

DATA 500

Lecture 10

Hypothesis Tests

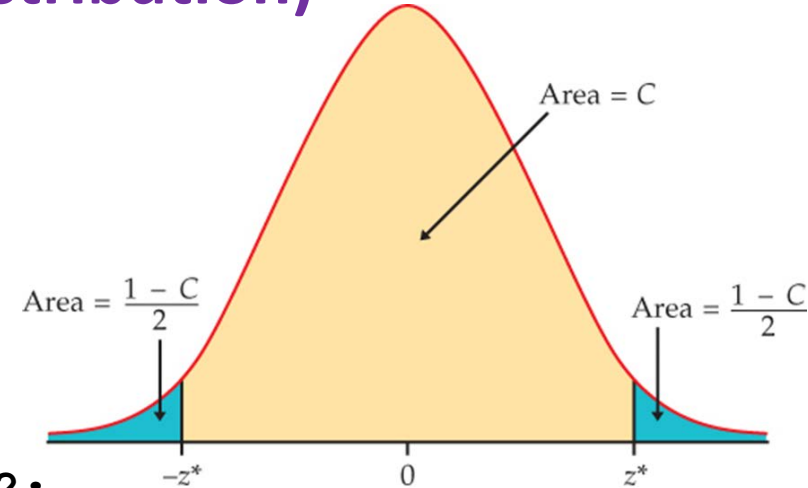
The t -distribution

The t -test

Getting confidence intervals using μ and σ qnorm (normal distribution)

$x = 100$, $SEM = 10$

How do we get a 80 % CI
on population mean?



Symmetry: 10% in each tail $(1 - CI)/2$;

```
CI <- 0.80;
```

```
# qnorm(1-CI)/2,mu, SEM) # lower bound;
```

```
qnorm(0.10,100,10) # 10the percentile;
```

```
[1] 87.18448
```

```
# qnorm(CI + (1-CI)/2,mu, se) # upper bound;
```

```
qnorm(0.90,100,10) # 90th percentile;
```

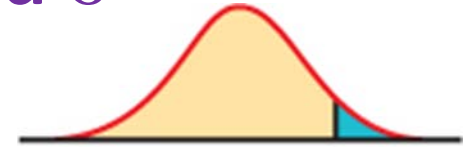
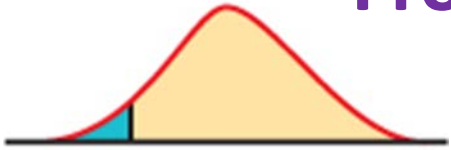
```
[1] 112.8155
```

```
112.8 - 87.2
```

```
[1] 25.6
```

the estimated population mean is 100 with 80% CI ± 12.8

From probabilities using μ and σ to conclusions



We can get probabilities for individuals

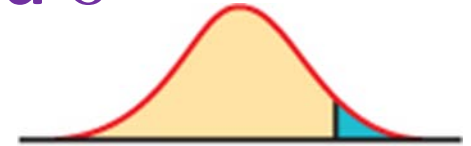
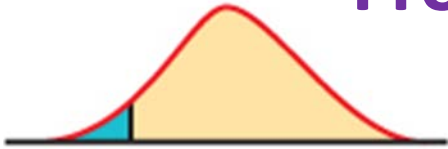
I know that *Buteo jamaicensis* female wingspan is $\sim N(168.75, 6.0)$.

I measured another female bird wingspan at 175 mm.

Could it come from the same species?



From probabilities using μ and σ to conclusions



We can get probabilities for individuals

I know that *Buteo jamaicensis* female wingspan is $\sim N(168.75, 6.0)$.

I measured another female bird wingspan at 175 mm.

Could it come from the same species?

(What is the probability that it comes from the same species?)

(we can still use an interval for a continuous probability)

```
pnorm(175,168.75,6) # value, mu, sigma
[1] 0.8512169
1-pnorm(175,168.75,6)
[1] 0.1487831
```

What do you conclude? Why?

I found another female and the wingspan was 154 mm.

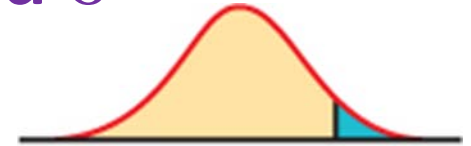
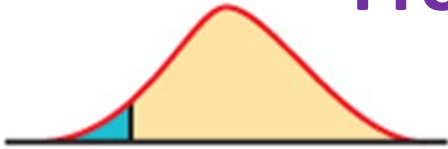
Could it come from the same species?

```
pnorm(154,168.75,6)
[1] 0.006979177
```

What do you conclude? Why?



From probabilities using μ and σ to conclusions



We can get probabilities for samples

I know that *Buteo jamaicensis* female wingspan is $\sim N(168.75, 6.0)$.

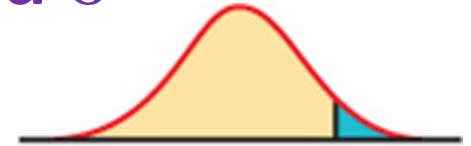
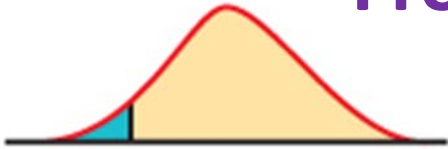
I measured 100 female bird wingspans. $\bar{x} = 169.4$

Could the sample come from the same species?

(What is the probability that the sample comes from the same species?)



From probabilities using μ and σ to conclusions



We can get probabilities for samples

I know that *Buteo jamaicensis* female wingspan is $\sim N(168.75, 6.0)$.

I measured 100 female bird wingspans. $\bar{x} = 169.4$

Could the sample come from the same species?

(What is the probability that the sample comes from the same species?)

(we can still use an interval for a continuous probability)

```
1-pnorm(169.4,168.75,6/sqrt(100)) # value, mu, sigma/sqrt(n)
[1] 0.1393302
```

What do you conclude? Why?

In another sample of 100, $\bar{x} = 170.4$

Could it come from the same species?

```
1-pnorm(170.4,168.75,0.6)
[1] 0.002979763
```

What do you conclude? Why?



Hypothesis testing helps us make conclusions - but it does NOT use deductive logic

Socrates is a man.

Premise/Fact

All men are mortal.

Premise/Fact

Socrates is mortal.

Conclusion

**Deductive logic involves certainties:
If the premises are true, the conclusion **MUST** be true**

Hypothesis testing uses **INDUCTIVE** Logic

- involves a **probability** of being correct

To test an idea, an assertion, **a hypothesis** that a coin is fair (random):

The **NULL*** hypothesis is that the coin is fair;

*** NO EFFECT, NO ASSOCIATION, EVERYTHING IS EQUAL, RANDOM**

**Should we try to find evidence to support the hypothesis?
(verification principle, Logical Positivists)**

No! We try to disprove / falsify the hypothesis!

(Karl Popper)

Probabilities less than 0.05 or 0.01 are usually rejected.

Probability quantifies our conclusions and certainty

Hypothesis Testing

Like a criminal trial?

Innocence is assumed (Null hypothesis)

Is there enough evidence, **beyond a reasonable doubt**, that the accused is guilty?

**Prosecutors try to make innocence nearly impossible
- with evidence of guilt**

Like a civil trial?

Is there are **preponderance of evidence** of something? (more likely than not)

Lawyers try to show that it is more likely than not that harm came to a plaintiff and a particular entity is responsible

Hypothesis Testing: Statistical Tests and Probability

- The NULL hypothesis (H_0) is that everything is the same; no relationship, no difference, any pattern is random
- The data at hand are used to calculate a test statistic, given test assumptions (requirements)
- The test statistic is used to calculate a probability under H_0
(In other words, **a significance test** is performed based on a test statistic)
- The probability (the p value) is of obtaining the value for a test statistic, or a more extreme value, under H_0

Hypothesis Testing: Statistical Tests and Probability

Put another way, p is the probability that **the data at hand** would be seen **IF** the NULL hypothesis were correct/ "true" /valid /plausible.

If $p > 0.05$, you **ACCEPT** H_0

If $p < 0.05$ (1/20), (or $p < 0.01$) then you **REJECT** H_0
(it is "FALSE")

- otherwise, you accept / fail to reject the hypothesis

If you reject H_0 then the test was (statistically) significant

Hypothesis Testing:

Things to remember

Rejecting is not inherently better than accepting H_0

Sometimes you would rather have one or the other

Example: tests for significant differences between groups

- good if you are looking for differences
- bad if you want to pool groups (for bigger n)

Accepting does not mean H_0 is "true"

- you may have small samples
- you may violate test assumptions
- your data may have outliers you missed
- you may be using a bad test

Type I and Type II Errors

(b) Test of an hypothesis H_0		
Conclusion of statistical test	True state of affairs	
	H_0 is true	H_0 is false
H_0 is not rejected	Correct conclusion	(β) Type II error
H_0 is rejected	(α) Type I error	Correct conclusion

Type I: a false negative; H_0 is true, but was rejected.

Type II: a false positive; H_0 is not true, but was accepted.

Type I Error (α)

Type I: Null H_0 is true, but is rejected. (false rejection)

Type I error (α , alpha) depends on:

- Our chosen significance level (α) , $\alpha = 1 - p$
(usually $\alpha = 0.01$ or 0.05)

So you can:

Choose significance only if $p < 0.000000000000000001$, **reject everything**, and have type I error all the time;

Choose $p < 0.90$, **accept nearly everything**, and have almost no type I error (but with high type II error)

At a trial: an innocent man is found guilty

Type I Error

Type I: Null H_0 is true, but is rejected. (false rejection)

Type I error (α) depends on:

- Our chosen significance level (alpha α , $1-p$)

Type I errors will **DECREASE** when:

You choose a higher p value

Type I errors will **INCREASE** when:

You choose a lower p value

p values

(Verzani, page 296)

p -value range	significance stars	common description
$[0, 0.001]$	***	extremely significant
$(0.001, 0.01]$	**	highly significant
$(0.01, 0.05]$	*	statistically significant
$(0.05, 0.1]$.	NOT significant
$(0.1, 1]$		not significant

Table 9.1: Level of significance for range of p -values.

p values are still being debated, especially whether 0.05 or 0.01 is a better cutoff

I prefer 0.01 for most tests because they are more likely to be replicated ("true")

Many strange things are found when p is near 0.05

What should a p value be for rejecting innocence in a trial? (costs)

Type II Error (β)

Type II: Null H_0 is false, but is accepted. (false acceptance)

Type II error (β) depends on:

- True value of parameter (what you are estimating)
- Significance level (α , type 1 error)
- Whether one- or two-tailed test
- Sample standard deviation (variation)
- Sample size

So you can:

Collect more data

Choose your tails carefully

At a trial: a guilty man is found innocent

Type II Error (β)

Type II errors will decrease when

- sample size is larger
- a one-tailed test is chosen over a two-tailed test
- differences between sample means increase
- Type I error α (alpha) is increased (p is decreased)
- population standard deviation is smaller

Type II errors will increase when

- opposites of above

Population Mean: Type II errors depend on

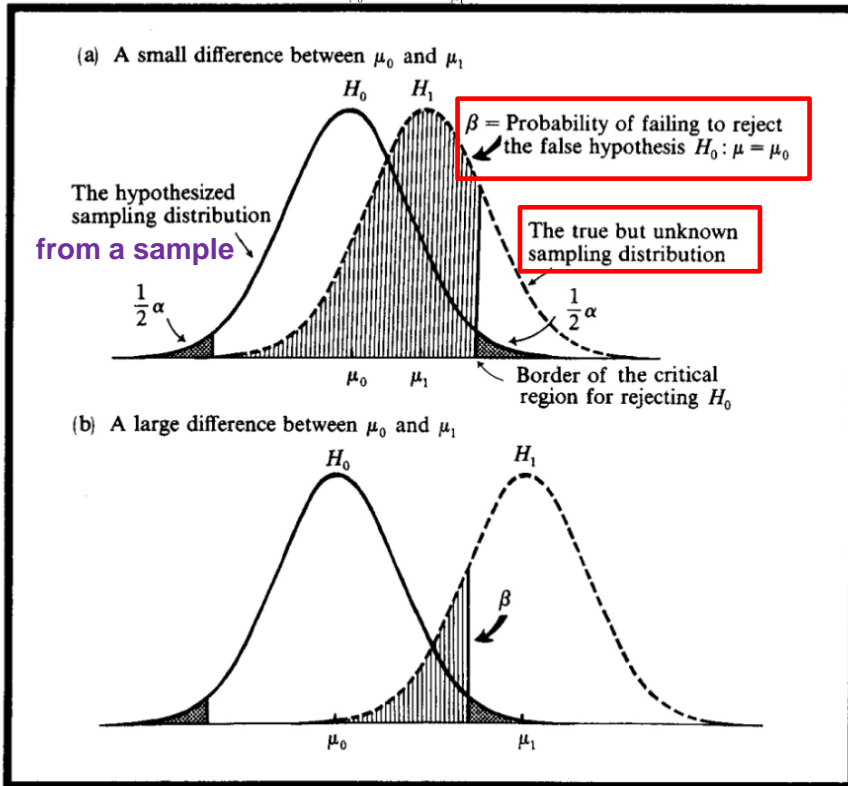
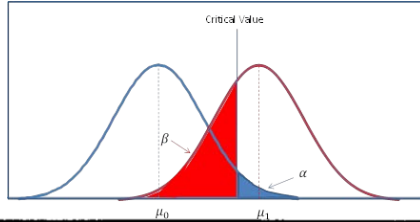


Figure 4 How the probability of a type II error, β , depends on the deviation of the true parameter value from the hypothesized value.

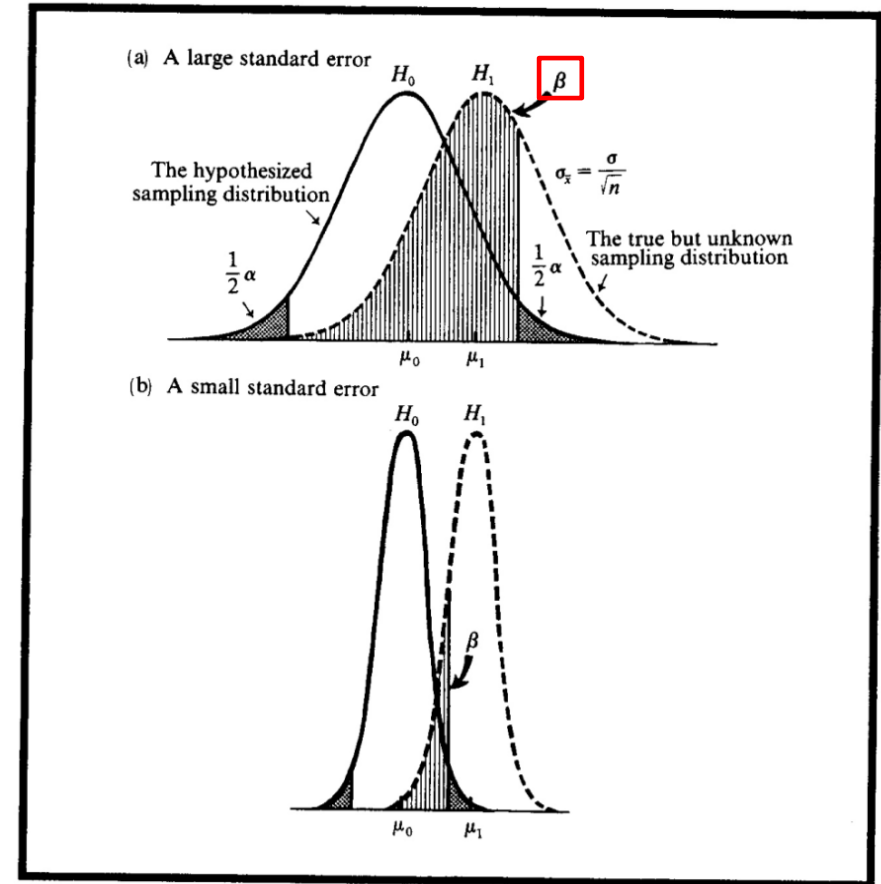


Figure 5 How the probability of a type II error, β , depends on the population standard deviation and the sample size.

Difference between sample estimate and true population parameters

Sample size (standard error) and population variation (standard deviation)

Type II errors depend on

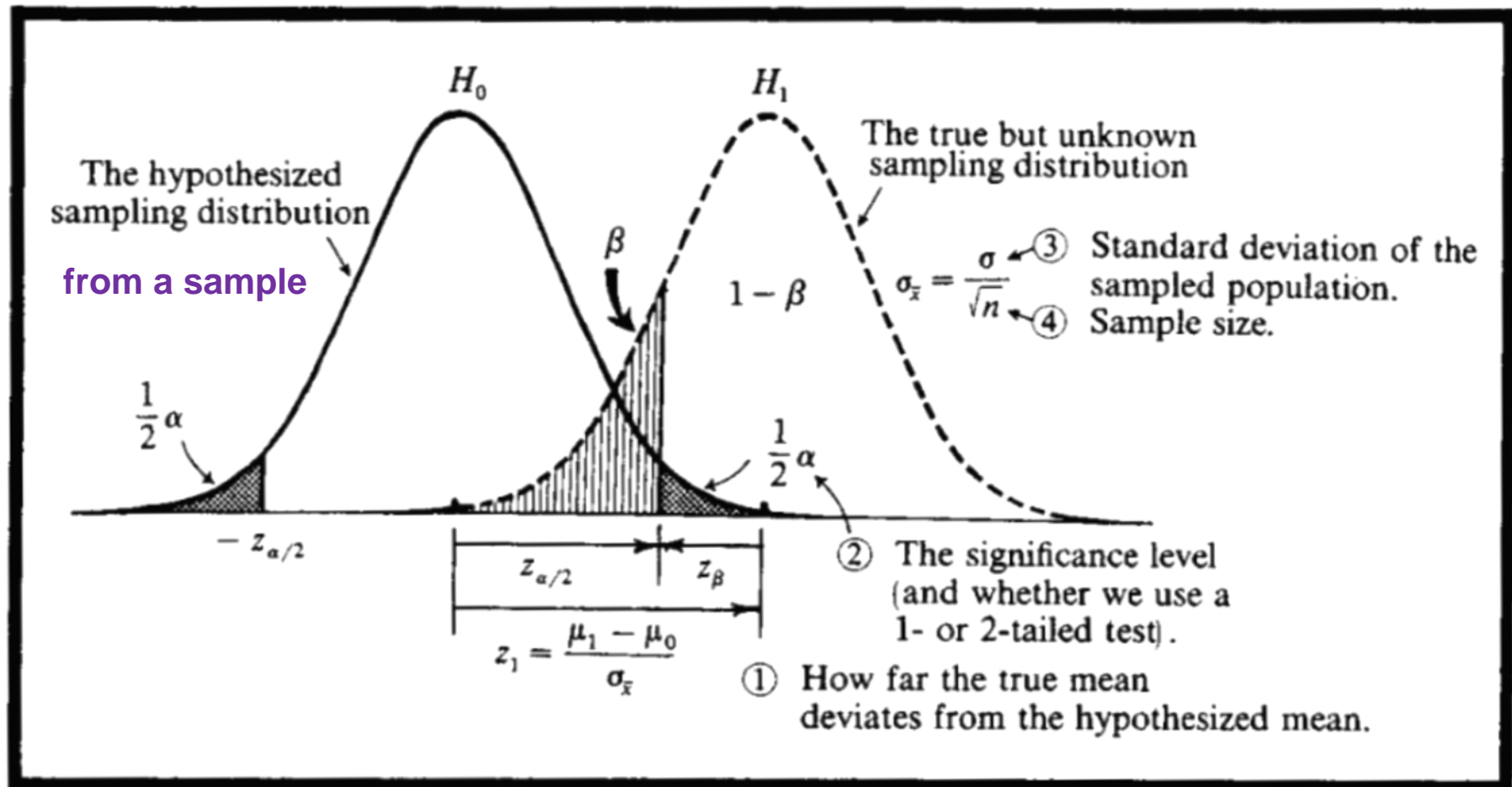


Figure 9 Summary of factors affecting the probability of a type II error, β (and the power of the test, $1 - \beta$).

a lot!

POWER

- the ability to detect TRUE differences between samples

Power is related to the Type II error (β):

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

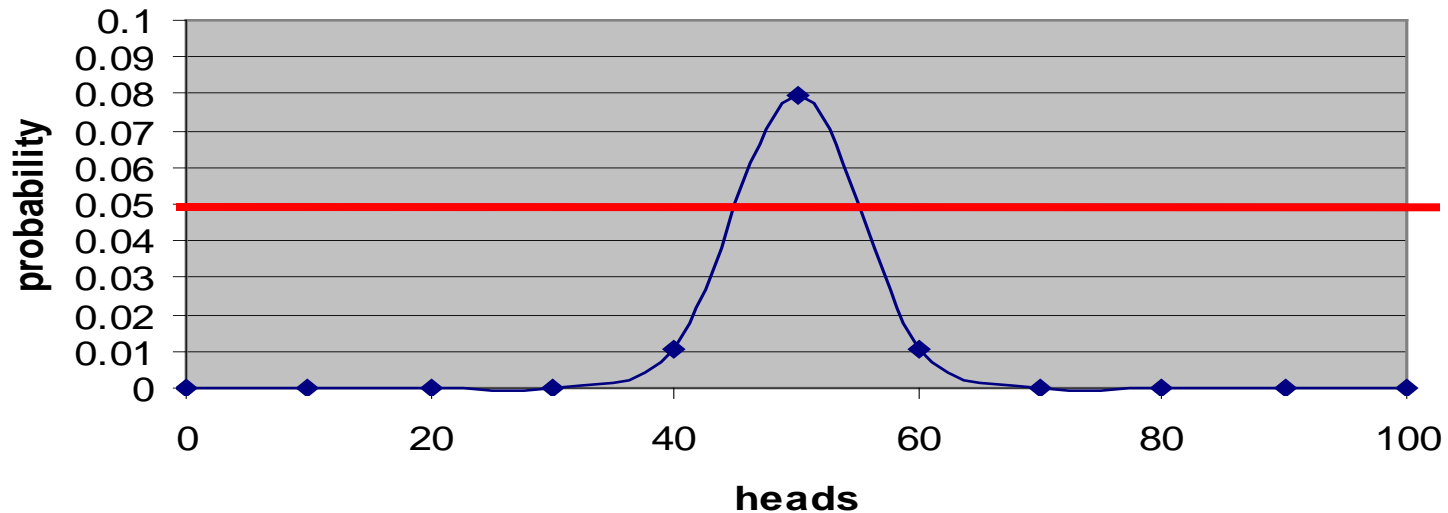
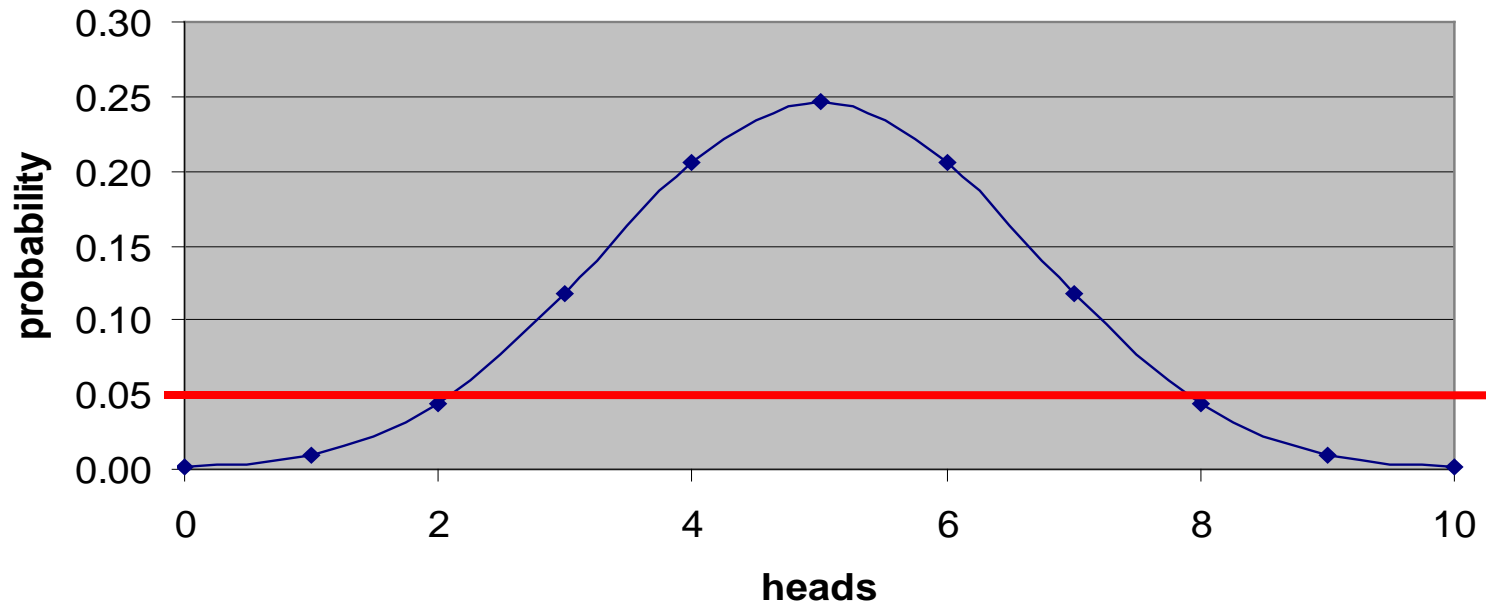
Higher power: better ability to detect TRUE differences between sample estimates and true underlying population parameters.

A test with higher power is more sensitive

TRUE differences are expressed in standard deviation units.

The TRUE difference is often called the **Effect Size**

**A larger sample size will have lower Type II error,
therefore greater POWER to detect true differences**



Power and sample size

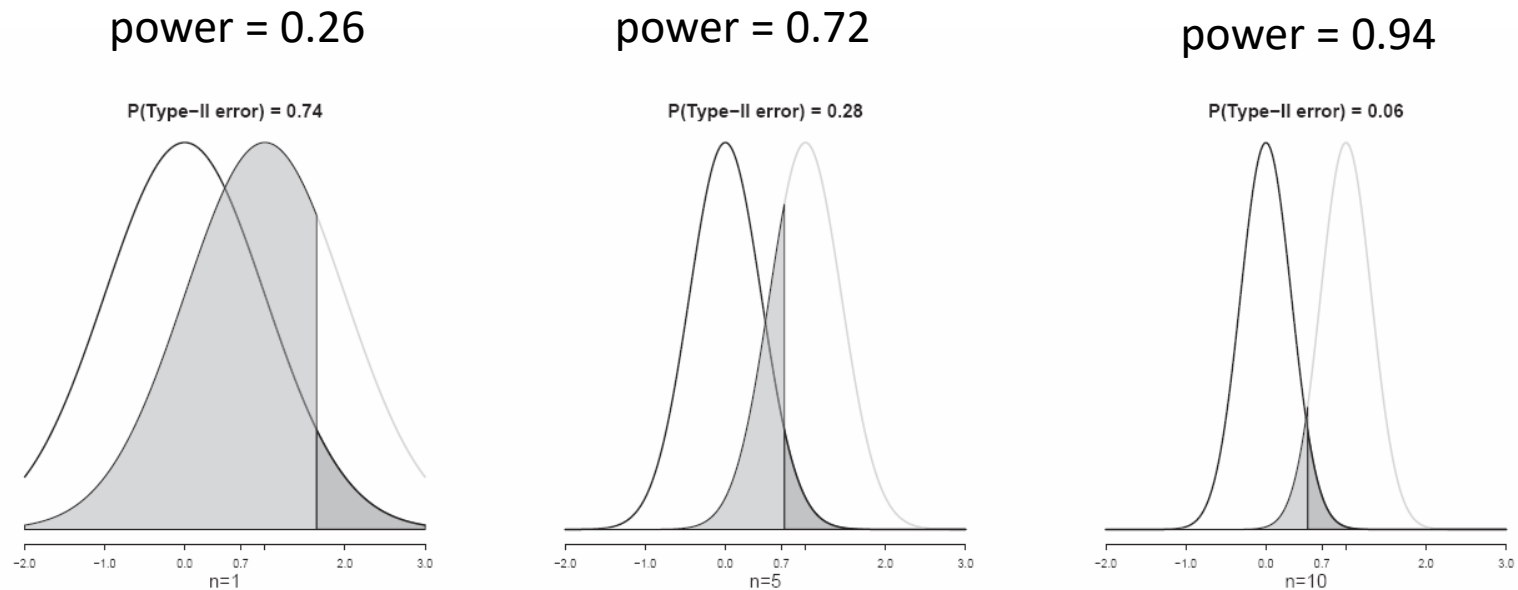
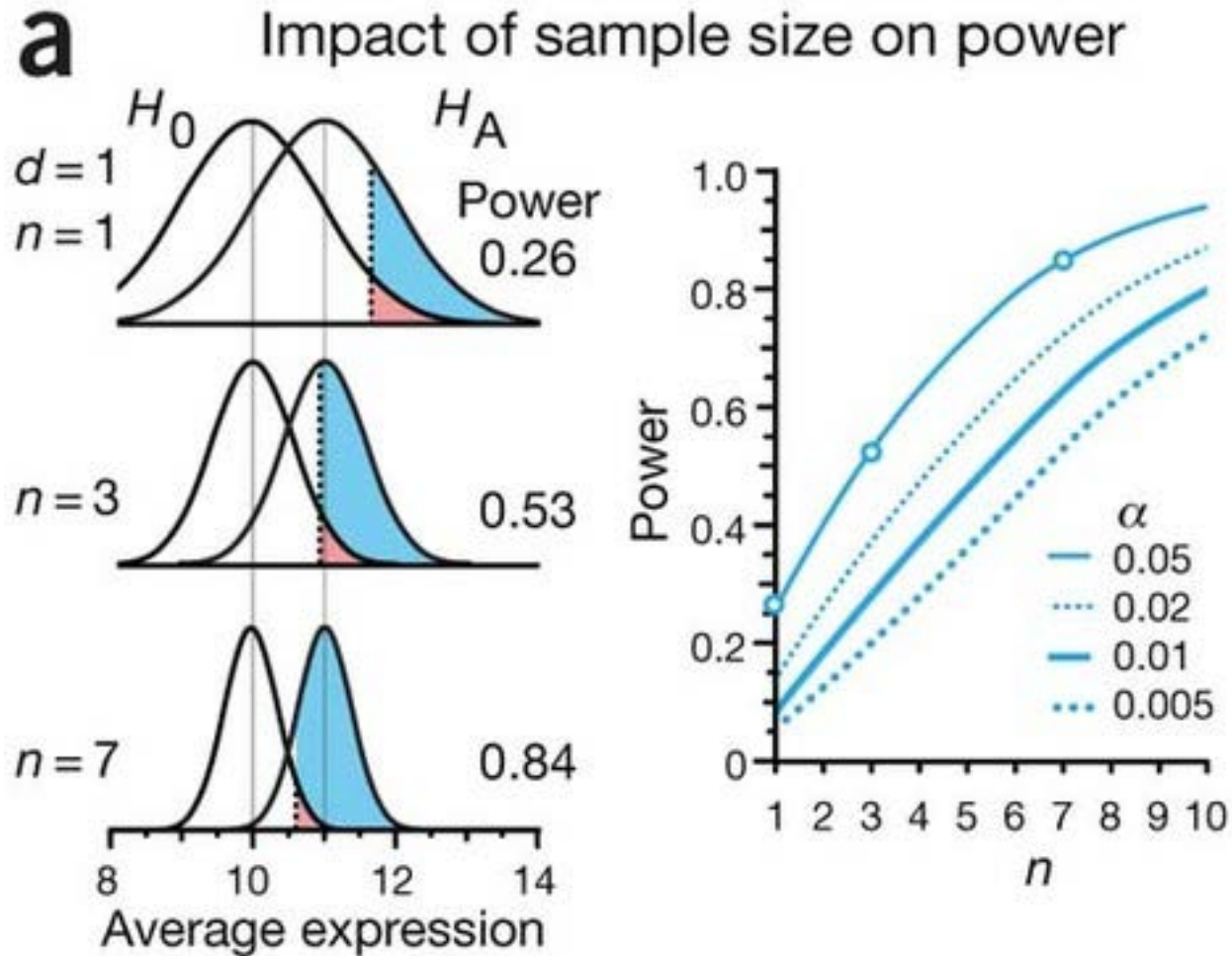


Figure 9.3: The probability of a Type-II error depends on the sample size. In the figure, this probability is represented by the area to the left of z^* for the probability distribution centered at 1. For $n = 1$, this error is likely, for $n = 10$ it is unlikely.

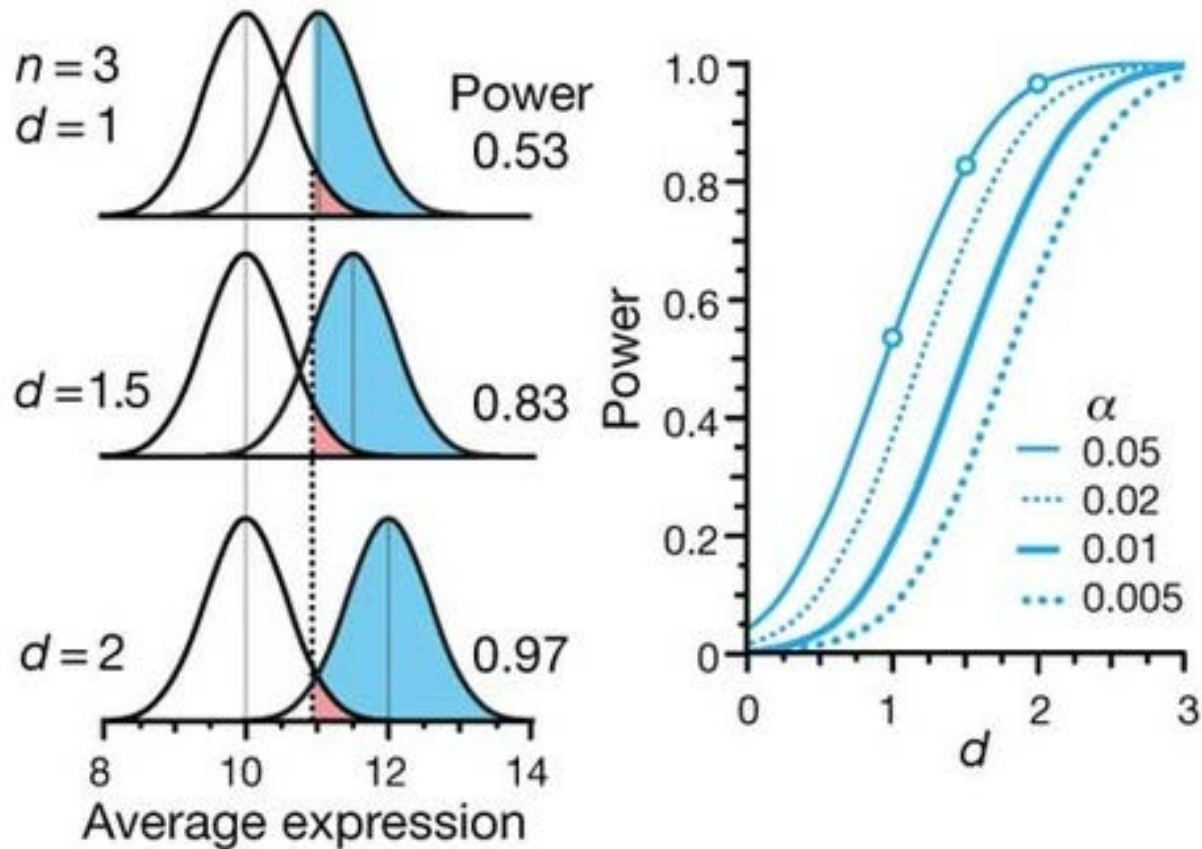
Power and Sample size: SEM



Power and Effect size: Sample Differences

b

Impact of effect size on power

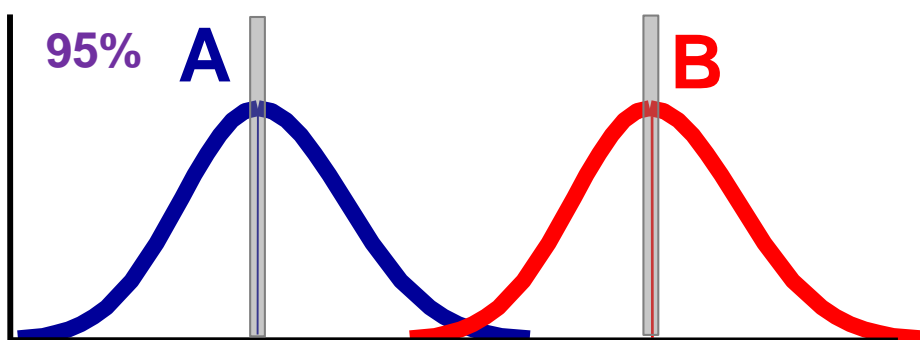
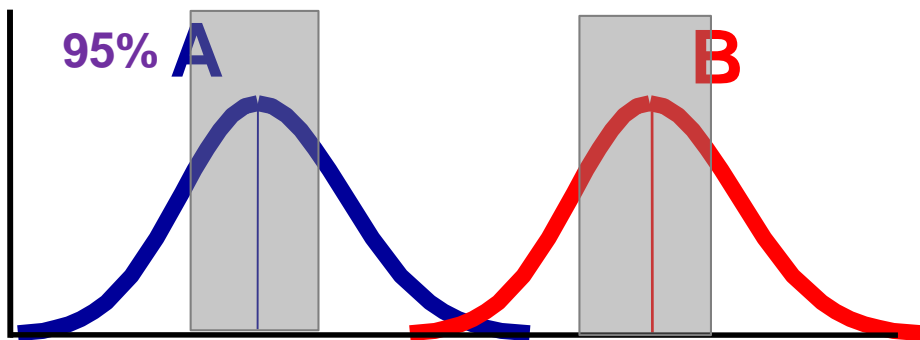


Larger differences are easier to detect

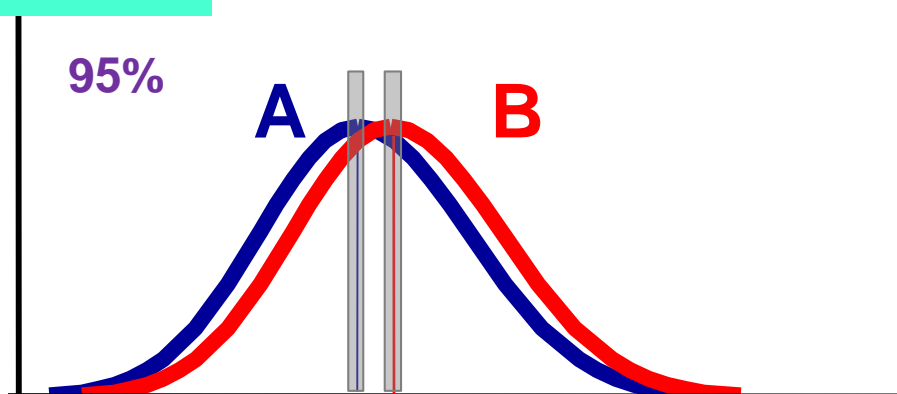
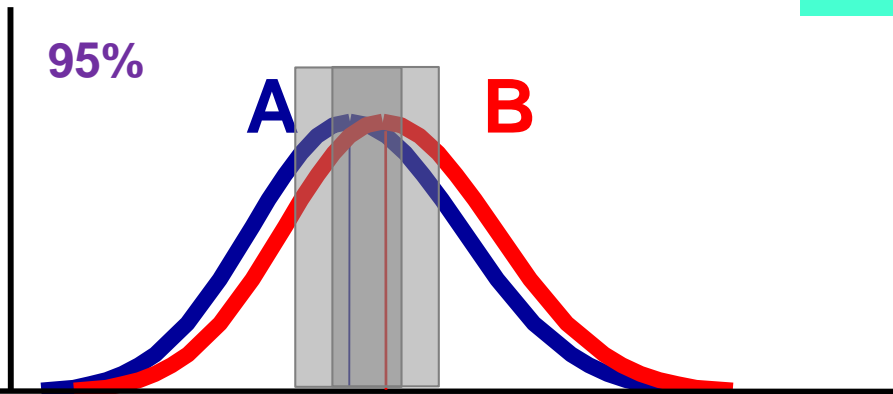
CIs for SEM,
small samples

LARGE differences
between groups

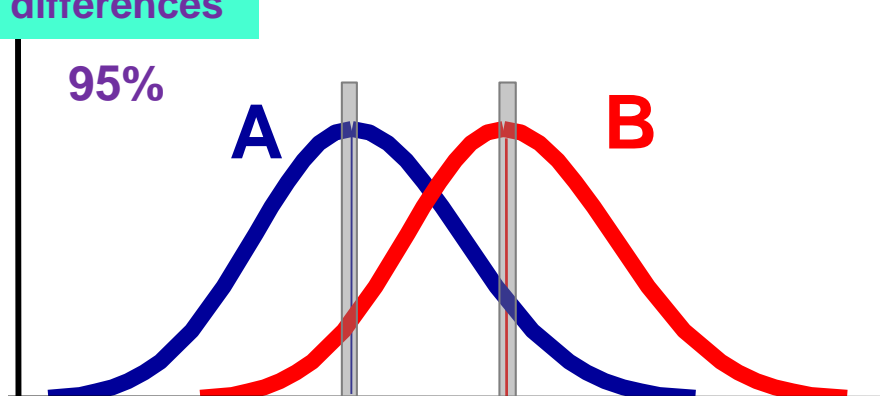
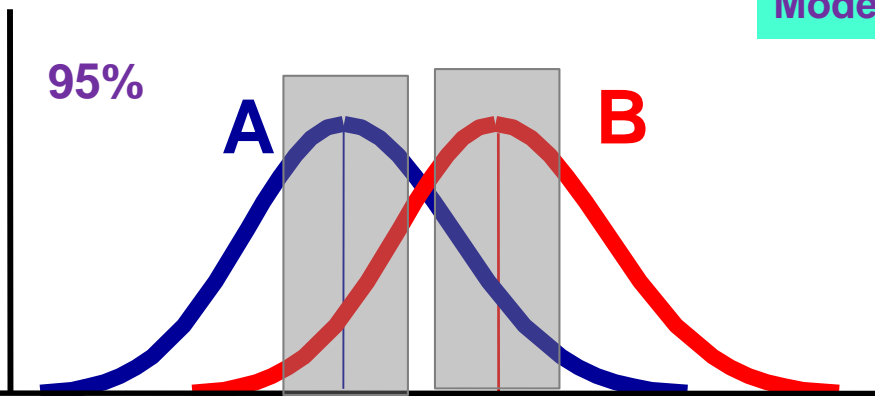
CIs for SEM,
LARGE samples



small differences



Moderate differences

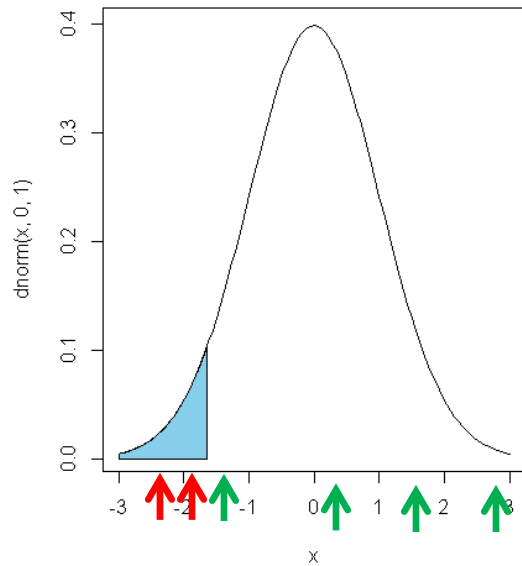


Creating a hypothesis test

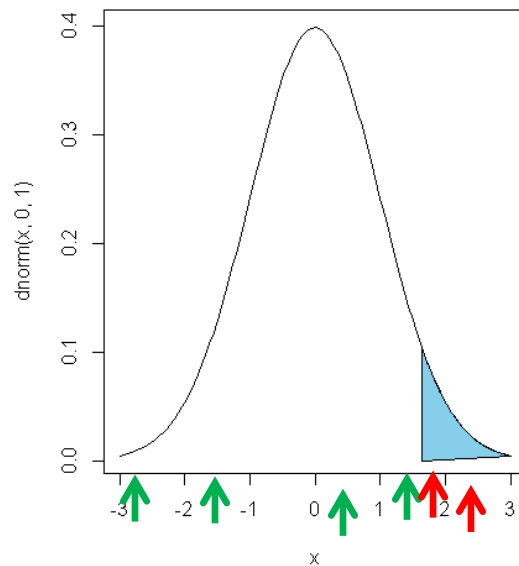
1. Specify some model for the underlying data.
 z or t distribution
2. Identify H_0 and keep in mind H_a - the alternate hypothesis.
3. Specify a **test statistic** that helps you decide between the two hypotheses, based on # 1.
4. Collect the data, then find the observed value of the test statistic.
5. **Calculate the p -value under the null hypothesis.** The p -value will be the probability of observing the test statistic or a more extreme test statistic, given the data, and accepting H_0 .
6. The smaller the p -value, the stronger the evidence against the null hypothesis.

Creating a hypothesis test

One-tailed lower probability, alpha = 5%



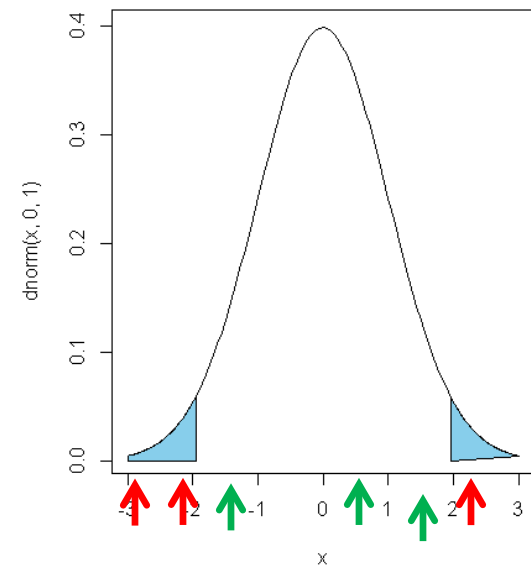
One-tailed upper probability, alpha = 5%



fail to reject H_0

reject H_0

Two-tailed probability, alpha = 5%



One-tailed, lower bound:

Are measurements from one sample **significantly lower** than another sample?

One-tailed, upper bound:

Are measurements from one sample **significantly greater** than another sample?

Two-tailed, upper and lower bounds:

Are measurements from one sample **significantly different** from another sample?

One- or two-tailed?

Question: How are two samples related to each other?

- based on a variable, Length (Len)

A. Are the samples different? (two-tailed test)

$$H_0 : S1Len = S2Len; \quad \bar{X}_1 = \bar{X}_2$$

The alternate hypothesis, H_A , is:

$$S1Len > S2Len \text{ OR } S2Len > S1Len \quad (\bar{X}_1 > \bar{X}_2 \text{ OR } \bar{X}_1 < \bar{X}_2)$$

or...

B. Does sample 1 have greater Len than sample 2? (one-tailed test)

$$H_0 : S1Len \leq S2Len; \quad \bar{X}_1 \leq \bar{X}_2$$

The alternate hypothesis, H_A , is:

$$S1Len > S2Len; \quad \bar{X}_1 > \bar{X}_2$$

The cost of errors

What if an innocent man is convicted?

What if a guilty man is set free?

What if a blood pressure drug has bad side effects but creates no benefits to the patient?

What if a cancer test is a false positive?

What if a cancer test is a false negative?

- all of these can require adjusting α

Example

GMAT scores worldwide (200,000):

$\mu = 525$, $\sigma = 100$

100 students go through special course (\$500), $\bar{x} = 541.4$

Is 541.4 significantly higher? (did they benefit from course?)

```
# get z value  
(541.4 - 525) / ( 100/sqrt(100) )  
1.64  
1-pnorm(1.64)  
[1] 0.05050258
```

Is 541.5 significantly higher?

```
# get z value (1.65)  
1-pnorm( (541.5-525)/(100/sqrt(100) ) )  
[1] 0.04947147
```

***** Report p value, not just "significant at $p < 0.05$ "**

Be ready to answer these next time

6.93 What do you know? A research report described two results that both achieved statistical significance at the 5% level. The P -value for the first is 0.048; for the second it is 0.0002. Do the P -values add any useful information beyond that conveyed by the statement that both results are statistically significant? Write a short paragraph explaining your views on this question.

6.100 Statistical versus practical significance. A study with 7500 subjects reported a result that was statistically significant at the 5% level. Explain why this result might not be particularly important.

6.101 More on statistical versus practical significance. A study with 14 subjects reported a result that failed to achieve statistical significance at the 5% level. The P -value was 0.051. Write a short summary of how you would interpret these findings.

Be ready to answer these next time

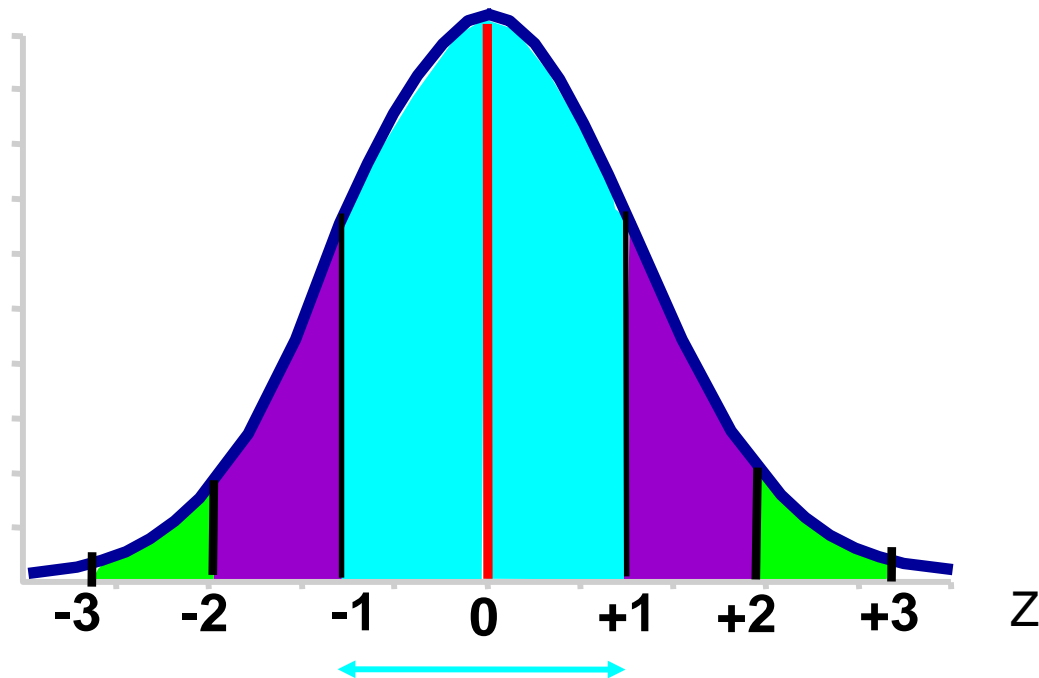
6.98 Do you agree? State whether or not you agree with each of the following statements and provide a short summary of the reasons for your answers.

- (a) If the P -value is larger than 0.05, the null hypothesis is true.
- (b) Practical significance is not the same as statistical significance.
- (c) We can perform a statistical analysis using any set of data.
- (d) If you find an interesting pattern in a set of data, it is appropriate to then use a significance test to determine its significance.
- (e) It's always better to use a significance level of $\alpha = 0.05$ than to use $\alpha = 0.01$ because it is easier to find statistical significance.

The Standard Normal Distribution (z)

Sampling distribution of x is centered on μ

We know σ



$\pm 1 \sigma$ = interval from $z = -1$ to $z = 1$ = 68 %

```
# probability/proportion/area between -1 and 1 (Z)
```

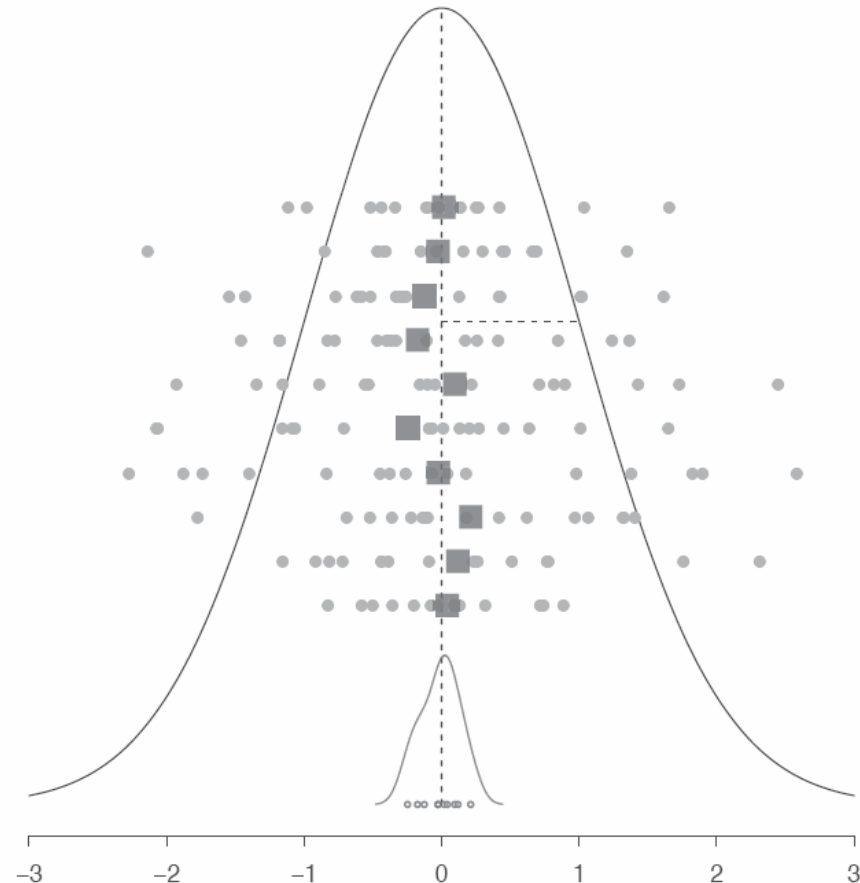
```
pnorm(1)-pnorm(-1)
```

```
[1] 0.6826895
```

```
diff(pnorm(c(-2,2))) # between -2 and 2 using diff function
```

```
[1] 0.9544997
```

Repeated samples and the SEM



\bar{x} is an unbiased estimator of μ

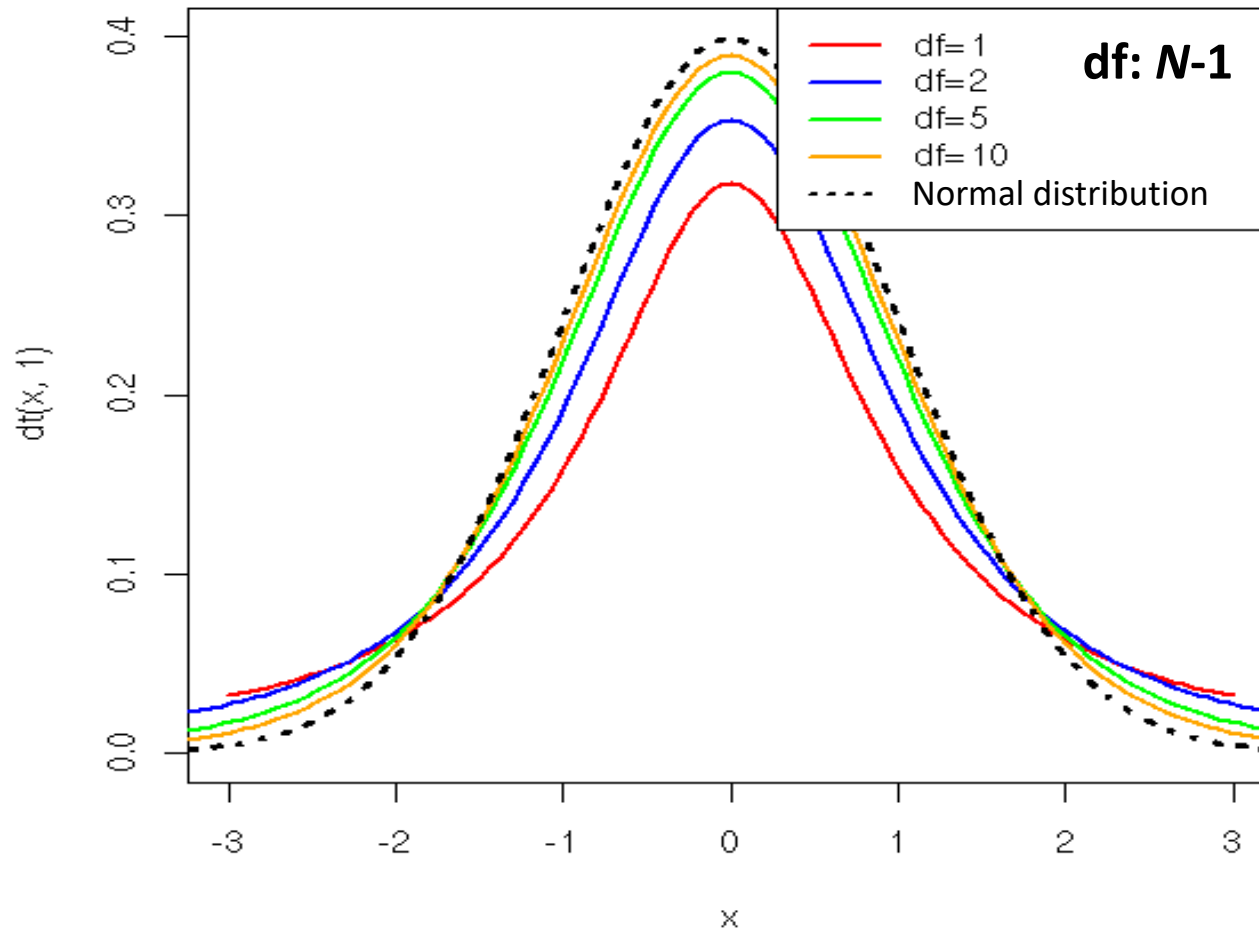
We rarely know σ

s is an unbiased estimator of σ

The t distribution

adjusts probabilities for sample size

t distribution vs. normal (Z distribution)



Small samples do not represent enough variation in the population

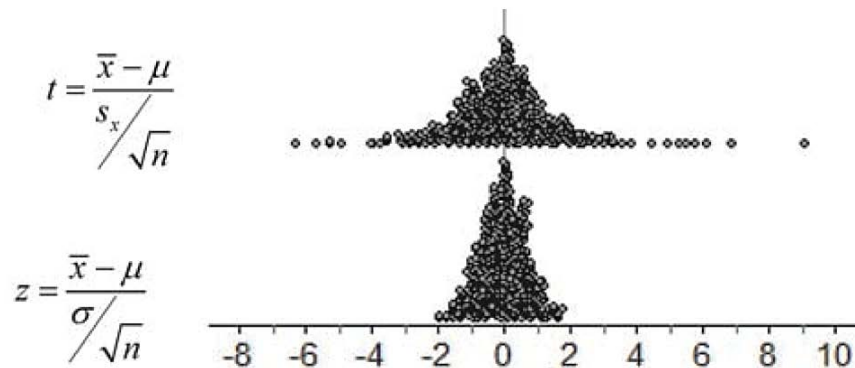
The t Distribution



When we standardize based on the sample standard deviation s_x , our statistic has a new distribution called a **t distribution**.

The t distribution has a shape similar to that of the standard Normal curve in that *it is symmetric with a single peak at 0*.

However, it differs from the Normal curve in that it has *more area in the tails*.



Like any standardized statistic, t tells us how far \bar{x} is from its mean μ in standard deviation units.

However, there is a different t distribution for each sample size, specified by its **degrees of freedom (df)**.

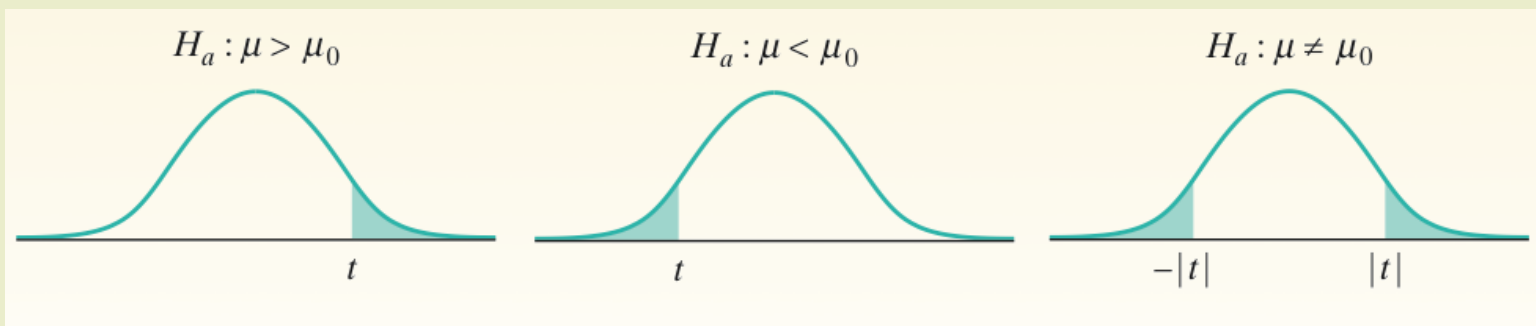
The One-Sample t Test

One-Sample t Test

Choose an SRS of size n from a large population that contains an unknown mean μ . To test the hypothesis $H_0 : \mu = \mu_0$, compute the one-sample t statistic:

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

Find the P -value by calculating the probability (at degrees of freedom = $n - 1$) of getting a t statistic this large or larger *in the direction specified by the alternative hypothesis* H_a .

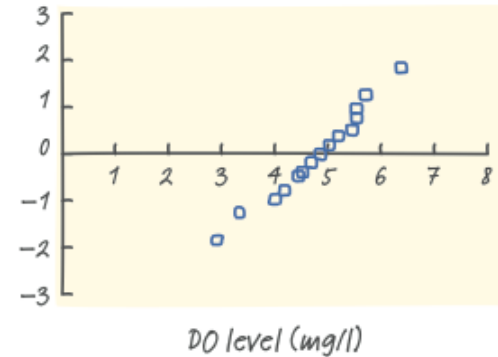
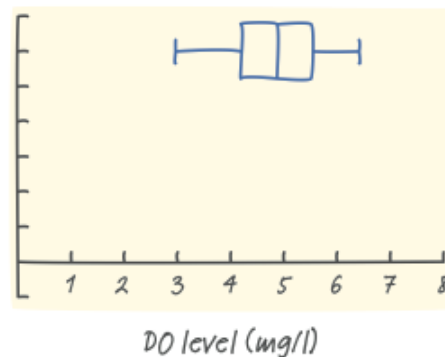
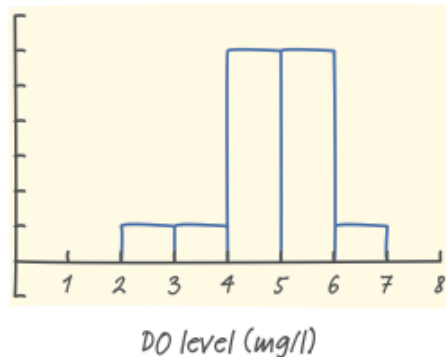


These P -values are exact if the population distribution is Normal and are approximately correct for large n in other cases.

Example

If conditions are met, we should do a one-sample t test for μ .

- ✓ **Random:** The researcher measured the DO level at 15 randomly chosen locations.
- ✓ **Normal:** We don't know whether the population distribution of DO levels at all points along the stream is Normal. With such a small sample size ($n = 15$), we need to look at the data to see if it's safe to use t procedures.



The histogram looks roughly symmetric; the boxplot shows no outliers; and the Normal probability plot is fairly linear. With no outliers or strong skewness, the t procedures should be pretty accurate even if the population distribution isn't exactly Normal.

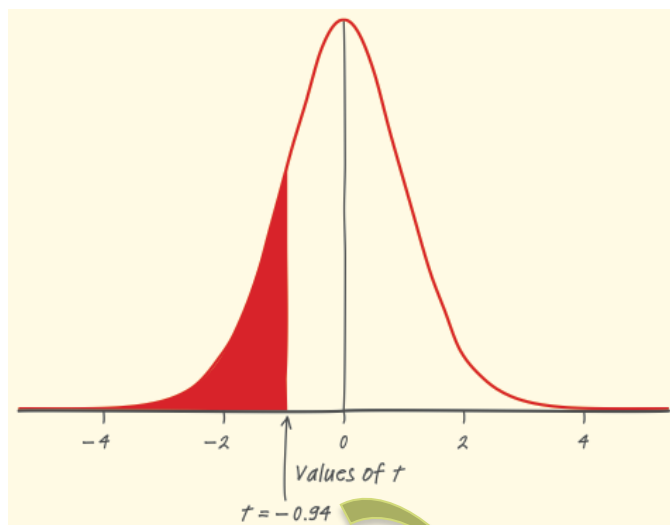
Example (it may help)

The sample mean and standard deviation are $\bar{x} = 4.771$ and $s_x = 0.9396$

Test statistic $t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{4.771 - 5}{\frac{0.9396}{\sqrt{15}}} = -0.94$

P-value The *P*-value is the area to the left of $t = -0.94$ under the *t* distribution curve with $df = 15 - 1 = 14$.

The *P*-value is between 0.15 and 0.20. Since this is greater than our $\alpha = 0.05$ significance level, we fail to reject H_0 . We don't have enough evidence to conclude that the mean DO level in the stream is less than 5 mg/l.



Upper-tail probability *p*

<i>df</i>	.25	.20	.15
13	.694	.870	1.079
14	.692	.868	1.076
15	.691	.866	1.074
	50%	60%	70%

Confidence level *C*

Since we decided not to reject H_0 , we could have made a Type II error (failing to reject H_0 when H_0 is false). If we did, then the mean dissolved oxygen level μ in the stream is actually less than 5 mg/l, but we didn't detect that with our significance test.

Robustness of t procedures



A confidence interval or significance test is called **robust** if the confidence level or P -value does not change very much when the conditions for use of the procedure are violated.

Using the t Procedures

Except in the case of small samples, the condition that the data are an SRS from the population of interest is more important than the condition that the population distribution is Normal.

- *Sample size at least 15*: The t procedures can be used except in the presence of outliers or strong skewness.
- *Sample size less than 15*: Use t procedures if the data appear close to Normal. If the data are clearly skewed or if outliers are present, do not use t .
- *Large samples*: The t procedures can be used even for clearly skewed distributions when the sample is large, roughly $n \geq 40$.

Student's t distribution and beer

- Published by in 1908 as "The Probable Error of a Mean" in *Biometrika*
- the author's name was given as "Student"
- actually William Gossett, who worked for Guinness
- Guinness had a non-disclosure policy

VOLUME VI

MARCH, 1908

No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

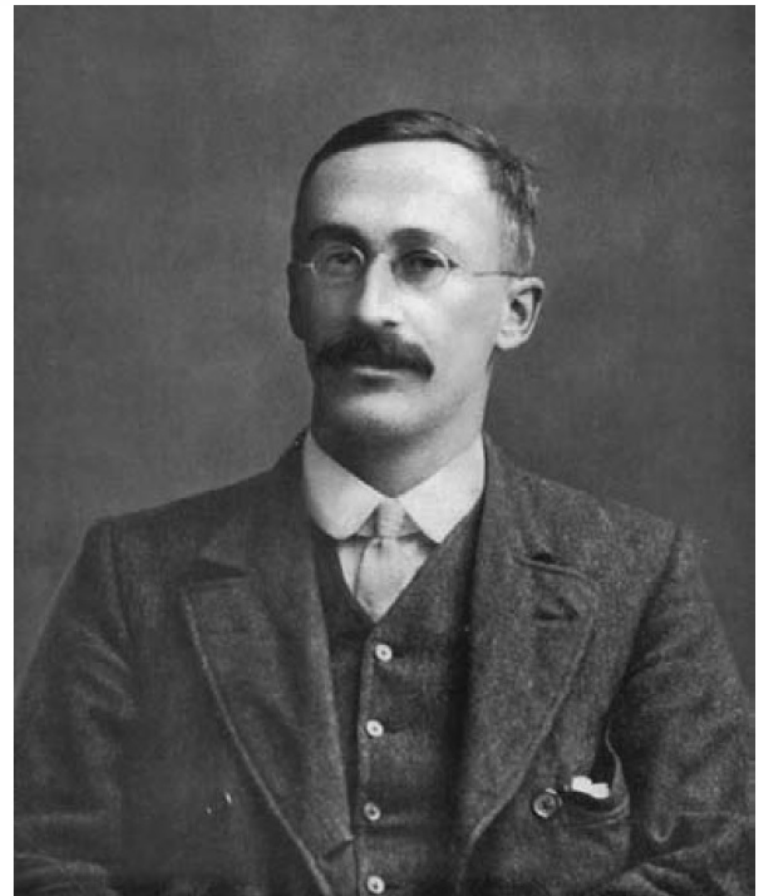
By STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information




In 1908 "Student," William Sealy Gosset (1876–1937), invented a statistical instrument that would change the life and social sciences. Now those

The t distribution

Gossett used simulation to demonstrate his theoretical argument

- repeated samples of $n = 4$ from 3,000 finger lengths
- adjusts the PDF (probability density function) for sample size
- we use it like z to get a probability and construct CIs

The basis for several statistical tests involving t

$$\text{SE}(\bar{x}) \text{ or SEM} = \frac{s}{\sqrt{n}}$$


p (individual value)

$$t\text{-score} = \frac{x_i - \bar{x}}{s}$$

p (sample vs. pop mean)

$$t^* = \frac{\bar{X} - \mu_0}{\text{SEM}}$$

p (two samples)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Let's resample Finger length means

```
MacD <- read.csv('http://math.mercyhurst.edu/~sousley/STAT_139/data/MacD.csv', header=T)
FingerL <- MacD$fingerL
```

```
# Use the replicate function to sample repeatedly
```

```
r500_4 <- replicate(500, # repeat 500 times
{
mean(sample(FingerL, 4) ) # samples of 4 ;
} )
```

```
r500_10 <- replicate(500, # repeat 500 times
{
mean(sample(FingerL, 10) ) # samples of 10;
} )
```

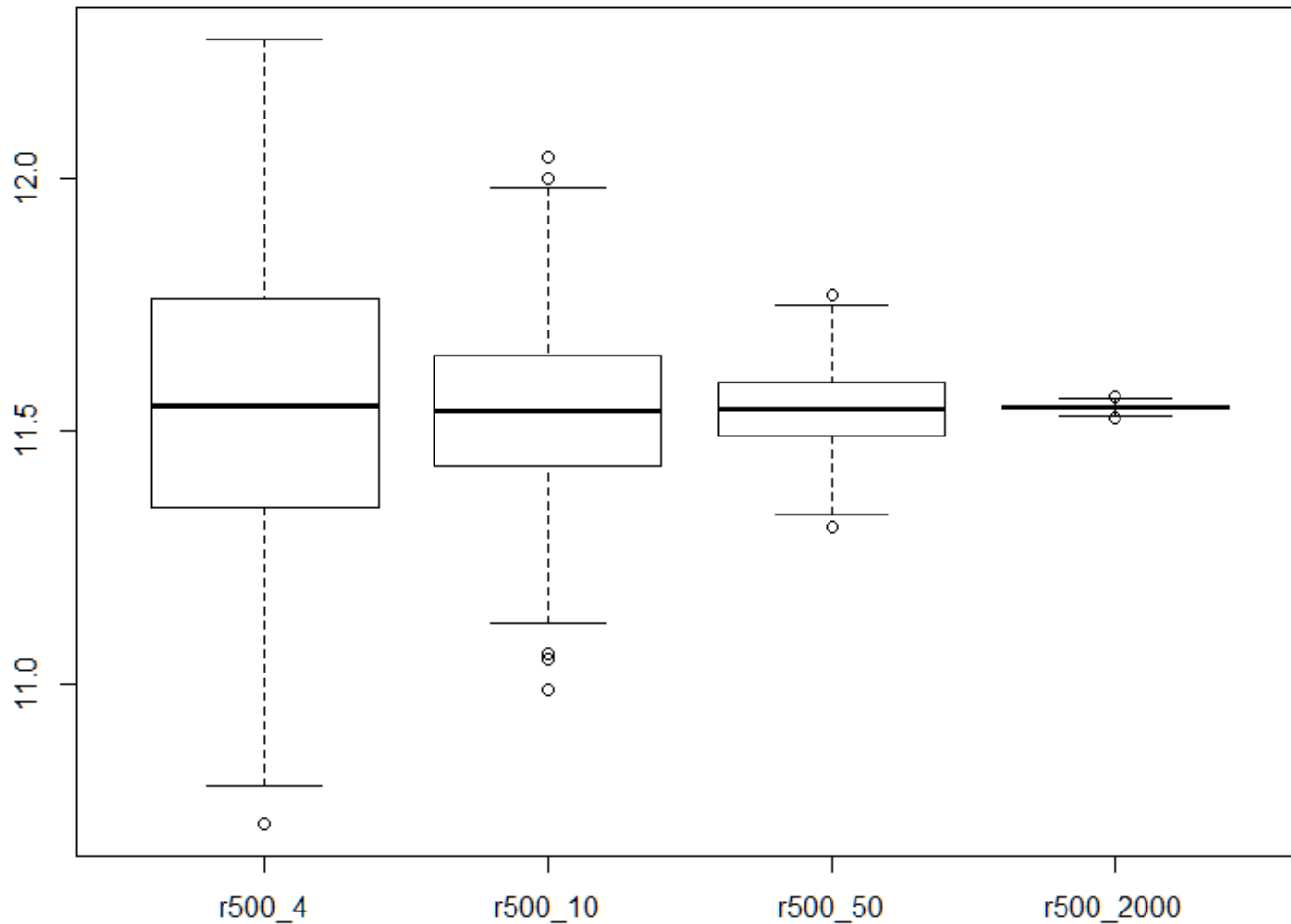
```
r500_50 <- replicate(500, # repeat 500 times
{
mean(sample(FingerL, 50) ) # samples of 50;
} )
```

```
r500_2000 <- replicate(500, # repeat 500 times
{
mean(sample(FingerL, 2000) ) # samples of 2000 ;
} )
```

```
boxplot(r500_4,r500_10,r500_50,r500_2000, names =
c('r500_4','r500_10','r500_50','r500_2000'), main = 'sampling variation
of the mean with different sample sizes' );
```

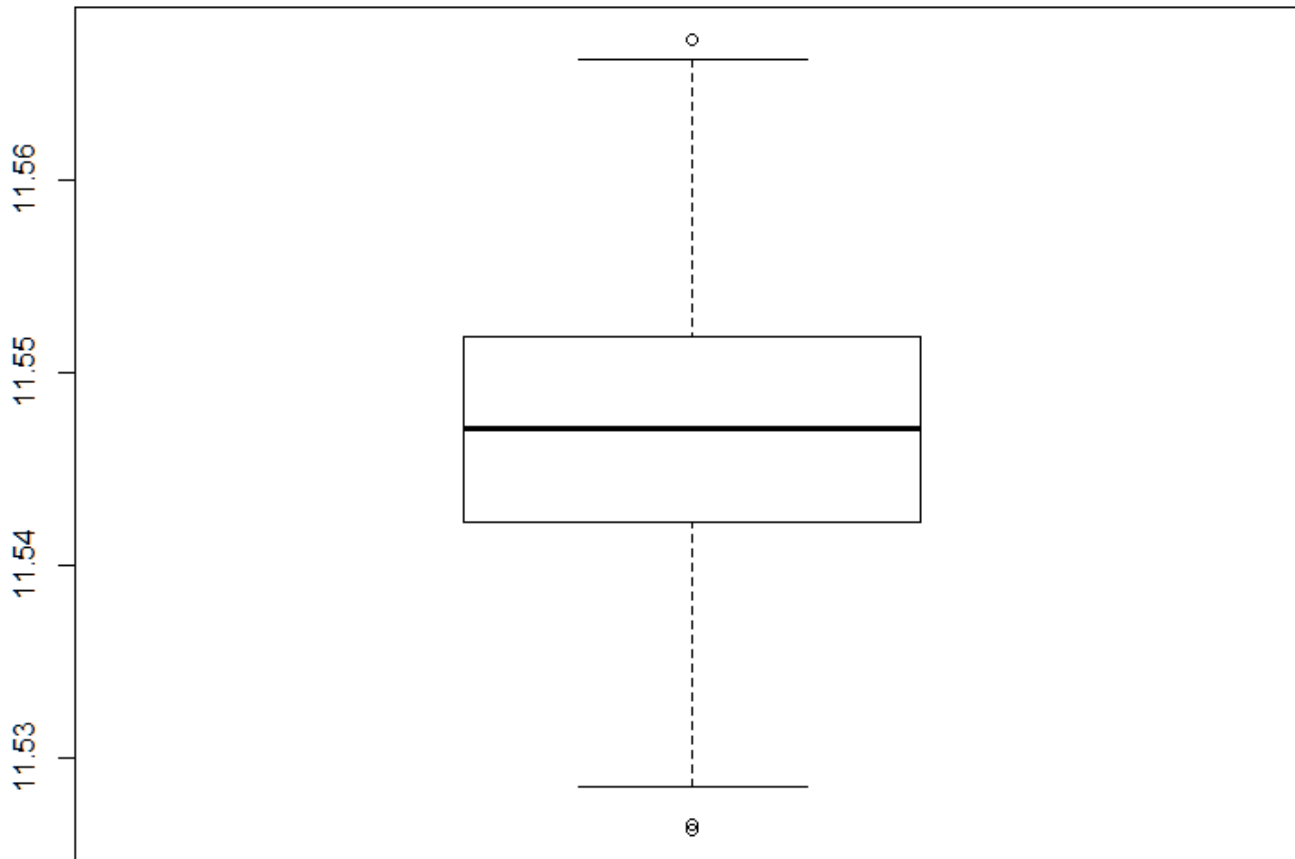
Resampled finger length means

sampling variation of the mean with different sample sizes



>>>> increasing precision of estimates of the mean >>>>

Resampled finger length means: zoom in on 500 samples of 2000

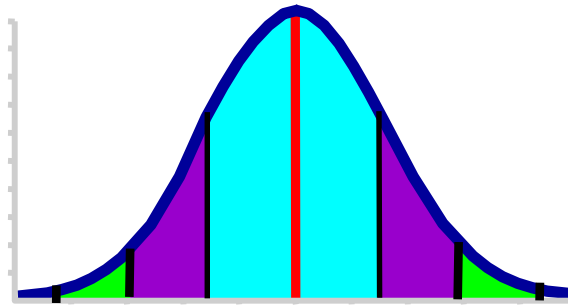


r500_2000

```
boxplot(r500_2000,xlab = 'r500_2000')
```

Getting probabilities based on t

pnorm is used for the probability (of a Z-score)



```
# probability/proportion/area between -1 and 1 (Z) = 68%
```

```
pnorm(1)-pnorm(-1)
```

```
[1] 0.6826895
```

```
diff(pnorm(c(-2,2))) # between -2 and 2 using diff function = 95%
```

```
[1] 0.9544997
```

pt is used for the probability (of a t -score) and requires the df (N-1)

pt(tvalue, df)

```
# probability/proportion/area between -1 and 1 (t)
```

```
pt(1,3)-pt(-1,3) # n = 4 , fewer individuals from -1 to +1 t
```

```
[1] 0.6089978
```

```
diff(pt(c(-2,2),3)) # n = 4, using diff function
```

```
[1] 0.860674
```

CI from a sample mean

\bar{X} is an unbiased estimator of μ , but it is not exact

How often will we be wrong? Choose error rate α , α is a proportion

Our CI = $1 - \alpha$. If we choose a 95% CI, $\alpha = 0.05$.

If we knew σ we could simply use $\bar{X} \pm 1.96 \text{ SEM}$ for our 95% CI

$\bar{X} \pm z^* \text{ SEM}$ (**zstar** = the z value for $1 - \alpha/2$)

$\bar{X} \pm z_{(1 - \alpha/2)} \text{ SEM}$ (`qnorm(0.975)` = 1.96)

- based on the z distribution

But we do not know σ

- our sample is finite ($N = 25$)

- we only know s , so we use "**tstar**"

$\bar{X} \pm t^* \text{ SEM}$ for our 95% CI

$\bar{X} \pm t_{(1 - \alpha/2)} \text{ SEM}$ (`qt(0.975, 24)` = 2.06)

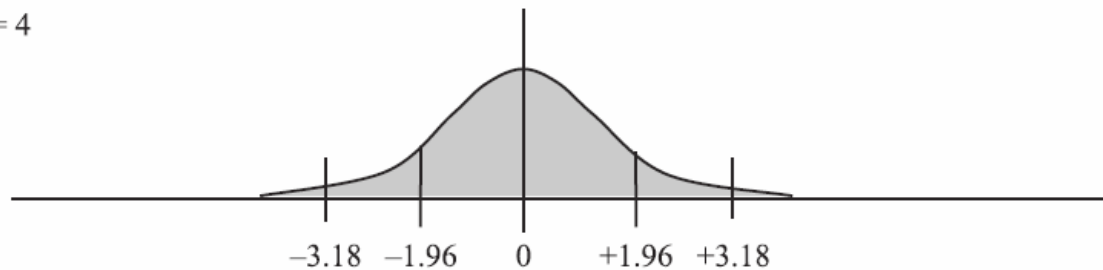
- based on the t distribution

The t distribution

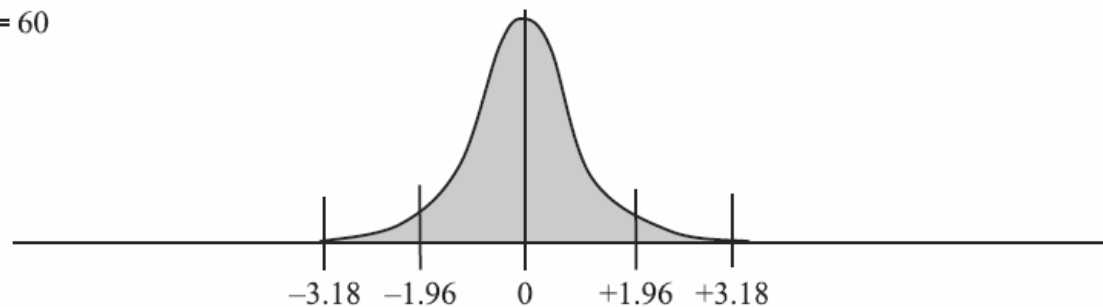
	Formula	Statistic	Sample size	95% confidence interval
(a)	$\frac{\bar{X}-\mu}{s_{\bar{X}}}$	t	4	$\pm 3 \cdot 182$
(b)	$\frac{\bar{X}-\mu}{s_{\bar{X}}}$	t	60	$\pm 2 \cdot 001$
(c)	$\frac{\bar{X}-\mu}{s_{\bar{X}}}$	t	200	$\pm 1 \cdot 972$
(d)	$\frac{\bar{X}-\mu}{s_{\bar{X}}}$	t	1000	$\pm 1 \cdot 962$
(e)	$\frac{\bar{X}-\mu}{s_{\bar{X}}}$	t	∞	$\pm 1 \cdot 96$

t scores need to be higher than z scores to include more of the estimated population variation

(a) $n = 4$



(b) $n = 60$



Getting t based on probability (qt)

use qt to get t-star from a probability and d.f.

`qt(p,df)`

```
# get two-sided 95% CI of a normal sample : qnorm
```

```
# use (1-0.95)/2 = 2.5% on each side
```

```
qnorm((1-0.95)/2)# Z score with 97.5% of sample lower
```

```
[1] -1.959964 # so +/- 1.959964 includes 95%
```

```
qt((1-0.95)/2, 3)# t score with 97.5% of sample lower, n = 4
```

```
[1] -3.182446
```

```
qt((1-0.95)/2, 9)# t score with 97.5% of sample lower, n = 10
```

```
[1] -2.262157
```

```
qt((1-0.95)/2, 29)# t score with 97.5% of sample lower, n = 30
```

```
[1] -2.04523
```

```
qt((1-0.95)/2, 99)# t score with 97.5% of sample lower, n = 100
```

```
[1] -1.984217
```

```
qt((1-0.95)/2, 999)# t score with 97.5% of sample lower, n = 1000
```

```
[1] -1.962341
```

t scores need to be higher than z scores

- to include more of the estimated population variation

Getting probability based on t (pt)

pt is used for the probability (of a t -score) and requires the df (N-1)

`pt(tvalue, df)`

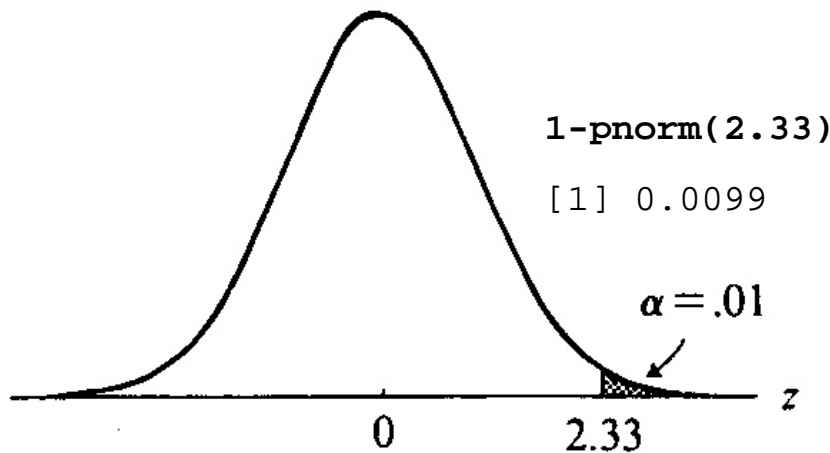
`-2 to +2: pnorm = 0.9544997`

```
diff(pt(c(-2,2),1)) # between -2 and 2, n = 2!!! extreme case
[1] 0.7048328
diff(pt(c(-2,2),19)) # between -2 and 2, n = 20
[1] 0.939998
diff(pt(c(-2,2),49)) # between -2 and 2, n = 50
[1] 0.9489409
diff(pt(c(-2,2),99)) # between -2 and 2, n = 100
[1] 0.9517603
diff(pt(c(-2,2),1999)) # between -2 and 2, n = 2000
[1] 0.9543647
```

The t score reflects more uncertainty in SEM with smaller sample sizes

Probability of a t score: extremes

- note that we are assuming a normal distribution



prob of t value, give df (n-1)

```
1-pt(2.33,9) # N = 10
```

```
[1] 0.02237356
```

```
-pt(2.33, 19)
```

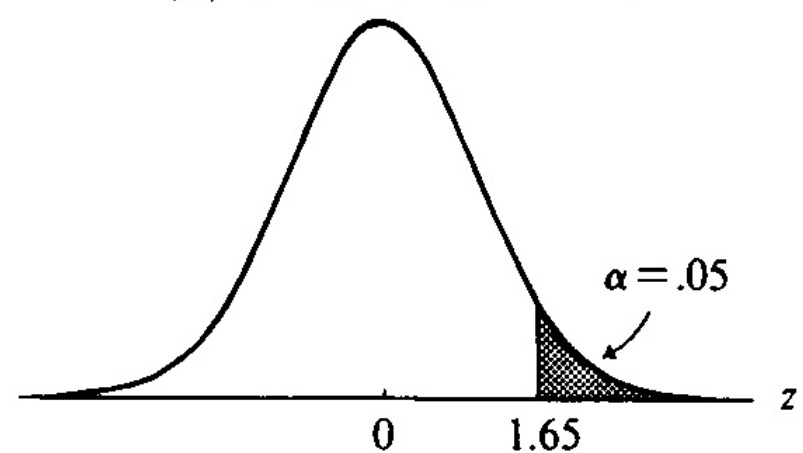
```
[1] 0.0155
```

```
1-pt(2.33,49)
```

```
[1] 0.01198455
```

```
1-pt(2.33,999)
```

```
[1] 0.01000229
```



```
1-pt(1.65, 9)
```

```
[1] 0.066670491
```

```
1-pt(1.65, 19)
```

```
[1] 0.05768864
```

```
1-pt(1.65, 49)
```

```
[1] 0.05266981
```

```
1-pt(1.65, 999)
```

```
[1] 0.04962864
```

Degrees of freedom (df) are used in many tests

Roughly, **df** is the minimum count of data available to calculate a statistic.

Practically, it is the number of independent observations in the data that are used for determining statistical significance.

Essentially, you lose one degree of freedom for every sample statistic you calculate.

To calculate the standard deviation, you must first calculate the mean.

The sum of deviations from the mean = 0.

Once you have calculated $n-1$ deviations, you know the last deviation.

So the degrees of freedom for most single sample, simple tests are $n-1$

... including the t test

x	x - mean	(x-mean) ²
10	-20	400
20	-10	100
30	0	0
40	10	100
50	20	400
total:	150	0
mean:	30	1000

Simple (one variable) chi-square tests have 1 df.

$s =$ 15.81

Why? Because EVERYTHING IS KNOWN (assumed to be exact)!

One-sample *t* test: Example

How much time is spent watching videos on a cell phone by college students?

Collect sample from 8 students, hours per month

```
VidTime <- c(11.9, 2.8, 3.0, 6.2, 4.7, 9.8, 11.1, 7.8)
```

```
mean(VidTime) # x-bar
```

```
[1] 7.1625
```

```
sd(VidTime)
```

```
[1] 3.558867
```

```
df = ???      N = 8
```

```
SEM <- sd(VidTime)/sqrt(7) # SEM = sd/N
```

```
[1] 1.345125 # SEM
```

```
tStar <- qt((1-0.95)/2, df) # t value (tStar) for alpha/2 (two-tailed CI)
```

```
[1] -2.364624
```

```
# CI = mean - tStar * SEM , mean + tStar*SEM (symmetrical)
```

```
mean(VidTime) - abs(tStar) * SEM # lower bound
```

```
[1] 3.981784
```

```
mean(VidTime) + abs(tStar) * SEM # upper bound
```

```
[1] 10.34322
```

More examples

Compare to another mean:

Is our sample mean of 7.16 significantly different from a nationwide average of 5.4 ?

Is it **higher** than the nationwide average?

- convert to t score, then get probability

```
mean(VidTime) - 5.4
```

```
[1] 1.7625
```

```
# divide by SEM to get t score
```

```
(mean(VidTime) - 5.4 ) /SEM
```

```
[1] 1.310287
```

```
# convert 1.310287 SEM units into probability of being at least as extreme
```

```
1-pt(1.31,7) # one-tailed probability (Ha:students > US average)
```

```
[1] 0.1157727
```

```
2 * (1-pt(1.31,7)) # two-tailed probability (Ha: students <> US average)
```

```
[1] 0.2315453
```

Compare sample to hypothesized mean: SUV mpg

Does my SUV get the advertised mpg of 17?

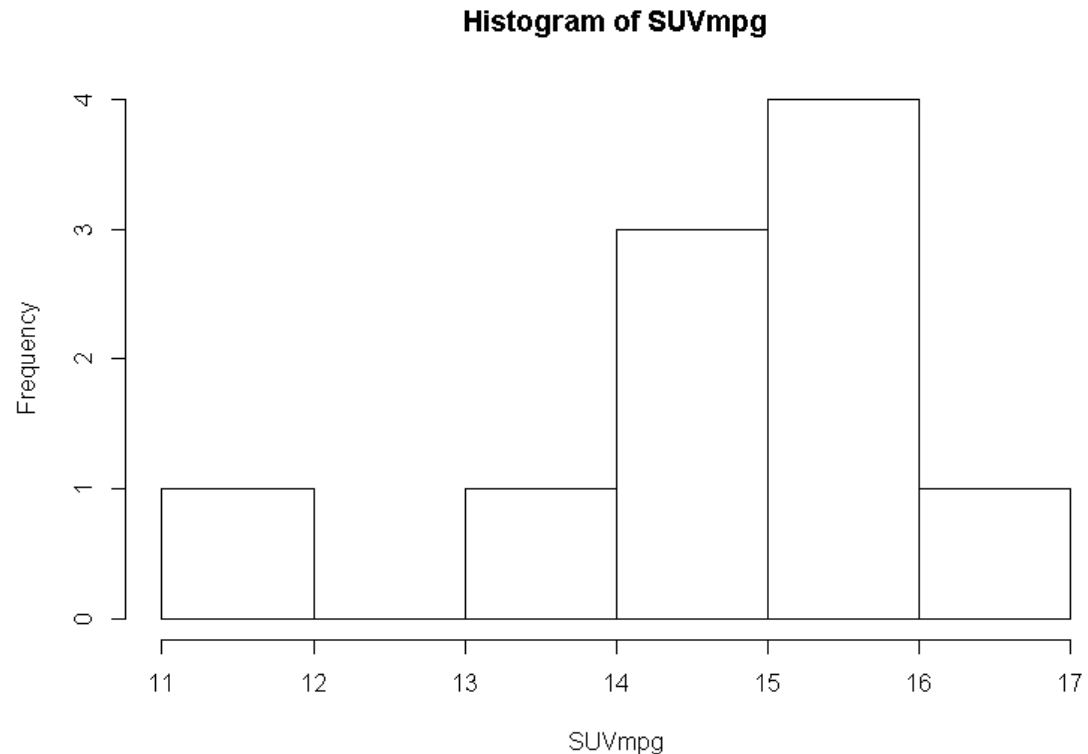
Fill up and record mileage ten times

```
SUVmpg <- c(11.4, 13.1, 14.7, 14.7, 15.0, 15.5, 15.6, 15.9, 16.0, 16.8);
```

```
mean(SUVmpg);
```

```
[1] 14.87
```

```
hist(SUVmpg);
```



Compare sample to hypothesized mean: SUV mpg

We are not interested in an estimate, we want to compare to a stated claim of 17 mpg, so we do a one-sided t-test.

The easier way to do it in R

```
# mu provides hypothesized mean, one-tailed "less": is vector < mu ;  
t.test(SUVmpg, mu = 17, alternative="less", conf.level = 0.95);
```

One Sample t-test

data: SUVmpg

t = -4.2847, df = 9, p-value = 0.001018

alternative hypothesis: true mean is less than 17

95 percent confidence interval:

-Inf 15.78127

sample estimates:

mean of x

14.87

Homework 7

7.1

Measured beak lengths = 6.08, 3.86, 2.28, 2.92, 3.21, 3.52, 4.45, 2.30, 2.83, and 2.24.

Find a 95% CI for the mean using the t distribution.

Is the mean beak length significantly different from 4.2?

7.2

measured leg lengths = 28.10, 21.78, 25.76 , 23.82 ,26.42 ,20.22 ,23.53 ,23.95 ,27.31 ,27.94 ,30.26 ,25.30 ,22.92 ,27.21, and 22.81.

Find a 95% CI and 99% CI for the mean using the t distribution.

Is the mean leg length significantly different from 28.0?

Make sure Email header contains "DATA 500 Homework 4"

Due before class Feb. 27

