that mu > mu_a). Notice we use the t quantile instead of the z. Also, since the proposed distribution is not centered at mu_0, we have to use the non-central t distribution.

=== 74%

R comes to the rescue again with the function power.t.test. We can omit one of the arguments and the function solves for it. Let's first use it to solve for power.

== 75%

We'll run it three times with the same values for n (16) and alpha (.05) but different delta and standard deviation values. We'll show that if delta (difference in means) divided by the standard deviation is the same, the power returned will also be the same. In other words, the effect size is constant for all three of our tests.

=== 76%

We'll specify a positive delta; this tells power.t.test that H_a proposes that mu > mu_0 and so we'll need a one-sided test. First run power.t.test(n = 16, delta = 2 / 4, sd=1, type = "one.sample", alt = "one.sided")$power .

```
> power.t.test(n = 16, delta = 2 / 4,sd=1, type = "one.sample", alt = "one.sided")$power
[1] 0.6040329
```

You are quite good my friend!

=== 77%

Now change delta to 2 and sd to 4. Keep everything else the same.

```
> power.t.test(n = 16, delta = 2,sd=4, type = "one.sample", alt = "one.sided")$power
[1] 0.6040329
```

Perseverance, that's the answer.

=== 78%

Same answer, right? Now change delta to 100 and sd to 200. Keep everything else the same.

```
> power.t.test(n = 16, delta = 100,sd=200, type = "one.sample", alt = "one.sided")$power
[1] 0.6040329
```

Excellent work!

=== 79%

So keeping the effect size (the ratio delta/sd) constant preserved the power. Let's try a similar experiment except now we'll specify a power we want and solve for the sample size n.

=== 80%

First run power.t.test(power = .8, delta = 2 / 4, sd=1, type = "one.sample", alt = "one.sided")$n .

```
> power.t.test(power=0.8, delta = 2/4,sd=1, type = "one.sample", alt = "one.sided")$n
[1] 26.13751
```

You got it!

=== 82%

Now change delta to 2 and sd to 4. Keep everything else the same.

```
> power.t.test(power=0.8, delta = 2,sd=4, type = "one.sample", alt = "one.sided")$n
[1] 26.13751
```

You're the best!

=== 83%

Same answer, right? Now change delta to 100 and sd to 200. Keep everything else the same.

```
> power.t.test(power=0.8, delta =100 ,sd=200, type = "one.sample", alt = "one.sided")$n
[1] 26.13751
```

You are amazing!

=== 84%

Now use power.t.test to find delta for a power=.8 and n=26 and sd=1

*> **power.t.test(power=0.8, n=26 ,sd=1, type = "one.sample", alt = "one.sided")$delta***
[1] 0.5013986

All that hard work is paying off!
=== 85%
Not a surprising result, is it? It told you before that with an effect size of .5 and power .8, you need a sample size a little more than 26. Now run it with n=27.
*> **power.t.test(power=0.8, n=27 ,sd=1, type = "one.sample", alt = "one.sided")$delta***
[1] 0.4914855

You are really on a roll!
== 86%
What do you think will happen if you doubled sd to 2 and ran the same test?

1: delta won't change
2: delta will halve
3: delta will double

*Selection: **2***

Not quite, but you're learning! Try again.
Since you're doubling the denominator (sd) you have to double the numerator (delta) in order to keep the effect size constant.

1: delta will halve
2: delta won't change
3: delta will double

*Selection: **3***

Perseverance, that's the answer.
=== 87%
Now for a quick review. We call this the power.u.test since it comes after the power.t.test. LOL.
1. The level of a test is specified by what?

1: delta
2: alpha
3: gamma
4: None of the others
5: beta

*Selection: **2***

That's the answer I was looking for.
== 89%
2. What is a Type II error?

1: accepting a true hypothesis
2: rejecting a false hypothesis
3: rejecting a true hypothesis
4: accepting a false hypothesis

Selection: **4**

That's correct!

=== 90%

3. What is power?

1: beta
2: delta
3: None of the others
4: alpha
5: gamma
6: thrilling

Selection: **3**

That's a job well done!

===91%

4. You're a perfectionist designing an experiment and you want both alpha and beta to be small. Can they both be 0 for this single test?

1: Yes
2: No

Selection: **2**

Excellent work!

=== 92%

5. Suppose H_0 proposes mu = mu_0 and H_a proposes that mu < mu_0. You'll test a series of mu_a with power != alpha. Which of the following is NOT true?

1: mu_a-mu_0=0
2: mu_0-mu_a > 0
3: huh?
4: mu_a-mu_0 < 0

Selection: **1**

Great job!

=== 93%

6. Suppose H_0 proposes mu = mu_0 and H_a proposes that mu < mu_0. Which of the following is true?

1: mu_0=mu_a maximizes the power
2: the smaller mu_a-mu_0 the more powerful the test
3: the smaller mu_0-mu_a the more powerful the test

*Selection: **1***

Keep trying!
Here mu_a < mu_0 and the smaller mu_a-mu_0 is, the easier it is to discriminate between mu_a and mu_0.

1: mu_0=mu_a maximizes the power
2: the smaller mu_a-mu_0 the more powerful the test
3: the smaller mu_0-mu_a the more powerful the test

*Selection: **2***

All that hard work is paying off!
=== 95%
7. Which expression represents the size effect?

1: (mu_a - mu_0) / sqrt(sigma)
2: (mu_a - mu_0) / sigma
3: (mu_a - mu_0) / sqrt(n)
4: (mu_a - mu_0) / n

*Selection: **2***

You are really on a roll!
===96%
8. True or False? More power is better than less power.

1: False
2: True

*Selection: **2***

You're the best!
=== 97%
9. True or False? A larger beta (call it beta_max) is more powerful than a smaller beta.

1: True
2: False
*Selection: **2***

All that practice is paying off!
=== 98%

10. True or False? The larger the sample size the less powerful the test.

1: True
2: False

Selection: ***2***

That's a job well done!

==   99%
Congrats! You finished this powerful lesson. We hope you feel emPOWERED.
=== 100%