

Class 12: Confidence Intervals

Alison Presmanes Hill

Reading

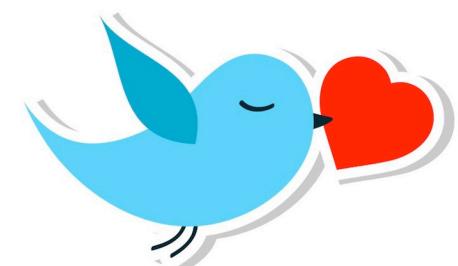
- Read pages 92-94 in All of Statistics:
- https://sakai.ohsu.edu/access/content/group/MATH-530-1-BB-F17/all_of_stats.pdf

Tools for classical inference

- Estimation (today)
 - Point estimates
 - Confidence intervals for those point estimates: need to know something about the (typically theoretical and unobserved) sampling distributions of those statistics in order to calculate standard errors

Remember, statistics...

- 2 main reasons we love them:
 - Sometimes they are **estimators** for population parameters we care about
 - Sometimes they are **test statistics**, i.e. the basis for a hypothesis test
- Our sample estimates are “best guesses” for population parameters
- Such a best guess is called a point estimate. We know, however, that point estimates – which are sample statistics, like the sample mean \bar{x} – vary from sample to sample.



Confidence intervals

- It is generally desirable to reflect the uncertainty due to sampling variation in an interval estimate, also called a confidence interval.
- Typically, a confidence interval takes the form of a point estimate \pm a margin of error.

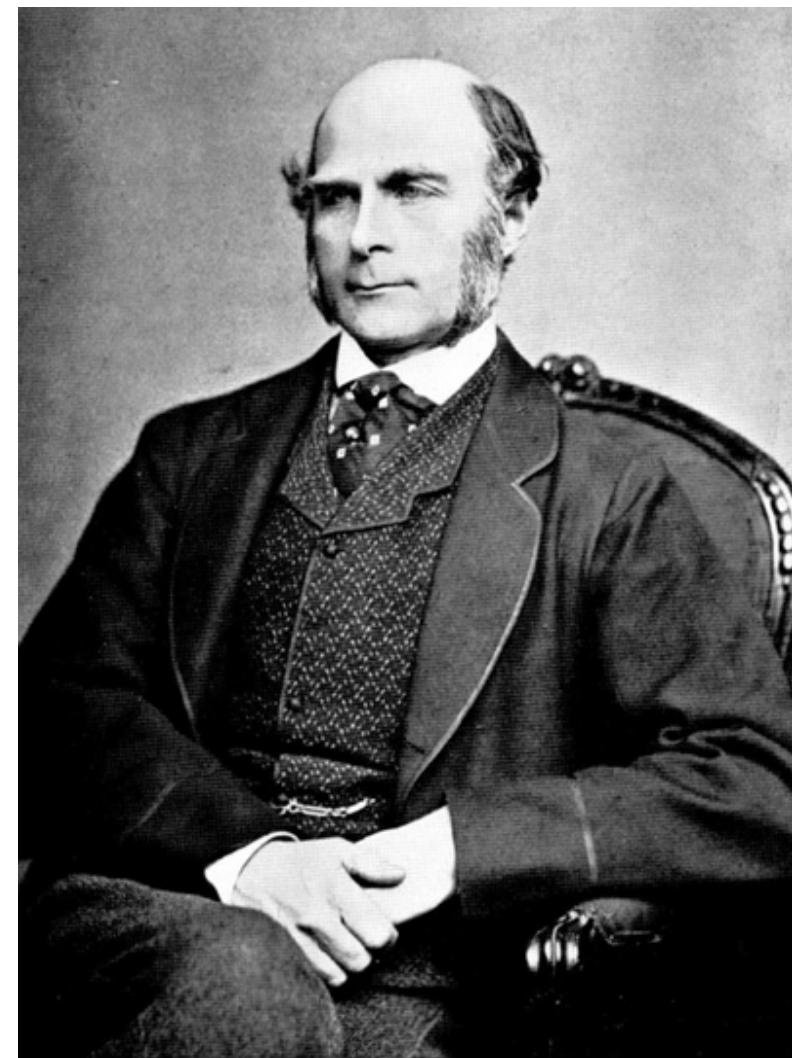
Confidence intervals for the population mean

Known*: μ, σ

*This is quite unusual to know both (or either)!
But bear with me here...

Sir Francis Galton (1822-1911)

- Created the concept of correlation and regression
- We'll use data from his 1885 study on the relationship between a parent's height and their children.
- **library (UsingR)
data (galton)**



Child height in Galton's sample

- Here are our sample statistics for heights among the 928 children:

$$\bar{x} = 68.09 \text{ inches}$$

$$s_x^2 = 6.34 \text{ (unit-less)}$$

$$s_x = 2.52 \text{ inches}$$



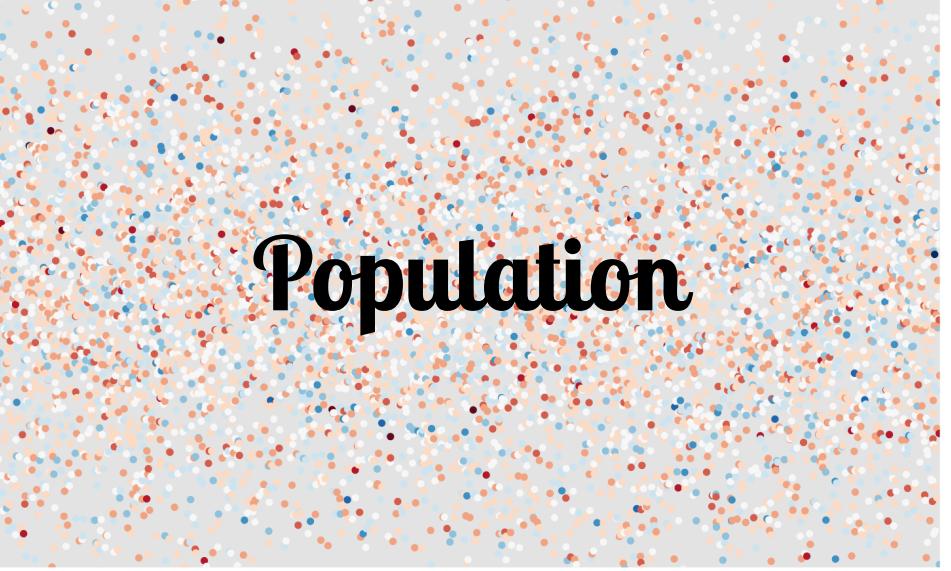
Imagine...

- We will begin by assuming, unrealistically, that we know the population mean $\mu = 70$ and standard deviation $\sigma = 3$, and, consequently that we know the sampling distribution of sample means, for samples of size n : that is,

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

This makes things a
whole lot easier!





Population

$$\leftarrow X_i \stackrel{iid}{\sim} N(\mu_X, \sigma_X^2)$$



**The sampling
distribution
of sample means**

$$\leftarrow \bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

Confidence intervals, based on normal distribution

- Based on the known population parameters, we know that the sampling distribution of the means is defined as:

$$\bar{x} \sim N\left(70, \frac{3^2}{928}\right)$$

- What we want to know is the confidence interval for the **population mean**:

(lower bound of CI, upper bound of CI)

$(\bar{x} - \text{margin of error}, \bar{x} + \text{margin of error})$

- How do we calculate the “margin of error” here?



Margin of error, based on normal distribution

- How do we calculate the “margin of error” here?

$$\bar{x} \sim N\left(70, \frac{3^2}{928}\right)$$

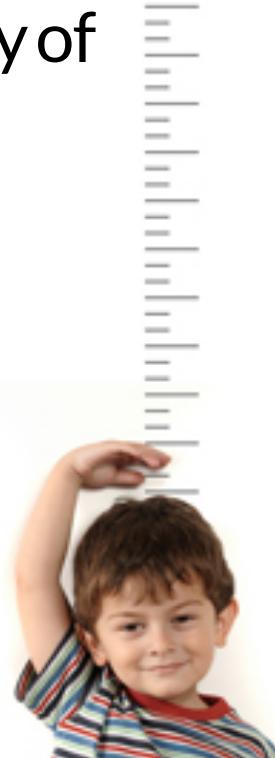
The square root of this is the SEM = .098

- We use the SEM here because it defines the variability of the sampling distribution for this specific statistic, the sample mean

$$(\bar{x} - \text{something} \times SEM, \bar{x} + \text{something} \times SEM)$$

What is this something?

What is this something?



Quantiles, based on normal distribution

- What is the something? It must be a **quantile** associated with how confident we wish to be!

$$\bar{x} \sim N \left(70, \frac{3^2}{928} \right)$$

The square root
of this is the
SEM = .098

- Most of the time, we want to be pretty confident about where we think the true population mean is, so we tend to aim for the 95% confidence interval
- Because we know the distribution (normal), μ , and σ , we can just use the unit normal distribution to get the **quantiles**



Quantiles, based on unit normal distribution

- We use `qnorm` in R to find out the **quantile** associated with the probability levels at each end of the **unit normal distribution** (because it is normal and hence symmetrical)

```
> qnorm(.025) # z score: mean = 0, sd = 1  
[1] -1.959964  
  
> qnorm(.975) # z score: mean = 0, sd = 1  
[1] 1.959964
```

- If we had a distribution $N(0, 1)$, only 2.5% of values are < -1.96 or > 1.96



Calculating 95% confidence intervals for μ

$$(\bar{x} - \underbrace{\text{something here!}}_{\times SEM}, \bar{x} + \underbrace{\text{something here!}}_{\times SEM})$$

$$(\bar{x} - q \frac{\sigma}{\sqrt{n}}, \bar{x} + q \frac{\sigma}{\sqrt{n}})$$

SEM = .098

Sample mean = 68.09

$q_z = 1.96$

Detailed description: The diagram illustrates the formula for a 95% confidence interval: $(\bar{x} - q \frac{\sigma}{\sqrt{n}}, \bar{x} + q \frac{\sigma}{\sqrt{n}})$. A blue dashed arrow originates from a box containing 'Sample mean = 68.09' and points to the central term \bar{x} . Another blue dashed arrow originates from a box containing 'SEM = .098' and points to the term $\frac{\sigma}{\sqrt{n}}$. A third blue dashed arrow originates from a box containing ' $q_z = 1.96$ ' and points to the multiplier q . Brackets are used to group the terms around the central mean \bar{x} .



Confidence interval for μ , σ known

```
> mu <- 70 # note that this isn't in any calculations!
> sigma <- 3
> n <- 928
> xbar <- 68.08847
> me <- qnorm(.975) * (3/sqrt(n))
> me
[1] 0.1930169
> lower <- xbar - me
> upper <- xbar + me
> c(lower, upper)
[1] 67.89545 68.28149
```

I am 95% confident that the population mean height for children (in this age range) is between 67.9 and 68.3 inches!



Confidence interval for μ , σ known

- Quicker way: use `qnorm` in R to find out the actual **quantile** associated with the probability levels at each end of our known sampling distribution of means

```
> confint <- qnorm(c( [redacted] , [redacted] , [redacted] ), [redacted])  
> confint  
[1] 67.89639 68.28055
```

I am 95% confident that the population mean height for children (in this age range) is between 67.9 and 68.3 inches!



Where is μ in all this?

- μ is smack dab in the center of our confidence interval- the CI is perfectly symmetrical around μ .

$$\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

$$P \left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right) = 0.95$$

We have been solving for μ all along...

$$P(-1.96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < 1.96) = 0.95$$

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

Interpreting confidence intervals (for μ , as an example)

- Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter
- We are $(1-\alpha) \times 100\%$ confident that the interval contains the true value of μ
- If we repeated the same process of drawing samples and computing intervals many times, in the long run, 95% of the intervals computed would include μ



Interpreting confidence intervals (for μ , as an example)

- μ is a constant- it is the *interval* that is random, with a 95% probability of including μ
- The true parameter value, μ , is either within or outside of CI; that is, $P(\text{CI lower bound} \leq \mu \leq \text{CI upper bound}) = 0$ or 1



Galton's (fictitious) intern attempts to summarize

We are 95% confident that the height for children in this sample is between 67.9 and 68.3 inches.

This is not correct:
A confidence interval is for a population parameter, and cannot be applied to individuals in the sample.



Galton's (fictitious) intern attempts to summarize

95% of all samples will give an average height between 67.9 and 68.3 inches.

This is not correct:
Each sample will give rise to a *different* confidence interval- 95% of those intervals will contain the true mean.



Galton's (fictitious) intern attempts to summarize

There is a **95% chance** that the true mean is between 67.9 and 68.3 inches.

This is not correct:
 μ is not random. The probability that μ is between 67.9 and 68.3 is either 0 or 1.



Galton's (fictitious) intern attempts to summarize

We are 95% confident that the population mean height of children (in this age range) is between 67.9 and 68.3 inches.

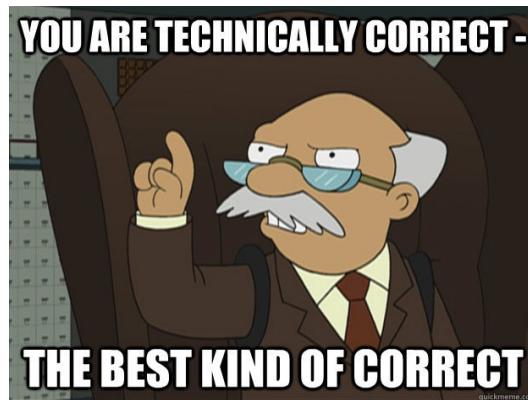
This is correct!:

The confidence interval is based on the sample, and hence is random. There is a 95% probability that the interval (67.9, 68.3) contains μ .



Galton's (fictitious) intern attempts to summarize

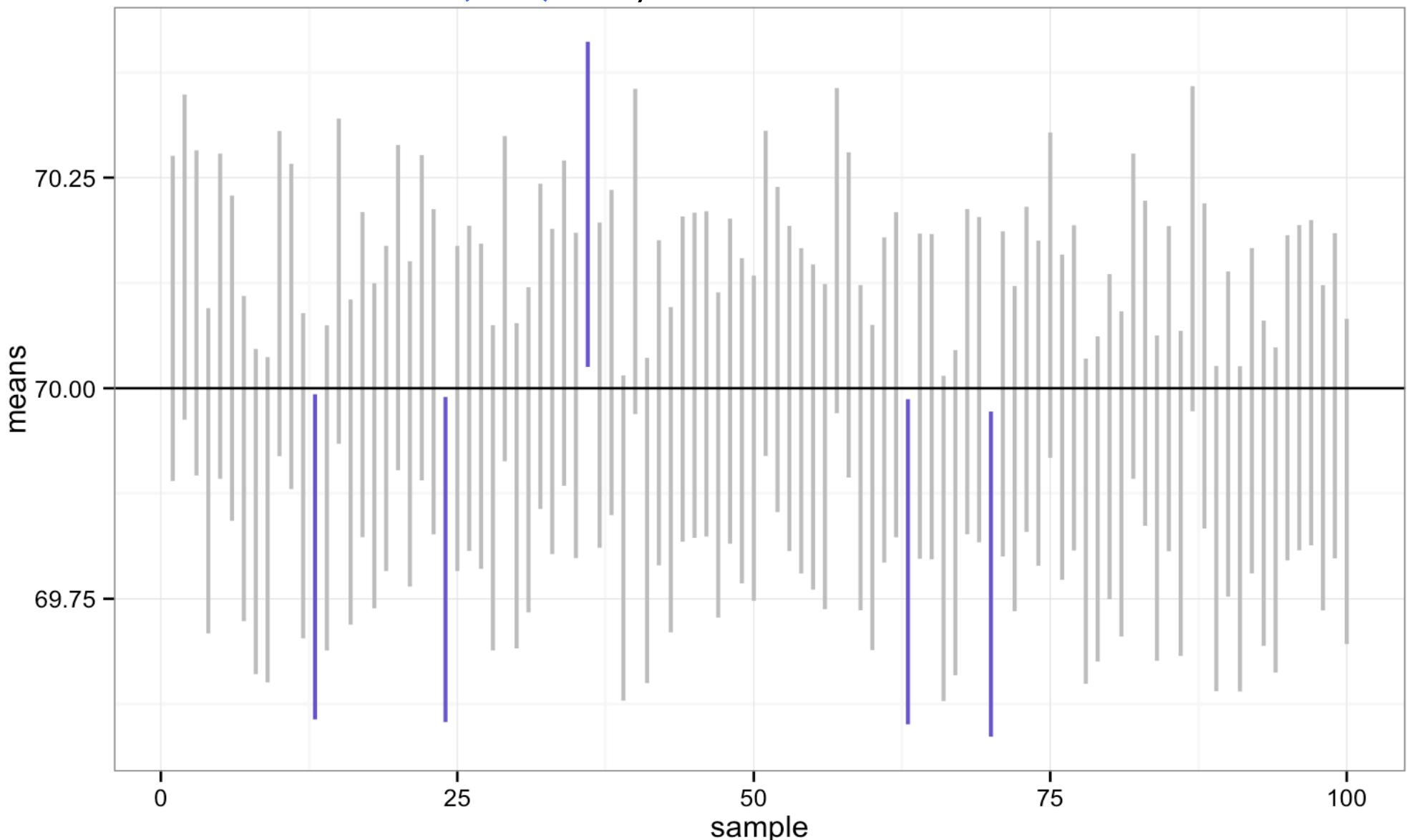
If we repeated the same process of drawing samples and computing confidence intervals, in the long run, 95% of those intervals would include μ .



In this highly fictitious example...

- Note that our 95% confidence interval (based on our observed sample mean) does not include the known population mean, μ !
 - $\mu = 70$
 - Our 95% confidence interval: (67.9, 68.3)
- What are the chances, *if our sample is indeed a random sample from that population?*

I drew 100 random samples of $n = 928$ observations from a normal distribution with $\mu = 70$ and $\sigma = 3$ (just like Galton's sample), calculated their means, then, for each sample, calculated the 95% confidence intervals for the population mean. **95 out of 100 (95%)** of my confidence intervals included 70! **5 did not.**



Confidence intervals for the population mean

Known*: μ
Unknown*: σ

*Again, quite unusual to know either, but especially to know one but not the other!

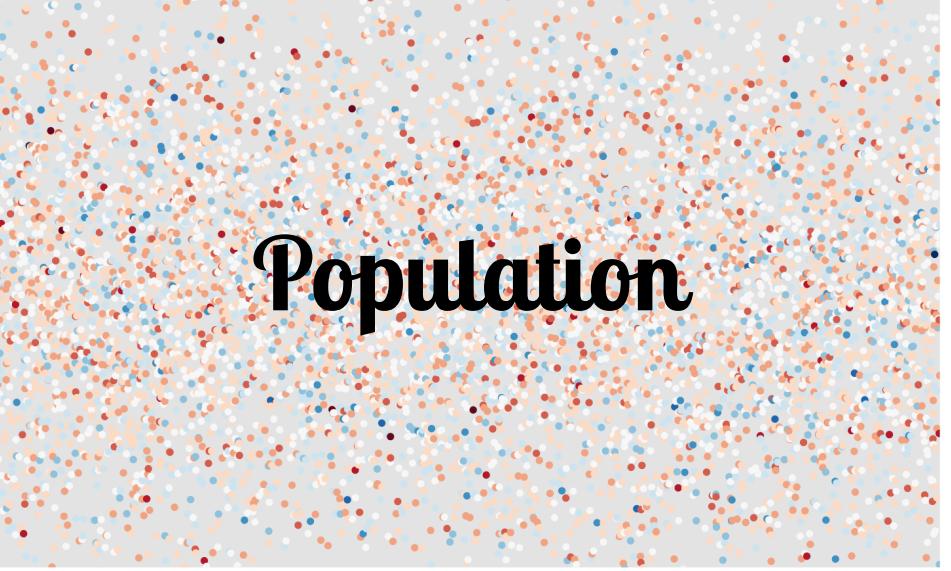
Now imagine...

- Let's assume, again unrealistically, that we know the population mean $\mu = 70$ but not the standard deviation σ . Consequently, we use our sample standard deviation, S , as an estimate of σ .

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

$$\bar{X}_n \sim ? \left(\mu_X, \frac{S_X^2}{n}\right)$$





Population

$$\leftarrow X_i \stackrel{iid}{\sim} N(\mu_X, \sigma_X^2)$$



The sampling distribution of sample means

$$\leftarrow \bar{X}_n \sim \mathcal{N}\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

Because: $E(S^2) = \sigma^2$ and $\hat{Var}(\bar{X}) = \frac{S^2}{n}$

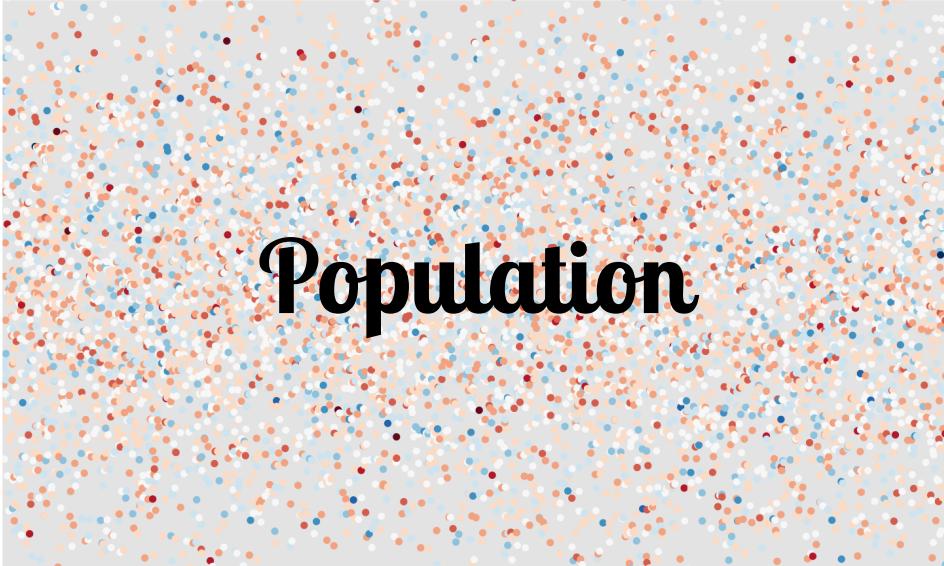
$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}}$$

↓

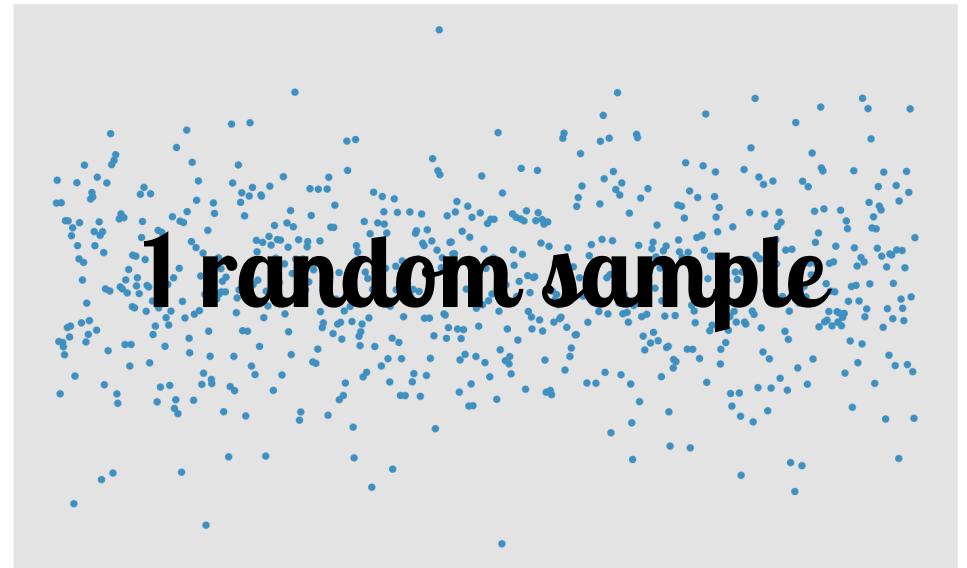
*“The standard error of the sample mean
can be estimated from the sample standard
deviation ($/\sqrt{n}$)”*

Does this change the distribution?



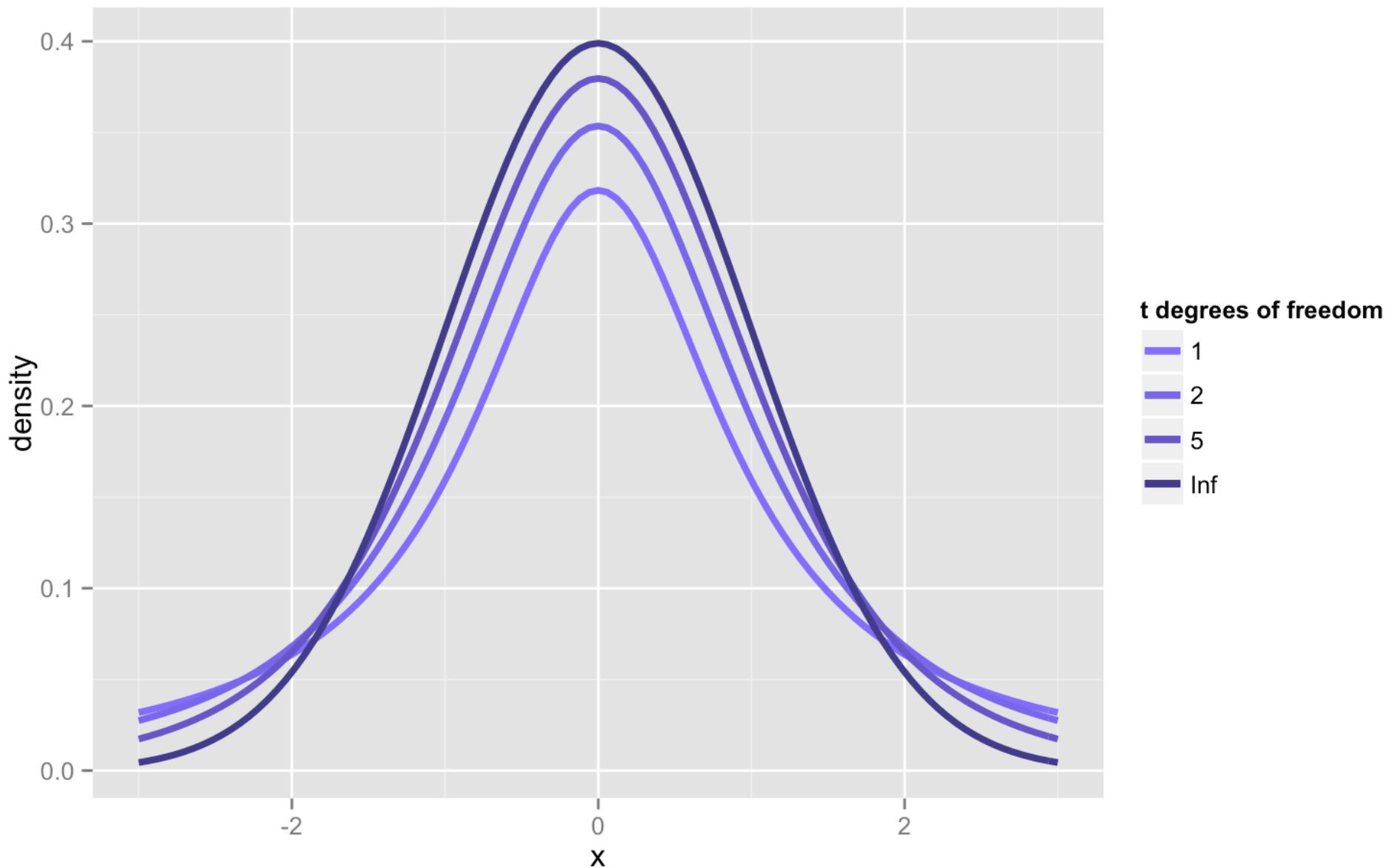


$$\leftarrow X_i \stackrel{iid}{\sim} N(\mu_X, \sigma_X^2)$$



$$\leftarrow \bar{X}_n \sim t_{n-1} \left(\mu_X, \frac{S_X^2}{n} \right)$$

Reminder: student's t

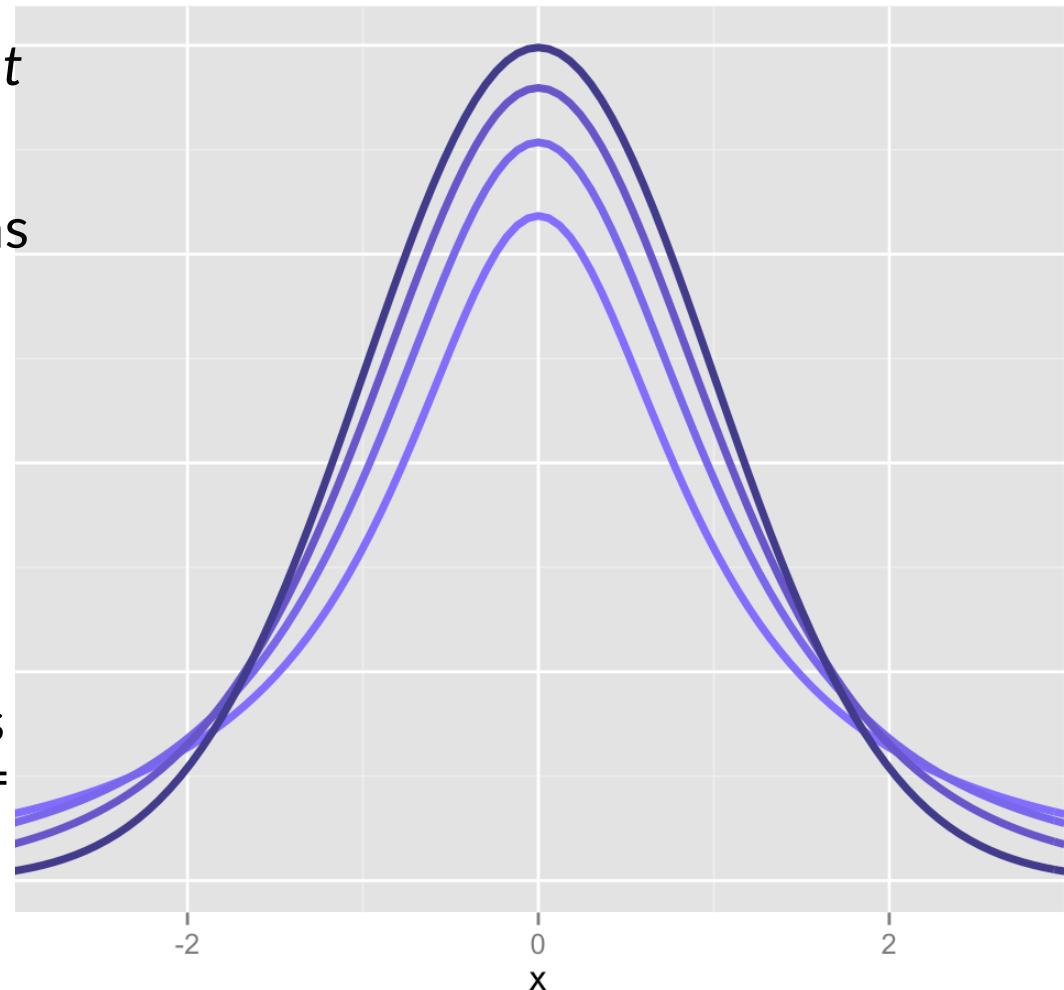


The t -distributions: PDFs

- Symmetric (skewness = 0)
- Bell-shaped, but notice that the t always has relatively more AUC in the tails vs. the unit-normal, and unit-normal has relatively more scores in the center; thus t -distribution is leptokurtic

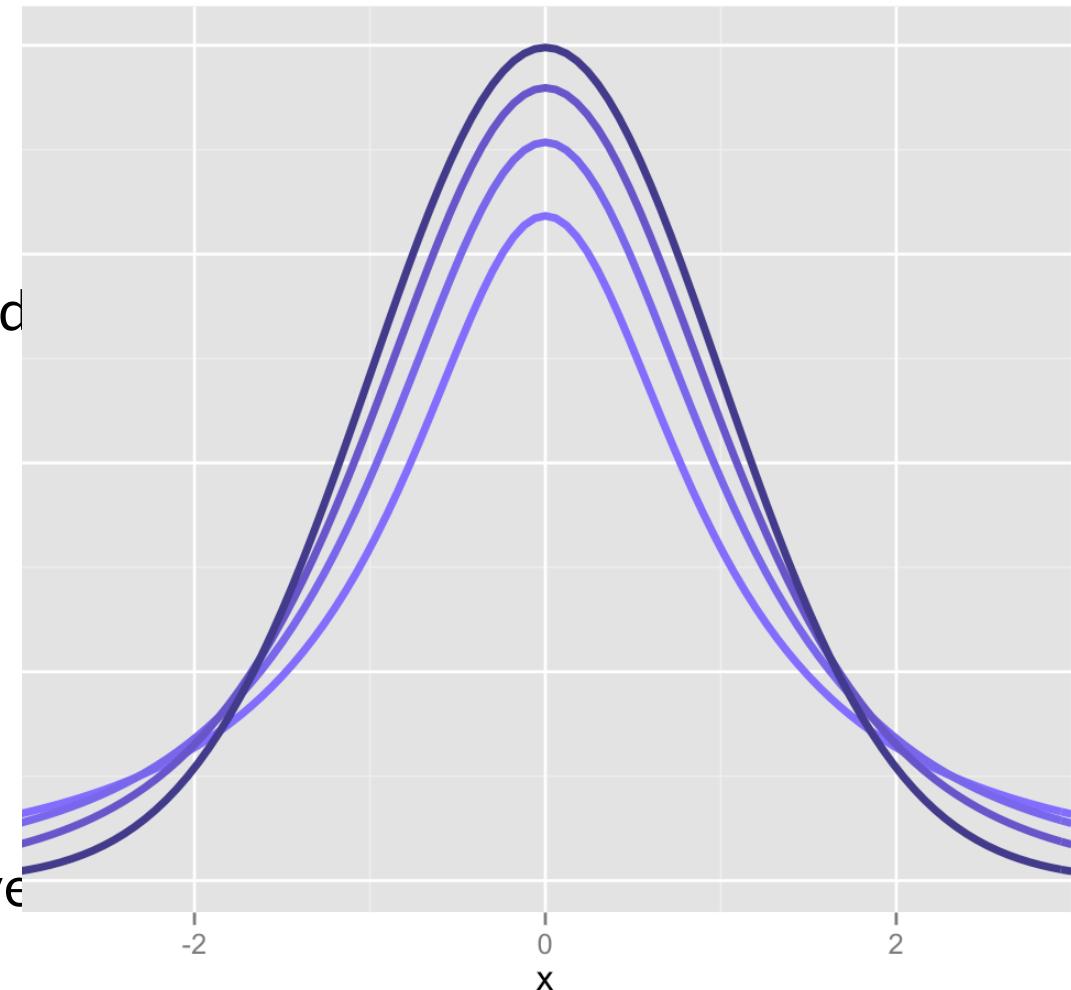
$$= \frac{3(df - 2)}{df - 4}$$

- Kurtosis is undefined for t -variables with $df < 4$
- On your own, at what N and df is the kurtosis of a t -variable = 3 (= to kurtosis of normal distribution)?



The t -distributions

- Let T denote a random variable with a t -distribution with df degrees of freedom. Then:
 - $E(T) = 0$; same as unit-normal
 - $\text{Var}(T) = df/(df-2)$; more spread out than unit-normal (\uparrow variance)
- As df increases, the t -distributions converge to the unit normal.
- t -distributions will be useful for statistical inference for one or more populations of quantitative variables.



Confidence intervals, based on t distribution

- So our new sampling distribution of the means is defined as:

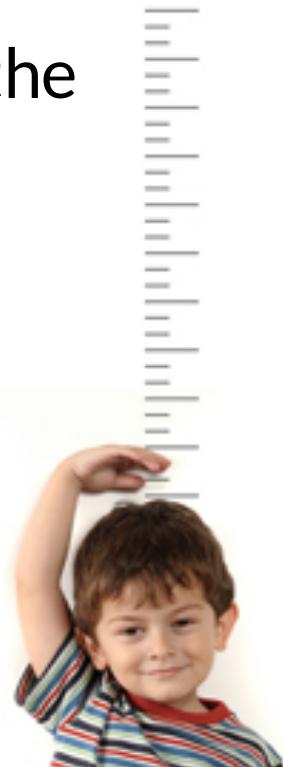
$$\bar{x} \sim t_{n-1} \left(70, \frac{2.52^2}{928} \right)$$

- What we want to know is the confidence interval for the **population mean**:

(lower bound of CI, upper bound of CI)

$(\bar{x} - \text{margin of error}, \bar{x} + \text{margin of error})$

- How do we calculate the “margin of error” here?



Margin of error, based on t distribution

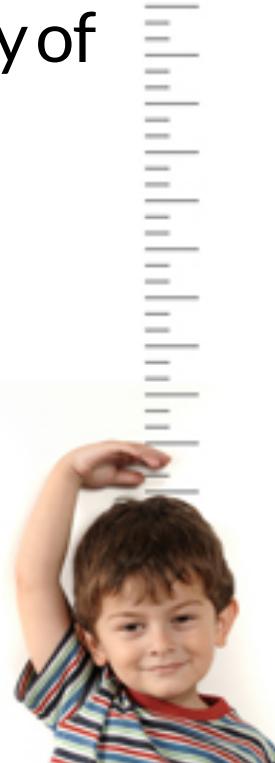
- How do we calculate the “margin of error” here?

$$\bar{x} \sim t_{n-1} \left(70, \frac{2.52^2}{928} \right)$$

The square root of this is the **estimated SEM = .083**

- We use the SEM here because it defines the variability of the sampling distribution for this specific statistic, the sample mean

$$(\bar{x} - \underbrace{\text{here!}}_{\substack{\text{something} \\ \text{What is this something?}}} \times SEM, \bar{x} + \underbrace{\text{here!}}_{\substack{\text{something} \\ \text{What is this something?}}} \times SEM)$$



Confidence interval for μ , σ unknown

```
> mu <- 70 # note that this isn't in any calculations!
> sigma <- 3 # note that this isn't in any calculations!
> n <- 928
> xbar <- 68.08847
> s_x <- 2.517941
> me_t <- qt(.975, n - 1)*(s_x/sqrt(n))
> me_t
[1] 0.1622135
> lower_t <- xbar - me_t
> upper_t <- xbar + me_t
> c(lower_t, upper_t)
[1] 67.92626 68.25068
```

I am 95% confident that the population mean height for children (in this age range) is between 67.9 and 68.3 inches!

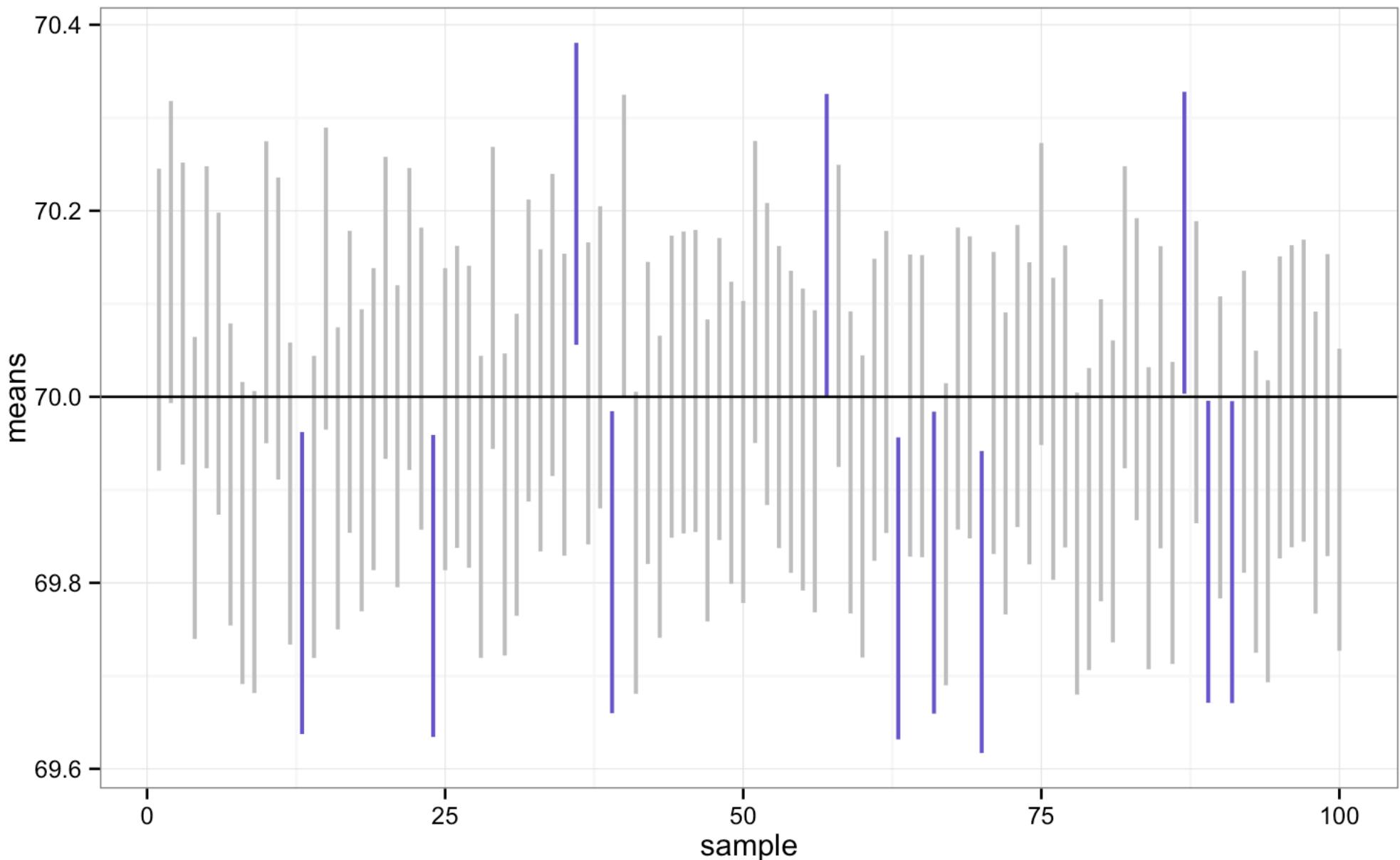


Again, where is μ ?

$$(\bar{x} - t_{\alpha/2} \frac{S_x}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \frac{S_x}{\sqrt{n}})$$

$$P(\bar{x} - t_{\alpha/2} \frac{S_x}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \frac{S_x}{\sqrt{n}}) = 1 - \alpha$$

Same data as before: now using the t-distribution and sample variance estimate to calculate 95% confidence intervals for 100 random samples of $\mu = 70$ and $\sigma = 3$. [89 out of 100 \(89%\)](#) of my confidence intervals included 70! [11 did not.](#)



What have we done?

- First, our highly contrived example here had 928 observations, so our two different confidence intervals were actually pretty similar- this is because the t-distribution is asymptotically normal- the biggest differences you would see between the t and the normal are for $n < 30$.
- In general, confidence intervals based on the t-distribution will always be more narrow than those based on the normal distribution.
- Same thing said a different way: confidence intervals based on the normal distribution, when we know μ and σ , will always be wider than those based on the t-distribution, when we don't know these population parameters, because we account for variability due to our sample estimates.
- In essence, we make it harder on ourselves on purpose.

Z confidence interval for normal mean with known σ

- For:

$$X_i \stackrel{iid}{\sim} N(\mu_X, \sigma_X^2)$$

- The $(1 - \alpha) \times 100\%$ confidence interval for μ is given by:

$$\left(\bar{x} - q \times \frac{\sigma_X}{\sqrt{n}}, \bar{x} + q \times \frac{\sigma_X}{\sqrt{n}} \right)$$

- Where q denotes the $(1 - \alpha/2)$ quantile of $N(0,1)$

T confidence interval for normal mean with unknown σ

- For:

$$X_i \stackrel{iid}{\sim} N(\mu_X, \sigma_X^2)$$

- The $(1 - \alpha) \times 100\%$ confidence interval for μ is given by:

$$\left(\bar{x} - q \times \frac{S_x}{\sqrt{n}}, \bar{x} + q \times \frac{S_x}{\sqrt{n}} \right)$$

- Where q denotes the $(1 - \alpha/2)$ quantile of the t -distribution with $(n - 1)$ degrees of freedom

And now: the easy way

Using the normal distribution, $\sigma = 3$

```
> library(TeachingDemos)
> z.test(galton$child, sd = 3)

One Sample z-test

data: galton$child
z = 691.3951, n = 928.000, Std. Dev. =
3.000, Std. Dev. of the sample mean =
0.098,
p-value < 2.2e-16
alternative hypothesis: true mean is not
equal to 0
95 percent confidence interval:
 67.89545 68.28149
sample estimates:
mean of galton$child
 68.08847
```

$$\Delta = 0.3860338$$

Using the t-distribution, $s = 2.52$

```
> t.test(galton$child)

One Sample t-test

data: galton$child
t = 823.7624, df = 927, p-value < 2.2e-
16
alternative hypothesis: true mean is not
equal to 0
95 percent confidence interval:
 67.92626 68.25068
sample estimates:
mean of x
 68.08847
```

$$\Delta_t = 0.3244271$$

Relation to hypothesis testing

- Both the z and t-tests were “significant”- what was the hypothesis we were testing? The most dull one in the world- that the sample mean is equal to a hypothesized value of the population mean. In this case, the hypothesized value is that $\mu = 0$. So the tests reflect that indeed our sample mean is significantly different from zero ($p < .05$).
- Clearly, it makes no sense to hypothesize that the population mean height of children in our age range of interest is 0.

Confidence intervals provide several advantages over NHST alone

- Provides bounded estimate of the population parameter
- Permits tests of all possible null hypotheses simultaneously
 - You can reject/not reject H_0 not only for value hypothesized in H_0 but also for other hypothesized values
- Interval width provides information about precision
 - Significance plus very narrow interval suggests power so great as to enable us to reject even a trivial effect
 - Non-significance with wide intervals suggests that our measures/procedures/experiment lacks precision
 - p -values do not give you this information, heavily influenced by sample size (n)

Constructing Confidence Intervals

- *Think backward: invert a significance test.* Let the confidence interval include all values of a sample statistic for which a two-sided significance test **would not reject** the null hypothesis
 - Level of significance (α) is equivalent to $(1-\alpha)\%$ confidence interval
- Intervals constructed around the *sample statistic, not the population parameter*
 - Although not constructed around the population parameter, the CI tells you exactly what you want to know about that parameter: it is the interval of reasonable values for that parameter given the sample data
- Does my hypothesized value fall within or outside of the CI?
 - *Within:* do not reject H_0 , hypothesized value is reasonable given sample
 - *Outside:* reject H_0 , hypothesized value is not reasonable given sample
- In one-sample tests, hypothesized value = population parameter

Hypothesis Tests and Confidence Intervals

- Suppose: z-test with $\alpha = .05$, 2-tailed
- $H_0: \bar{Y}_\cdot = \mu$
- $H_1: \bar{Y}_\cdot \neq \mu$

$$z_{obs} = \frac{\bar{Y}_\cdot - \mu}{\sigma / \sqrt{n}}$$

- I will reject H_0 if:

$$\frac{\bar{Y}_\cdot - \mu}{\sigma / \sqrt{n}} < -1.96$$

$$\frac{\bar{Y}_\cdot - \mu}{\sigma / \sqrt{n}} > 1.96$$

- Some algebra to re-arrange, I will reject H_0 if:

$$\mu < \bar{Y}_\cdot - 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\mu > \bar{Y}_\cdot + 1.96 \frac{\sigma}{\sqrt{n}}$$

- And I will not reject H_0 if μ lies between those 2 rejection points:

$$\left(\bar{Y}_\cdot - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y}_\cdot + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

This is my 95% confidence interval for μ

Hypothesis Tests and Confidence Intervals

- The confidence interval also tells when to reject H_0 - I will reject H_0 if μ is outside of this interval:

$$\left(\bar{Y}_\bullet - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y}_\bullet + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

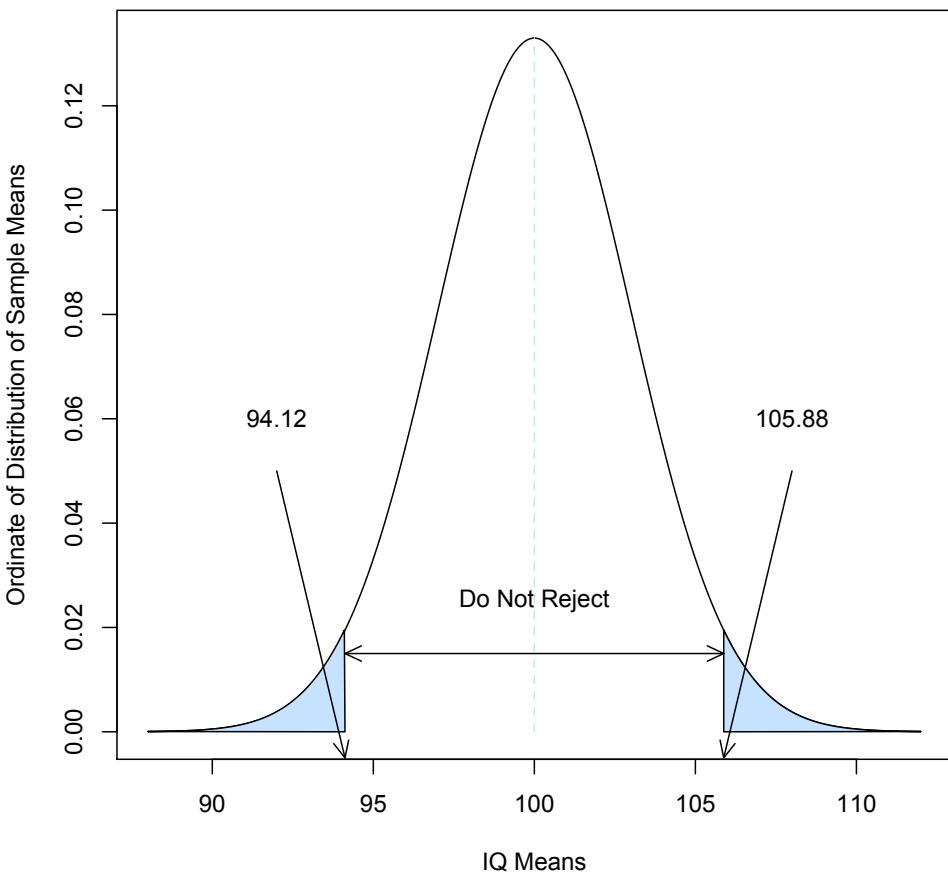
- But what if instead of solving for μ , I solve for the sample mean? I will reject H_0 if:

$$\bar{Y}_\bullet < \mu - 1.96 \frac{\sigma}{\sqrt{n}} \quad \quad \quad \bar{Y}_\bullet > \mu + 1.96 \frac{\sigma}{\sqrt{n}}$$

- These two values define both of my regions of rejection (2-tailed).
- My region of non-rejection is all values in between those 2 rejection points- I will not reject H_0 if my sample mean is within:

$$\left(\mu - 1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{Y}_\bullet \leq \mu + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

Regions of rejection/non-rejection: What if I observe a sample mean = 106?



- **Rejection regions**

I reject if:

$$\bar{Y}_\bullet < 100 - 1.96 \frac{15}{\sqrt{25}} = 94.12$$

or if:

$$\bar{Y}_\bullet > 100 + 1.96 \frac{15}{\sqrt{25}} = 105.88$$

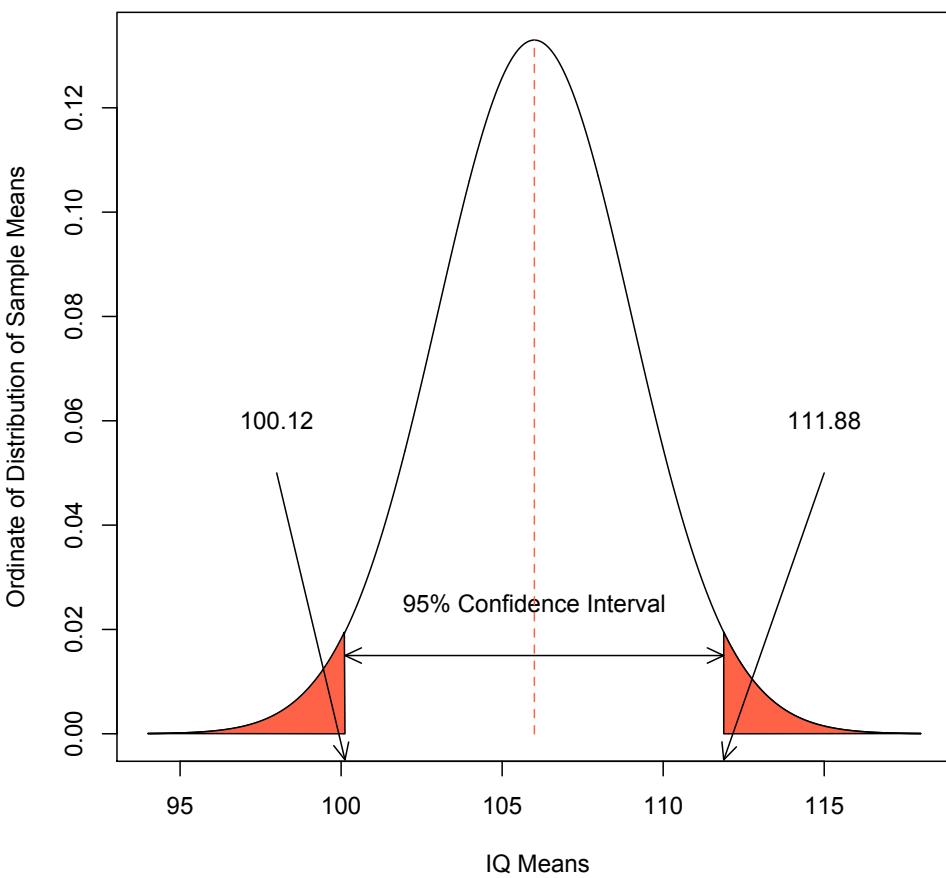
- **Non-rejection region**

I will not reject if:

$$(94.12 \leq \bar{Y}_\bullet \leq 105.88)$$

- With observed sample mean=106, I reject H_0

Confidence interval: What if I had hypothesized $\mu = 100$?

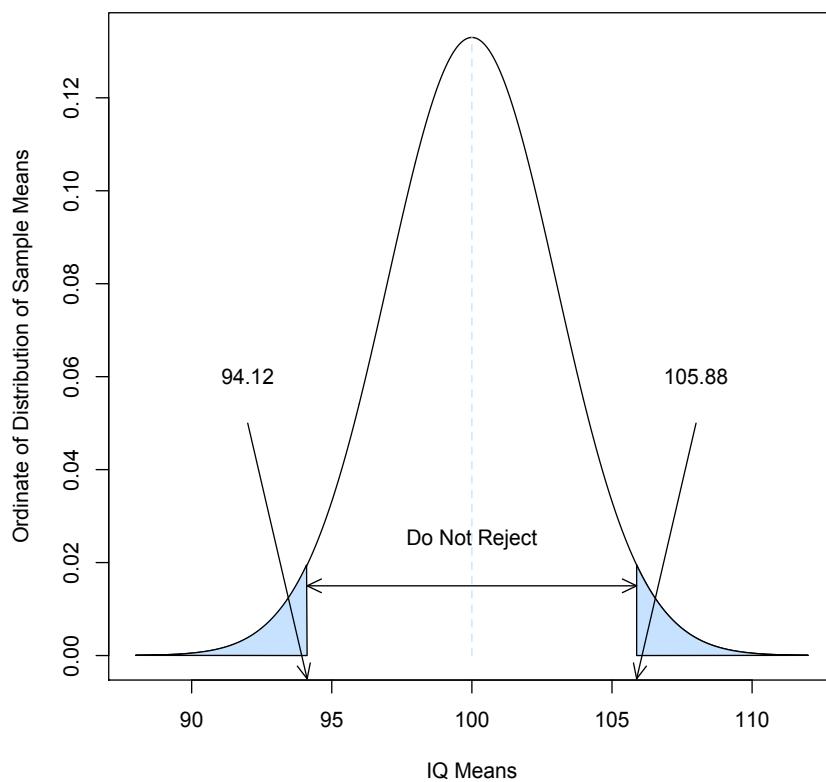


- **95% confidence interval**
I will not reject H_0 if μ is within:
$$\left(106 - 1.96 \frac{15}{\sqrt{25}} \leq \mu \leq 106 + 1.96 \frac{15}{\sqrt{25}} \right)$$
$$(100.12 \leq \mu \leq 111.88)$$
- Because $\mu=100$ is not contained in this interval, I reject H_0 as it does not contain my population value
- Note that this is the same decision I reached with my regions of rejection
- I also know more than I did before: I know that I am 95% confident that the true value of μ is between 100.12 and 111.88

Decision Rule: Reject H_0 if (a) sample mean is outside of RoR
 OR if (b) μ is not within CI
 (note: if (a) is true, so is (b)-these are equivalent)

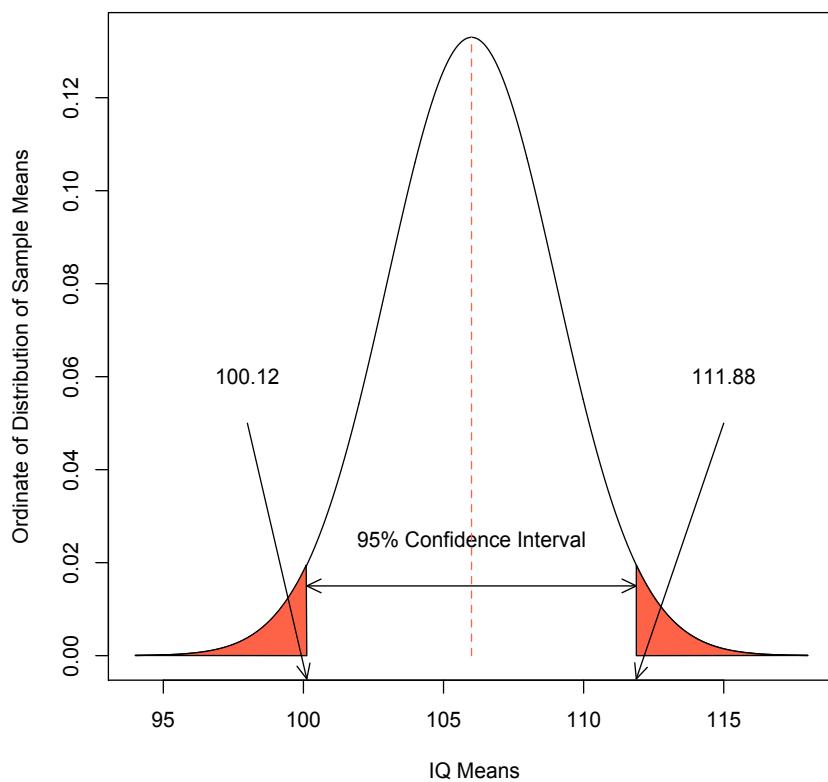
RoR: $\mu = 100$ (population parameter/hypothesized value)

Given that $\mu = 100$, is my sample mean a reasonable value from that population?



Confidence interval: $\mu = 106$ (sample statistic)

Given that my sample mean = 106, is μ a reasonable value for the population mean?



Note: The standard deviation used to create both sampling distributions is 3. Why?

Decision-making in statistical inference

	Reject H_0 /Fail to reject H_1	Fail to reject H_0 /Reject H_1
RoR	If statistic value is <i>within</i> RoR	If statistic value is <i>outside</i> RoR
CI	If population parameter is <i>outside</i> CI	If population parameter is <i>within</i> CI
p -value	p -value will be $< \alpha$	p -value will be $\geq \alpha$

Power

- Power is the probability of correctly rejecting a false null hypothesis

$$1 - \beta = P(\text{reject } H_0 \mid H_1 \text{ true})$$

- Suppose we wish to test ($\alpha = .025$, 1-tailed):
 - $H_0: \mu \leq 100$
 - $H_1: \mu > 100$
- Let's use the same sample of $n=25$ aspiring astronauts (we know $\sigma = 15$)
- First, what is β ?

$$\beta = P(\text{fail to reject } H_0 \mid H_1 \text{ true})$$

Let's do a one-tailed t-test...

```
> aat_1 <- t.test(iq_aa, mu = 100, alternative =  
+ c("greater"))  
> aat_1
```

One Sample t-test

```
data: iq_aa  
t = 1.9227, df = 24, p-value = 0.03323  
alternative hypothesis: true mean is greater than 100  
95 percent confidence interval:  
100.5509      Inf  
sample estimates:  
mean of x  
105
```



If the null is true...

- The test statistic under the null will have a central t distribution with $v = n - 1 = 24$ degrees of freedom.
- The (one-tailed) critical value will be:

```
> qt(.95, 24) # tcritical, null dist  
[1] 1.710882
```



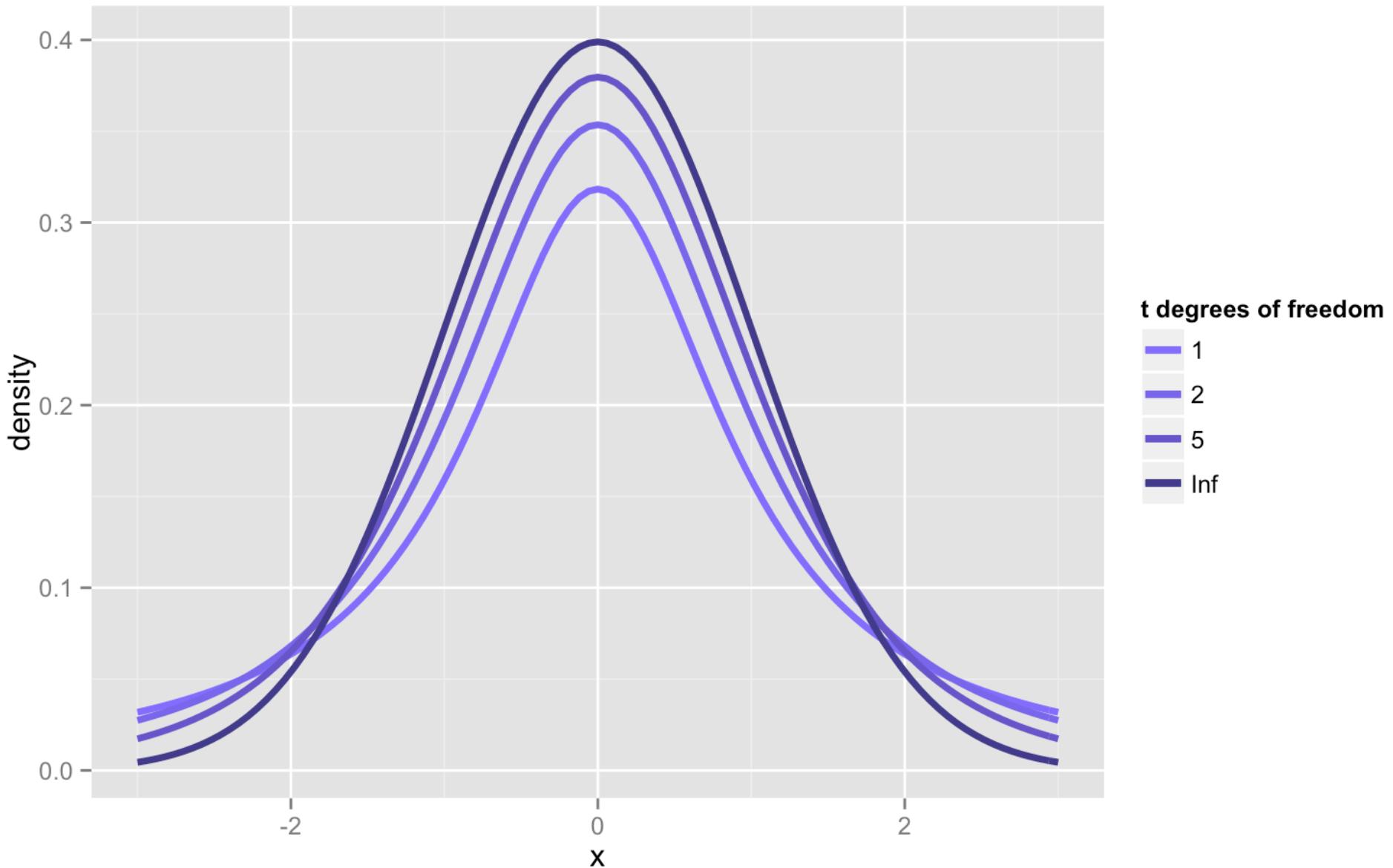
- $P(\text{false positive}) = \alpha = \text{Type I error rate} = .05$
- $P(\text{false negative}) = \beta = \text{Type II error} = ?$
- $P(\text{true positive}) = 1 - \beta = \text{power} = ?$



Finding β (and $1 - \beta$)

- Need to know exact **null** distribution (just as with NHST)
- Also need to know exact **alternative** distribution of the test statistic
 - Often requires some specialized statistical knowledge
- In general, it is much more likely that expressions for the null distribution of the test statistic will be available than expressions for the non-null distribution.

Recall student's t -distribution



The Student t Distribution

Description

Density, distribution function, quantile function and random generation for the t distribution with `df` degrees of freedom (and optional non-centrality parameter `ncp`).

Usage



```
dt(x, df, ncp, log = FALSE)
pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)
qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)
rt(n, df, ncp)
```

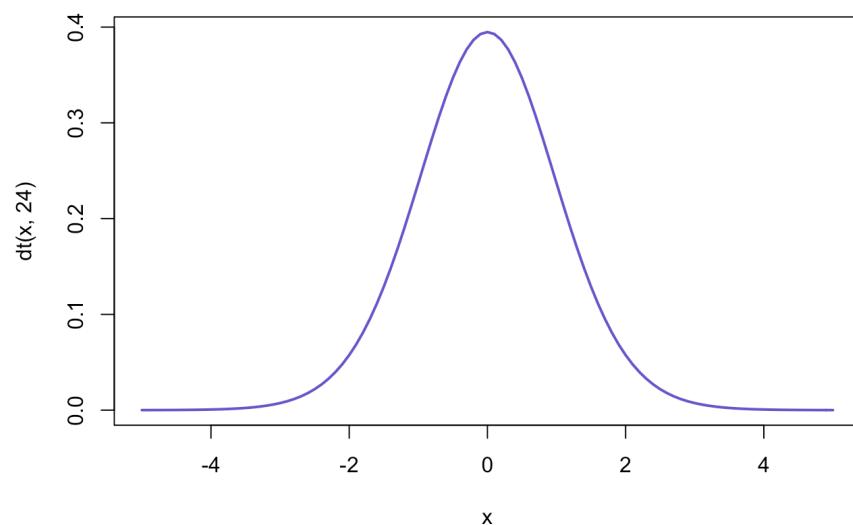
Arguments

- `x, q` vector of quantiles.
- `p` vector of probabilities.
- `n` number of observations. If `length(n) > 1`, the length is taken to be the number required.
- `df` degrees of freedom (> 0 , maybe non-integer). `df = Inf` is allowed.
- `ncp` non-centrality parameter *delta*; currently except for `rt()`, only for $\text{abs}(ncp) \leq 37.62$. If omitted, use the central t distribution.



Central t-distribution (v = degrees of freedom)

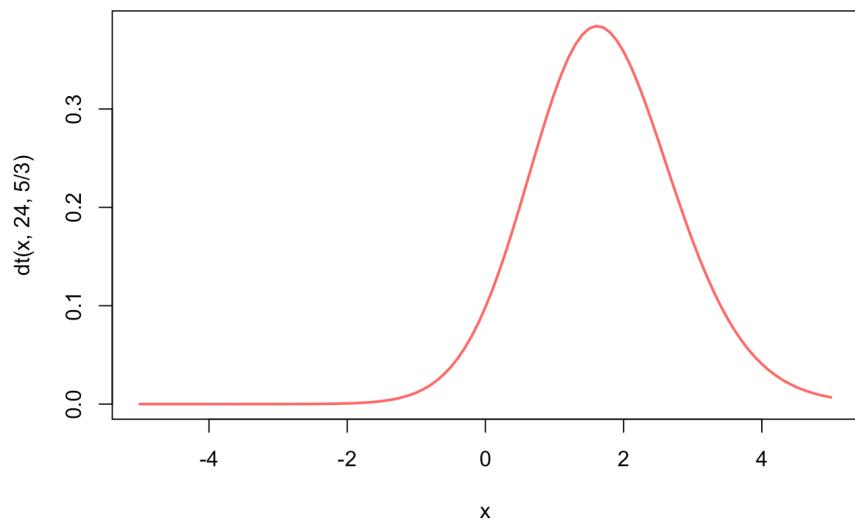
- $\delta = 0$
- Mean = 0
- Variance slightly > than $N(0, 1)$
- Kurtosis (biased) is > 3
- Symmetric

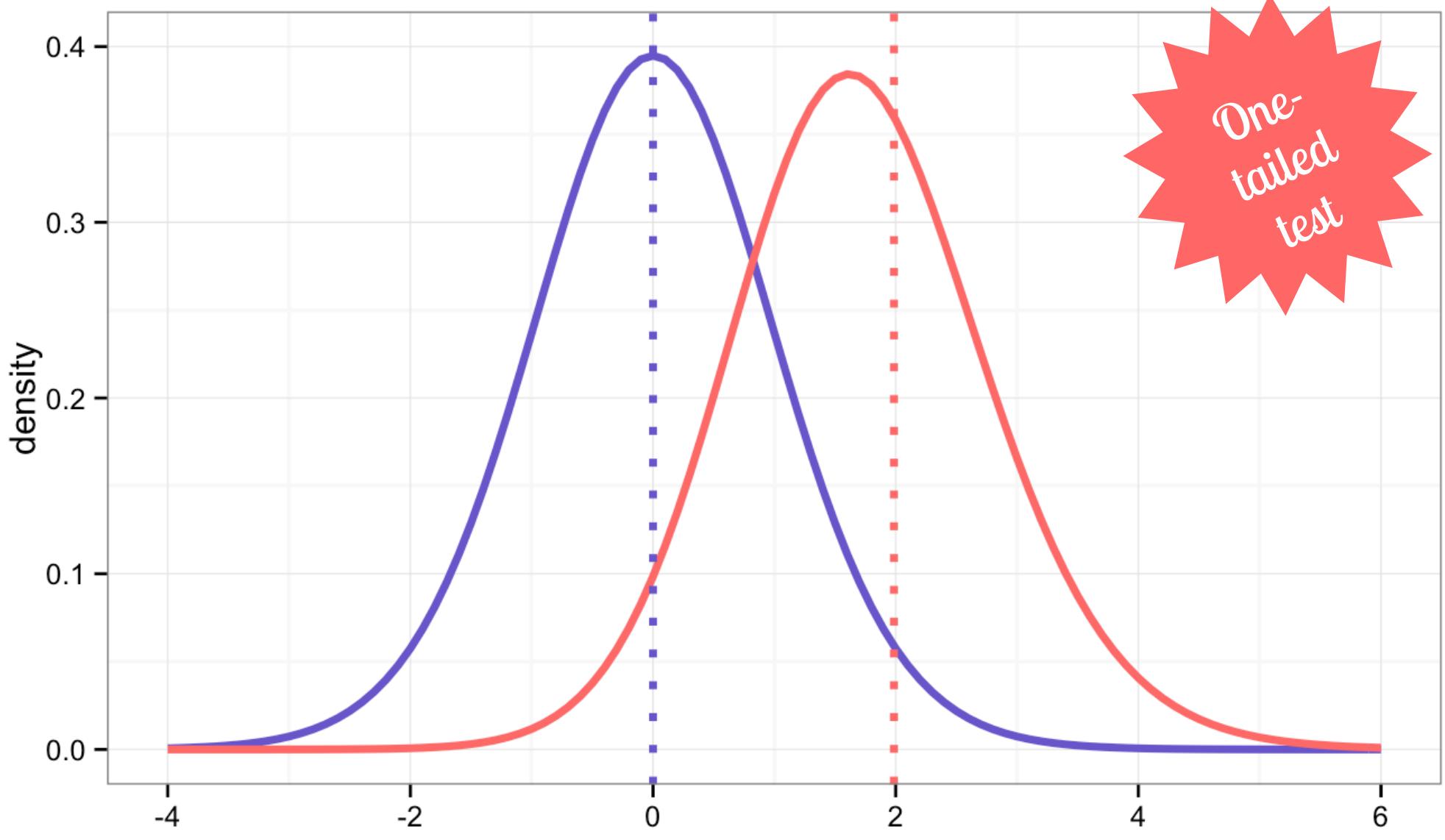


Noncentral t distribution (ν = degrees of freedom)

- $\delta \neq 0$
- Asymmetric: skewed in the direction of δ

$$E(T) = \begin{cases} \delta \sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} & \text{if } \nu > 1 \\ \text{Does not exist} & \text{if } \nu \leq 1 \end{cases}$$





Null:
Central t Alt:
Noncentral t

How do we calculate the ncp?

- The noncentrality parameter (ncp) is defined as:

$$\delta = \sqrt{n}E_s$$

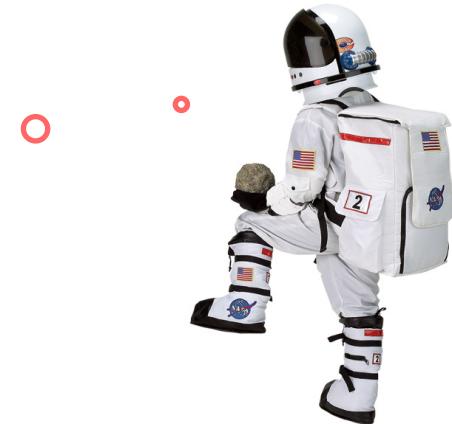
- Where E_s is the standardized measure of effect size...how do we calculate this?

Effect size

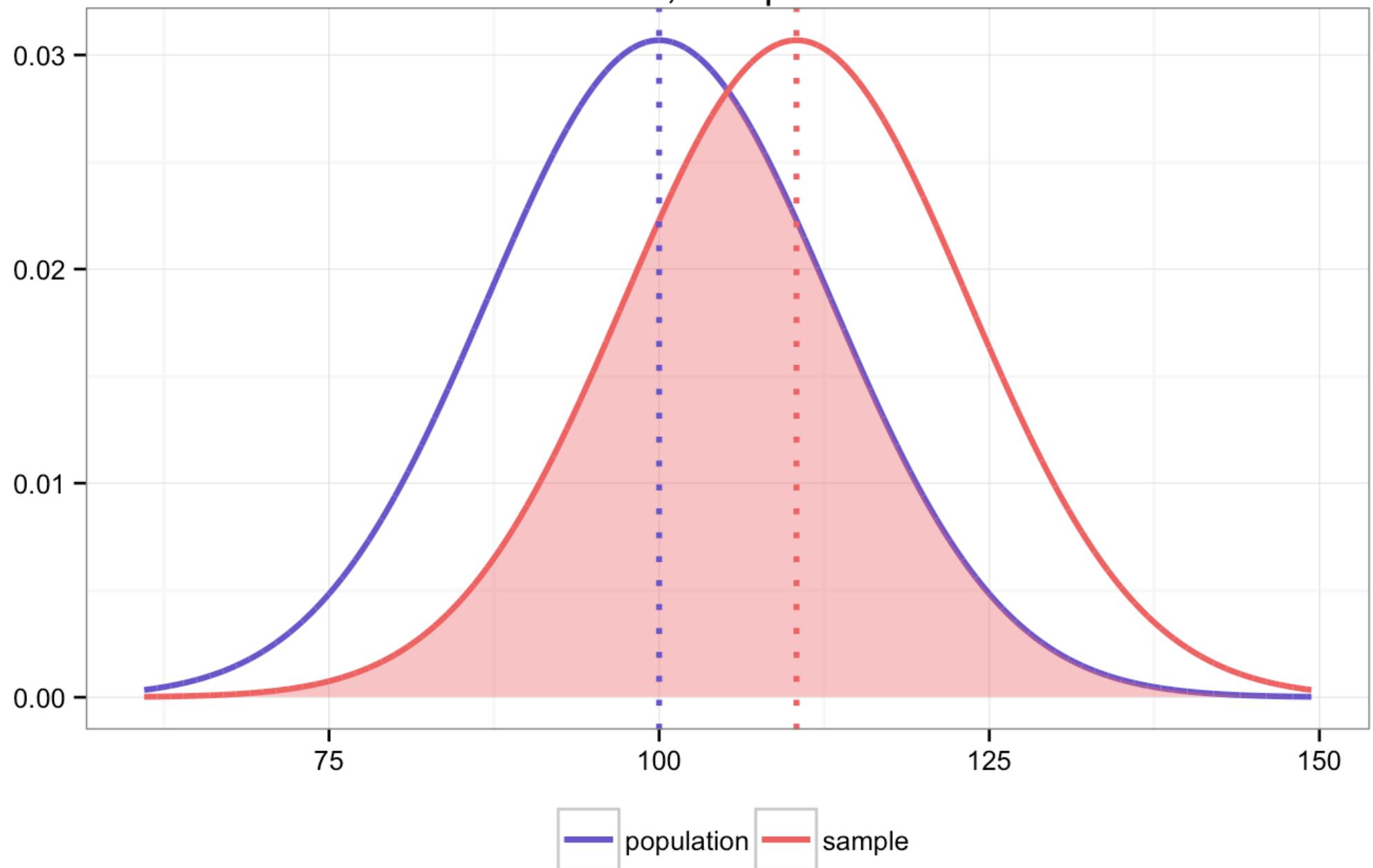
$$t_{\nu} = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

$$E_s = \frac{\mu_1 - \mu_0}{s}$$

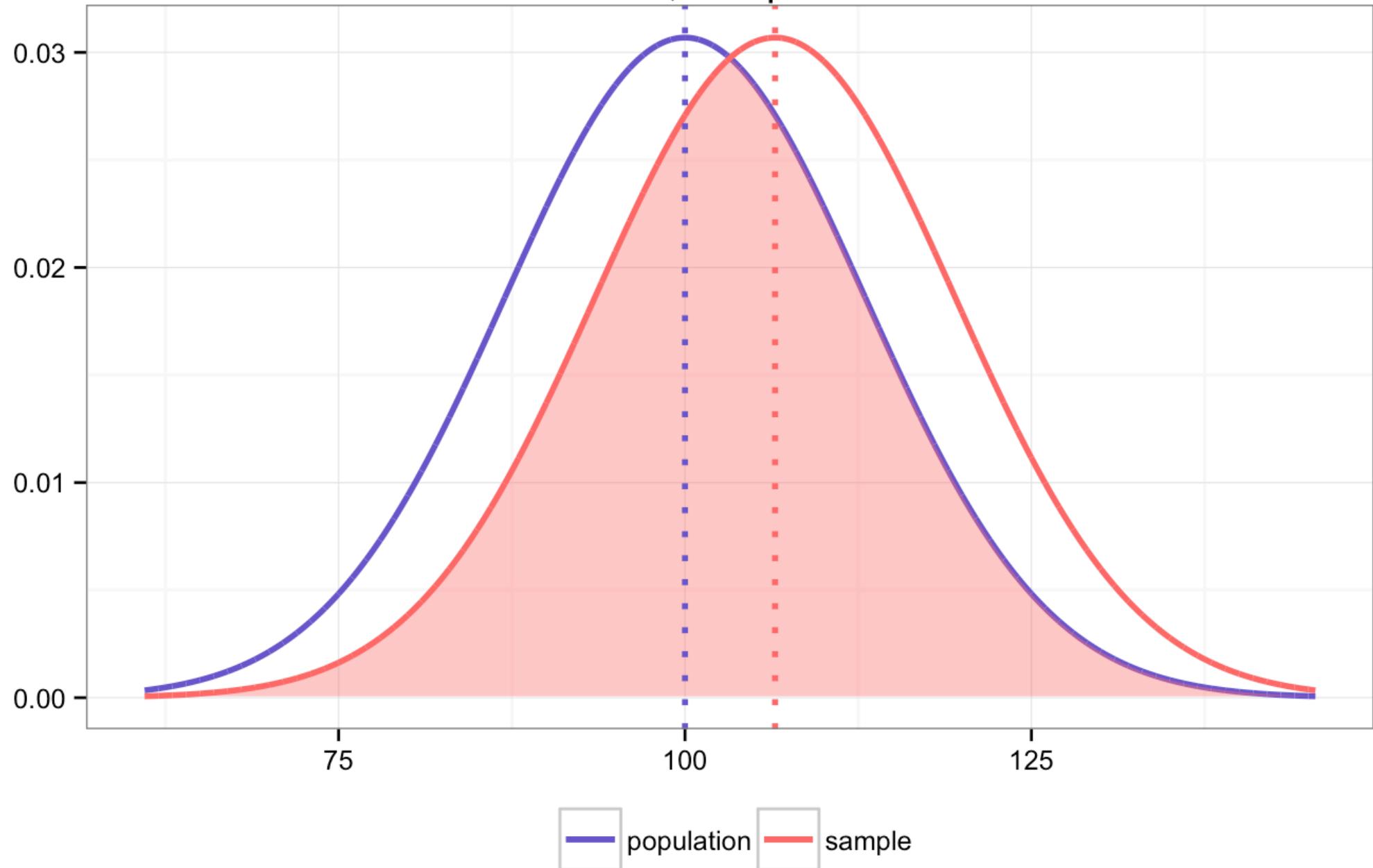
What is **not** in this formula?



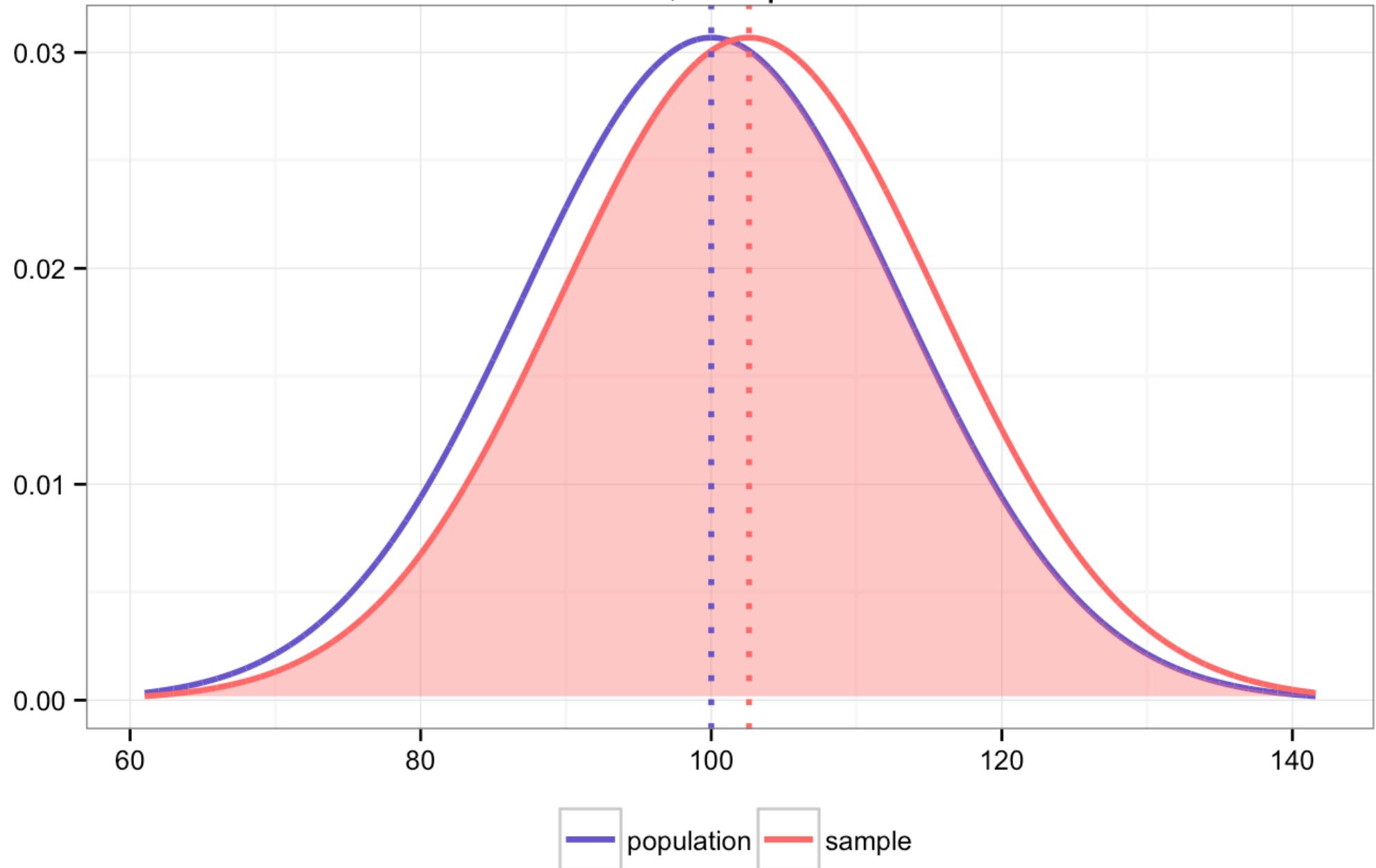
Effect size = 0.8; sample mean = 110.4



Effect size = 0.5; sample mean = 106.5



Effect size = 0.2; sample mean = 102.6



Calculating effect size and ncp

- In our example, the E_s is defined just by the sample mean, null mean, and the sample s.d., so the ncp is:

$$\begin{aligned}\delta &= \sqrt{n}E_s \\ &= \sqrt{25} \times \frac{105 - 100}{13} \\ &= 5 \times \frac{5}{13} = 1.923077\end{aligned}$$



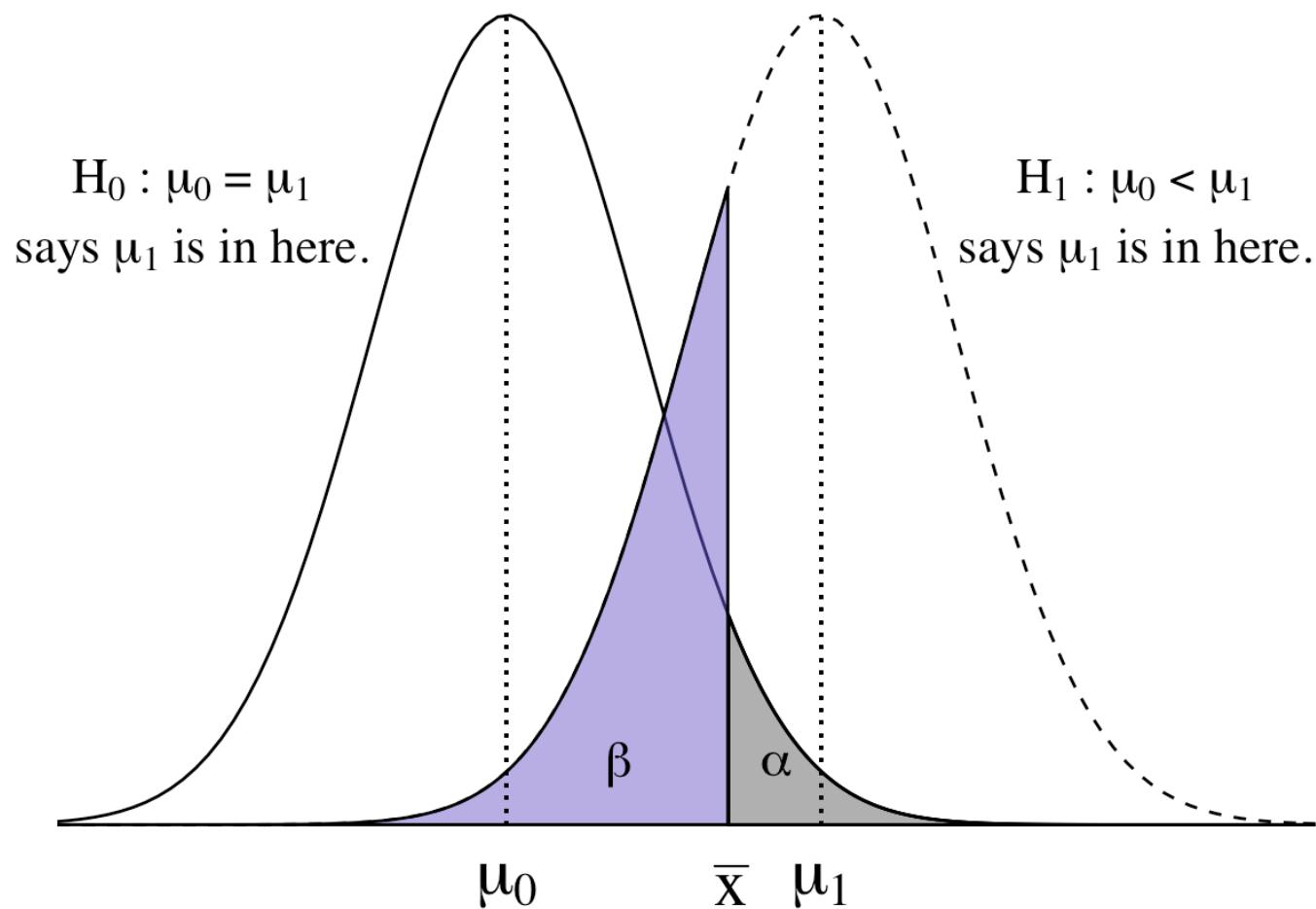
The null and alternative t-distributions

Distribution of t under H_0 is $t_{\nu=24, \delta=0}$

Distribution of t under H_1 is $t_{\nu=24, \delta=1.923}$



TYPE II ERROR: $p(\text{false negative}) = \beta$



TYPE II ERROR: $p(\text{false negative}) = \beta$

For a given decision,

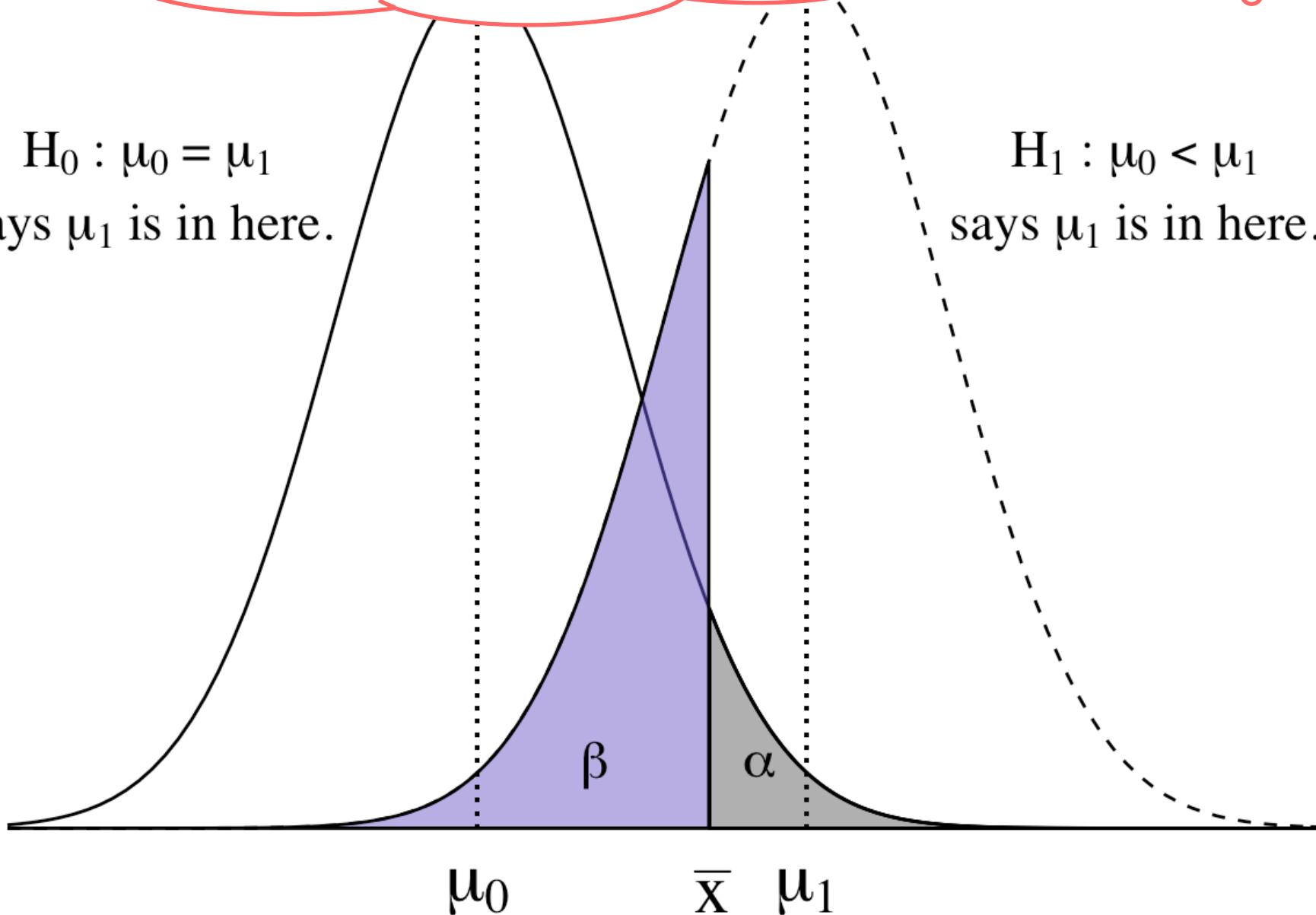
- $P(\text{false positive}) = \alpha = \text{Type I error}$
- $P(\text{false negative}) = \beta = \text{Type II error}$

What will happen to β if I make α smaller?

What will happen to β if I make
 α smaller?

$H_0 : \mu_0 = \mu_1$
says μ_1 is in here.

$H_1 : \mu_0 < \mu_1$
says μ_1 is in here.



$\alpha \uparrow \rightsquigarrow \beta \downarrow$

$\alpha \downarrow \rightsquigarrow \beta \uparrow$

TYPE II ERROR: $p(\text{false negative}) = \beta$

- Probability is defined as the chances of observing a t-statistic more extreme than the one we observed given the alternative t-distribution
- “How much more wrong could we have been?”



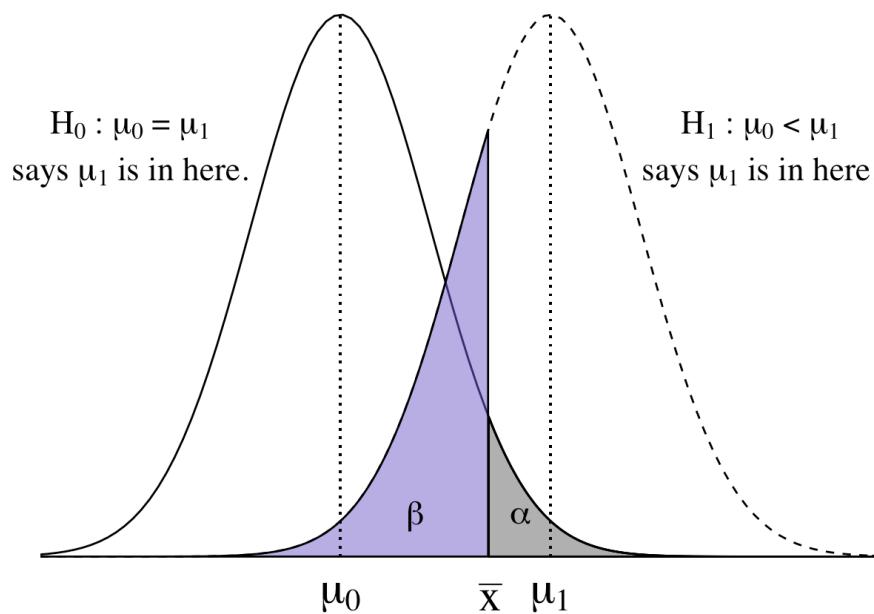
In our aspiring astronauts example...

```
qt(.95, 24)
```

```
[1] 1.710882
```

```
pt(qt(0.95, 24), 24, 25/13)
```

```
[1] 0.4115342 # beta
```



Type II error (β)
False negative



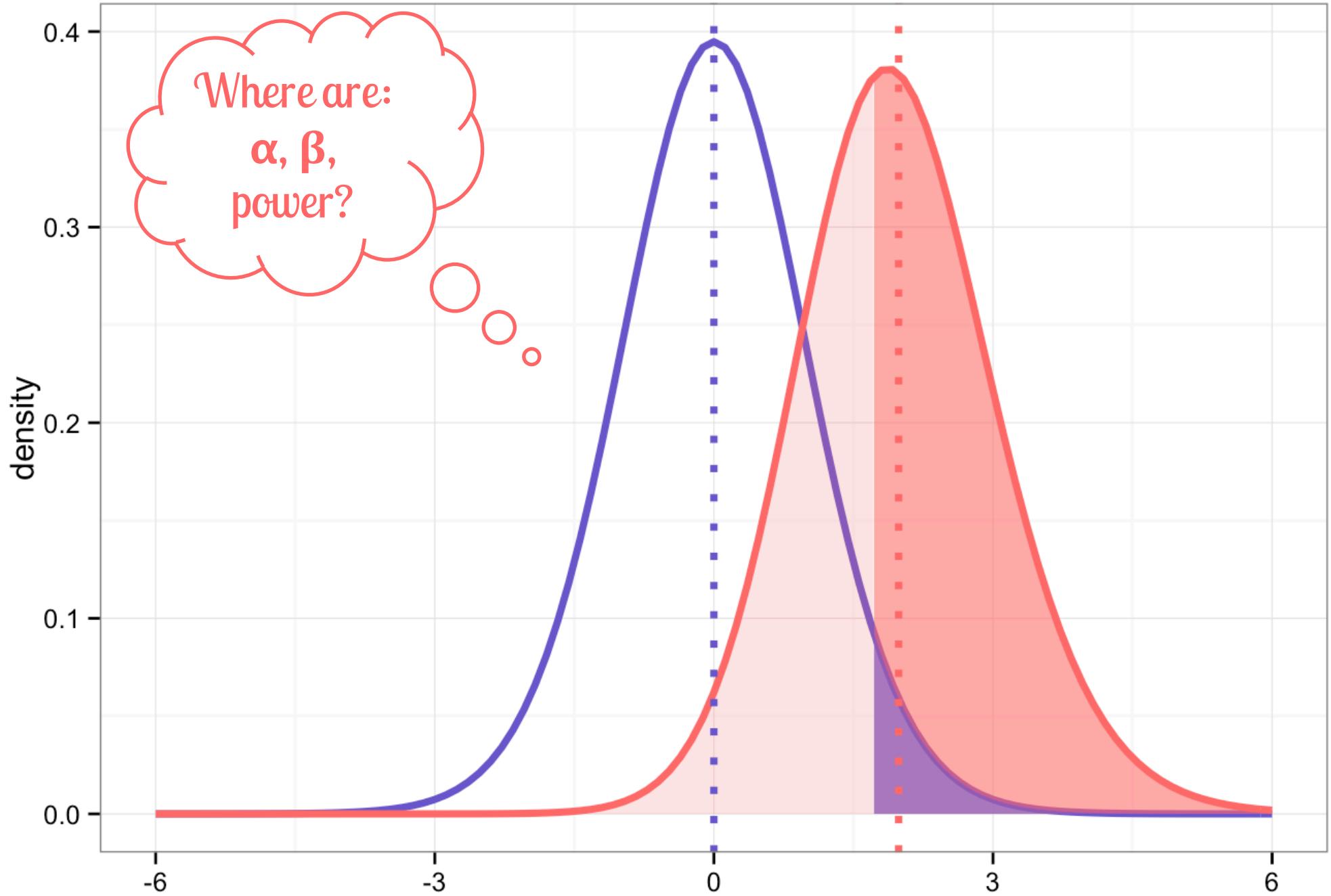
Decide:
“You’re **not** smarter than average!”
Reality:
but you **are** actually

POWER: $p(\text{true positive}) = 1 - \beta$

- So power is the probability of exceeding the rejection point in this noncentral t distribution.

```
qt(.95, 24) # tcritical, null dist  
[1] 1.710882  
  
pt(qt(0.95, 24), 24, 25/13) # beta  
[1] 0.4115342  
  
1 - pt(qt(0.95, 24), 24, 25/13) # power  
[1] 0.5884658
```





Power

```
power.t.test(n = 25, delta = 5, sd = 13, type = "one.sample",
alternative = c("one.sided"))
```

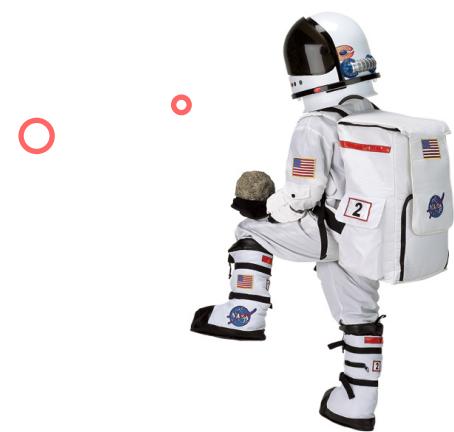
One-sample t test power calculation

```
n = 25
delta = 5
sd = 13
sig.level = 0.05
power = 0.5884658
alternative = one.sided
```



Good for: “post-mortem” power analysis

delta here is confusing: it is neither the ncp nor the effect size- it is the raw difference between means you wish to detect



Factors that affect Power ($1 - \beta$)

- Sample size
 - Increased n reduces SE_{mean}
- Level of significance
 - Power increases as α increases
- Reliability of your measure
 - Classical test theory:
total variance = true score variance + error variance
- Effect size (sds between the true mean & the one hypothesized in H_0 ; $\mu - \mu_0$)
- Population variance
 - Decreased variance reduces SE_{mean}

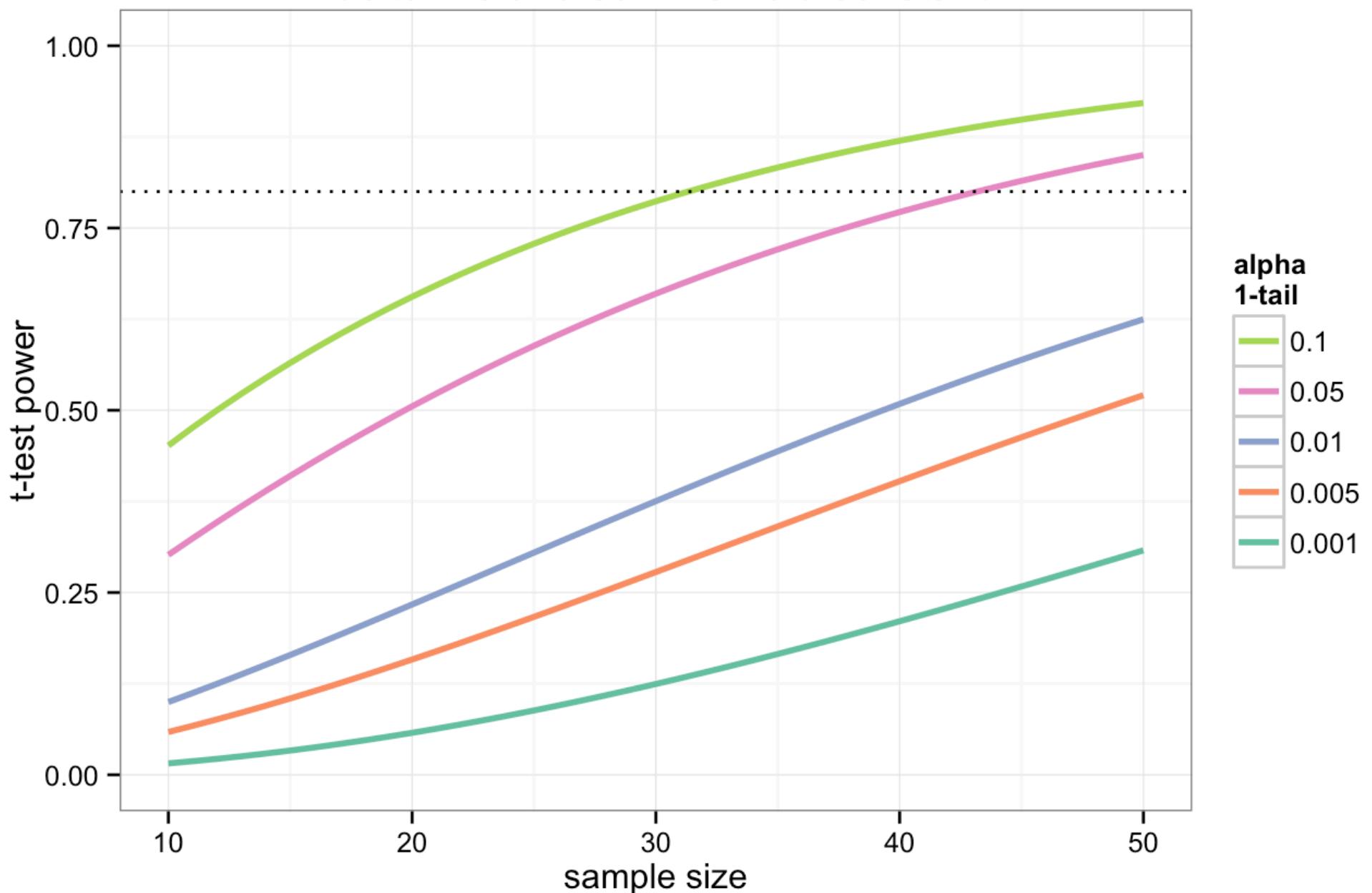
Power

For a given decision,

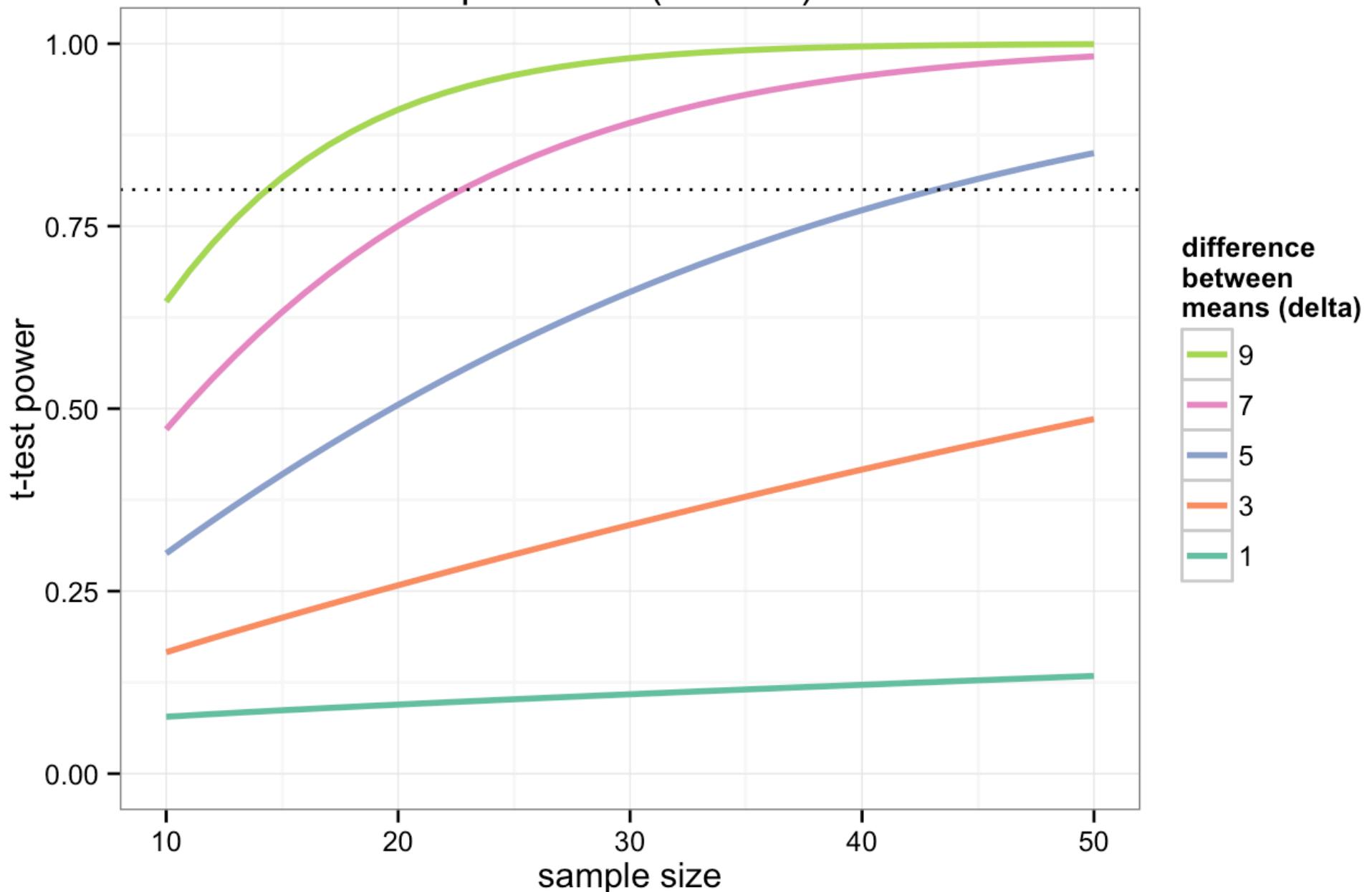
- $P(\text{false positive}) = \alpha = \text{Type I error}$
- $P(\text{true positive}) = 1 - \beta = \text{power}$

What will happen to $1 - \beta$ (power) if I make α smaller?

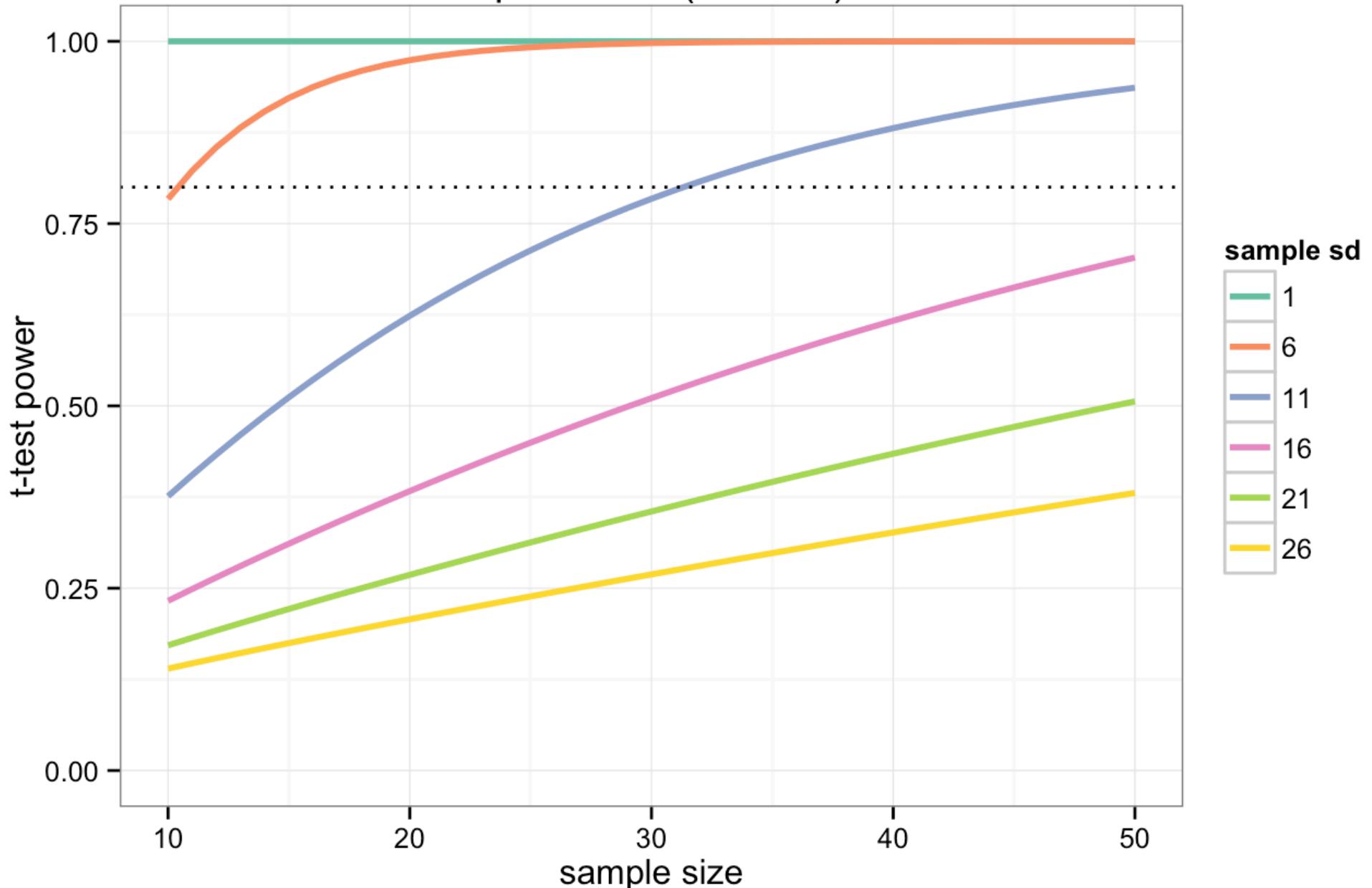
Power increases as n and alpha (1-tailed) increase
delta = 5 and sd = 13 held constant



Power increases as n and delta increase
sd = 13 and alpha = .05 (1-tailed) held constant



Power increases as n increases and sample sd decreases
delta = 5 and alpha = .05 (1-tailed) held constant



How large would our “n” have to be?

To detect:

- $\Delta = 5$
- $1 - \beta = .80$
- $\alpha = .05$



With s.d. = 13

Good for: a priori sample size determination



Sample size determination

```
power.t.test(n = [red box], delta = [red box], sd = [red box], sig.level = [red box], power  
= [red box], type = [red box] alternative = [red box])
```

One-tailed test



Sample size determination

```
power.t.test(n = NULL, delta = 5, sd = 13, sig.level = .05, power = .80, type = "one.sample", alternative = c("one.sided"))
```

One-sample t test power calculation

```
n = [REDACTED]
delta = 5
sd = 13
sig.level = 0.05
power = 0.8
alternative = one.sided
```



One-tailed test



Sample size determination

```
power.t.test(n = NULL, delta = 5, sd = 13, sig.level = .05, power = .80, type = "one.sample", alternative = c("one.sided"))
```

One-sample t test power calculation

```
n = 43.17957  
delta = 5  
sd = 15  
sig.level = 0.05  
power = 0.8  
alternative = one.sided
```



One-tailed test



How small of an effect could we detect...

If we knew we could get:

- $n = 100$ high school girls who are aspiring astronauts

And we wanted:

- $1 - \beta = .80$
- $\alpha = .05$

With $s.d. = 13$

One-tailed test



Sample size determination

```
power.t.test(n = [REDACTED], delta = [REDACTED], sd = [REDACTED], sig.level = [REDACTED].5,  
power = [REDACTED], type = [REDACTED], alternative = [REDACTED])
```

One-tailed test



Effect size determination

```
power.t.test(n = 100, delta = NULL, sd = 13, sig.level = .05,  
power = .80, type = "one.sample", alternative = c("one.sided"))
```

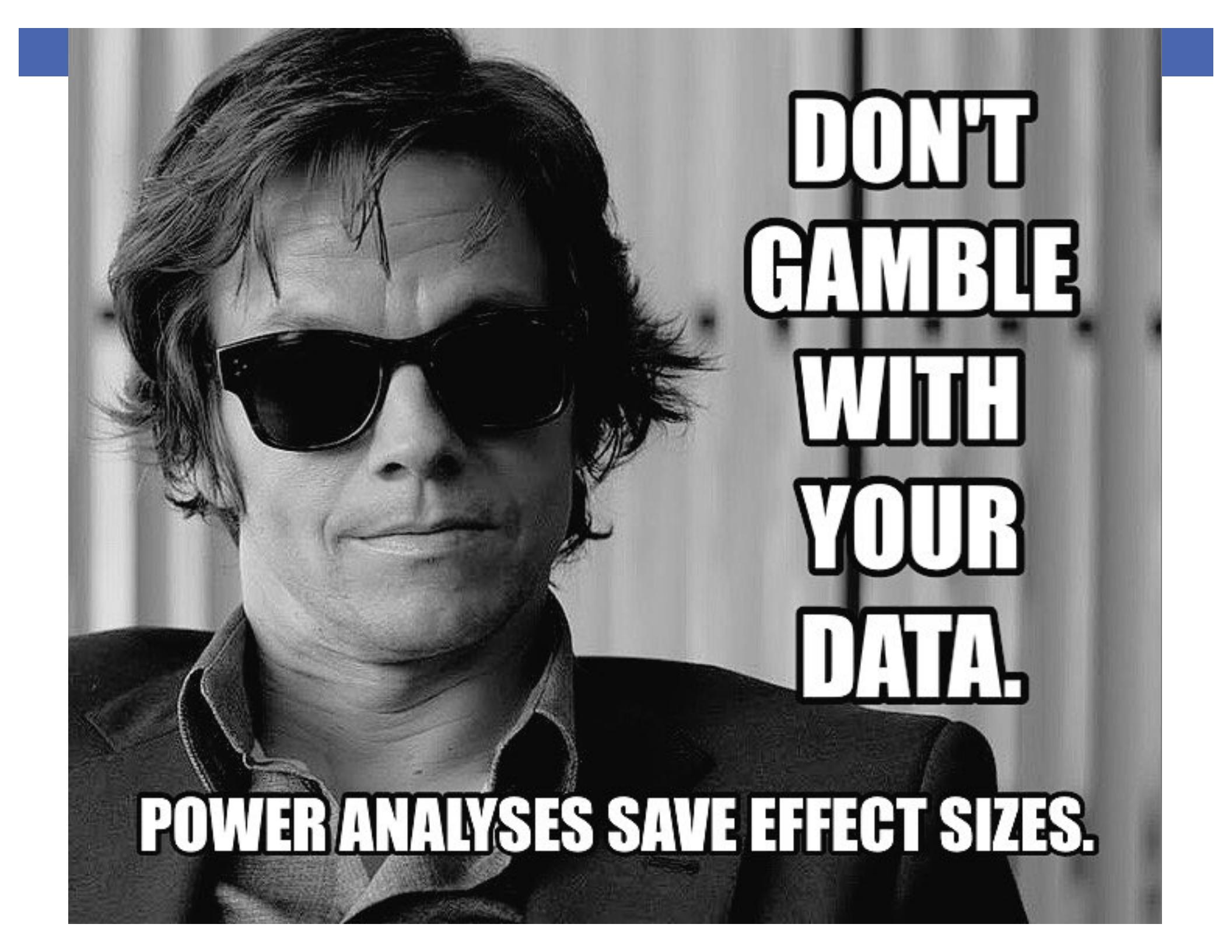
One-sample t test power calculation

```
n = 100  
delta = 3.254735  
sd = 13  
sig.level = 0.05  
power = 0.8  
alternative = one.sided
```



One-tailed test





**DON'T
GAMBLE
WITH
YOUR
DATA.**

POWER ANALYSES SAVE EFFECT SIZES.