

ST 411/511 Outline 2

Reading assignment: Chapter 2. This chapter reviews the details of one- and two-sample t-tests.

Chapter 2 Inference Using t-Distributions

part of brain
associated with
learning

Case Study 2.1.2 Twin study. How big is the difference in volumes of left hippocampus where one monozygotic twin has schizophrenia and the other does not?

identical

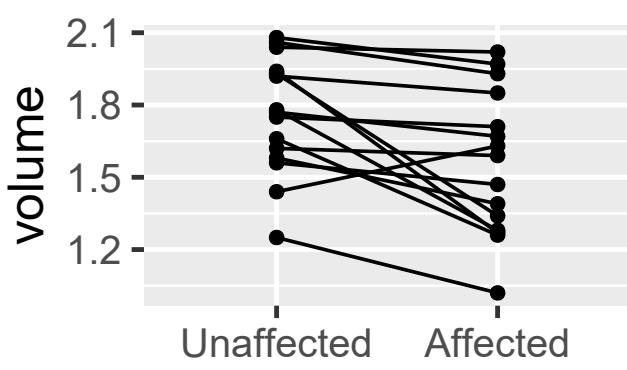
psychological disease

DISPLAY 2.2 Differences in volumes (cm^3) of left hippocampus in 15 sets of monozygotic twins where one twin is affected by schizophrenia

Pair #	Unaffected	Affected	Difference	Differences
1	1.94	1.27	0.67	-2 9
2	1.44	1.63	-0.19	-1 0 23479
3	1.56	1.47	0.09	-0 1 0139
4	1.58	1.39	0.19	2 3
5	2.06	1.93	0.13	3 4 0
6	1.66	1.26	0.40	5 09
7	1.75	1.71	0.04	6 7
8	1.77	1.67	0.10	7
9	1.78	1.28	0.50	
10	1.92	1.85	0.07	
11	1.25	1.02	0.23	
12	1.93	1.34	0.59	
13	2.04	2.02	0.02	
14	1.62	1.59	0.03	
15	2.08	1.97	0.11	

Average: 0.199
Sample SD: 0.238
 $n: 15$

Legend: | 6 | 7 represents 0.67 cm^3



Important: Data are paired. 30 numbers but only 15 subjects.

Using twins helps remove genetic variation so we can better detect a diff. in volume. We have only one sample, not two.

Twin Study Hypothesis Test

Notation:

Y_i = diff. in volume for pair i ($i=1, \dots, 15$)

\bar{Y} = sample mean of $Y_1, \dots, Y_{15} = 0.199$

μ = pop'n mean diff.

s = sample std. dev. = 0.238

σ = pop'n std. dev.

n = sample size = 15

Research question: Is there a difference in volume?

Null hypothesis: $H_0: \mu = 0$ $H_A: \mu \neq 0$

Test statistic: general form of a t-statistic is

$$t\text{-stat} = \frac{\text{point estimate of parameter} - \text{null hypothesized value}}{\text{SE of point estimate}}$$

parameter = μ = pop'n mean (see H_0)

point estimate of parameter = $\bar{Y} = 0.199$ = sample mean

null hypothesized value = 0

SE of point estimate = $SE(\bar{Y})$ = est. of std. dev.
of \bar{Y} .

standard deviation of \bar{Y} = $\frac{\text{popn std. dev.}}{\sqrt{\text{sample size}}} = \frac{\sigma}{\sqrt{n}}$

not obvious but makes sense.

Std. dev of \bar{Y} larger when σ
larger, but smaller when sample
size is larger.

$$SE(\bar{Y}) = \frac{\sigma}{\sqrt{n}} = \frac{0.238}{\sqrt{15}}$$

$$t\text{-stat} = \frac{0.199 - 0}{0.238 / \sqrt{15}} = 3.238$$

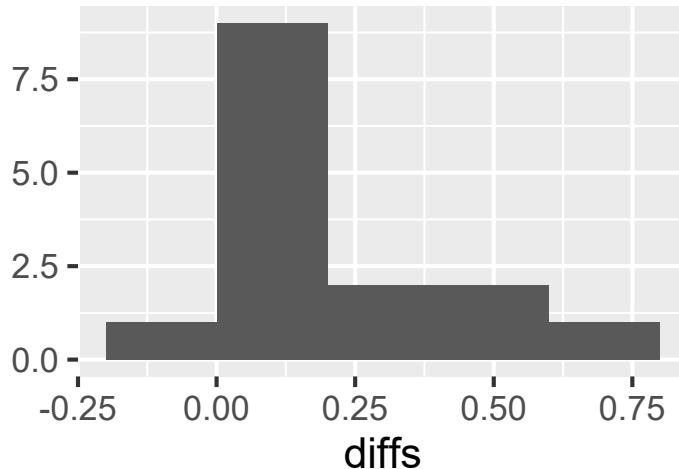
Sampling distribution of the test statistic under the null hypothesis:

Under assumptions we'll see in ch. 3
 t -stat always has a t dist'n.
with df = $n - \# \text{ mean parameters}$
 $15 - 1 = 14$

R Code

You did a 2-sample t-test in Lab 1. This is one-sample t-test.

```
> # Load packages
> library(ggplot2)
> library(Sleuth3)
>
> # View the first few lines of data
> head(case0202) ←
  Unaffected Affected
1      1.94     1.27
2      1.44     1.63
3      1.56     1.47
4      1.58     1.39
5      2.06     1.93
6      1.66     1.26
>
> # Calculate the differences in volume.
> diffs ← case0202$Unaffected - case0202$Affected
  ↗ assigns variable "diffs" a value
>
>
> head(diffs)
[1]  0.67 -0.19  0.09  0.19  0.13  0.40 ← differences
>
> # Check a histogram of the differences
> qplot(diffs, geom="histogram", breaks=c(-0.2,0,0.2,0.4,0.6,0.8))
```



```

> # Calculate the mean and standard deviation of diffs
> mean(diffs)
[1] 0.1986667 ←  $\bar{Y}$ 
> sd(diffs)
[1] 0.2382935 ←  $S$ 
>
> # Sample size:
> nrow(case0202)
[1] 15 ←  $n$ 
>
> # nrow() doesn't work on diffs because diffs isn't a data frame.
> nrow(diffs)
NULL
>
> # Do a t-test.
> t.test(diffs)

```

Since `diffs` is a single column of numbers, R figures we want a 1-Sample t-test.

```

One Sample t-test
data: diffs t -stat
t = 3.2289, df = 14, p-value = 0.006062
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.0667041 0.3306292
sample estimates:
mean of x
0.1986667

```

t -stat

$n-1$

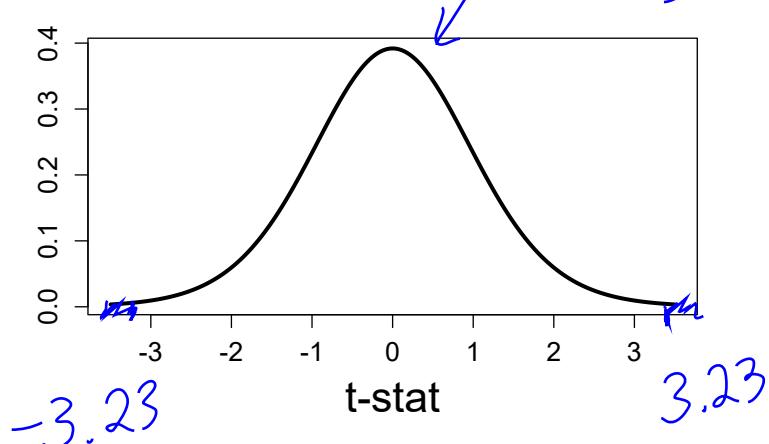
p-value compares obs. t -stat to Sampling dist'n if H_0 true

$H_A: \mu \neq 0$

\bar{Y}

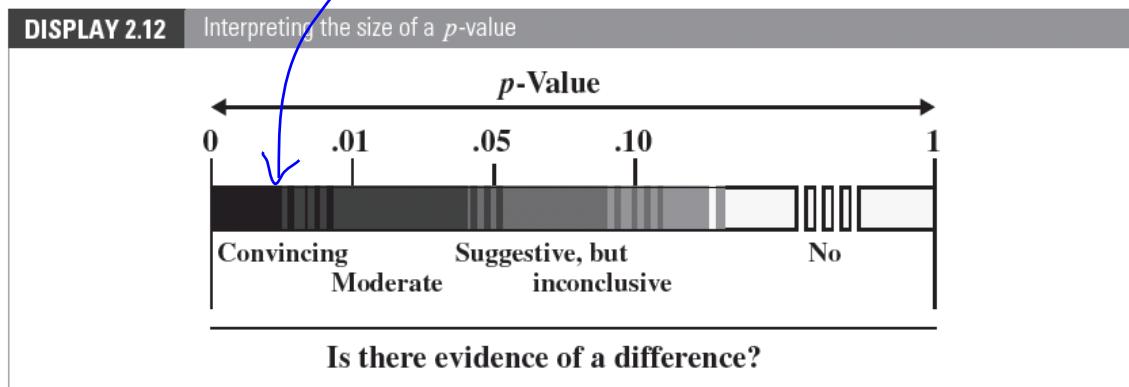
CI. We will discuss more later.

Interpreting the p-value



$$p = 0.006062$$

$$p \approx 0.006$$



Writing a statistical conclusion for a hypothesis test.

Elements of a statistical conclusion for a hypothesis test:

- (A) A quantification of the evidence against the null in favor of the alternative
- (B) A paraphrase of the hypotheses
- (C) The p-value and mention of the test performed

Statistical conclusion for hypothesis test in twin study:

There is convincing evidence [that
the pop'n mean diff in volume of
left hippocampus between affected
+ unaffected twins is not 0]
(p-value ≈ 0.006 , one-sample t-test.)

Scope of inference for twin study:

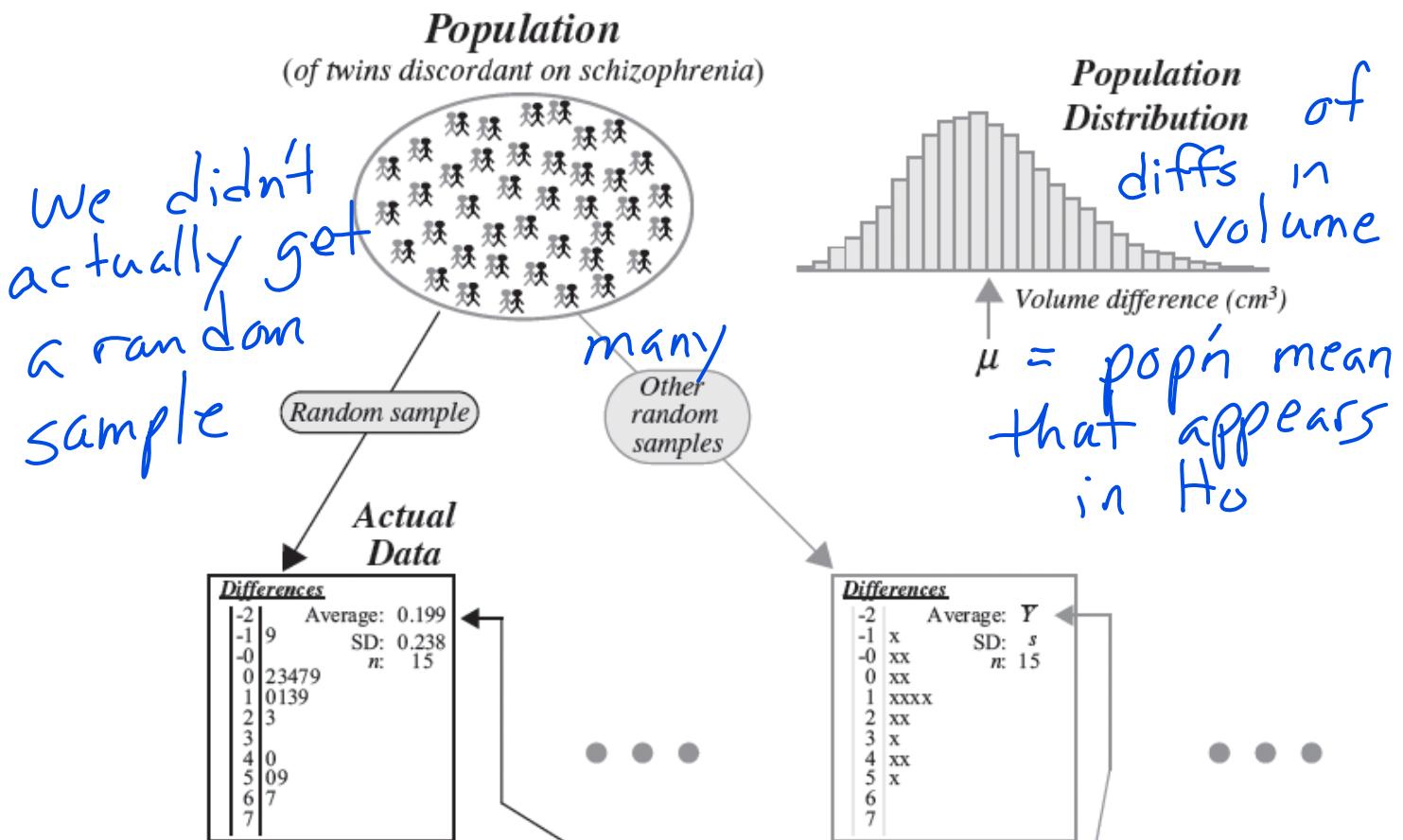
Since subjects were not randomly
assigned to groups, can't infer
causation. Since twins not randomly
drawn from a population, can only
infer results to subjects observed.

Behind the scenes of a t-test:

Ideal situation - more conceptual than actual

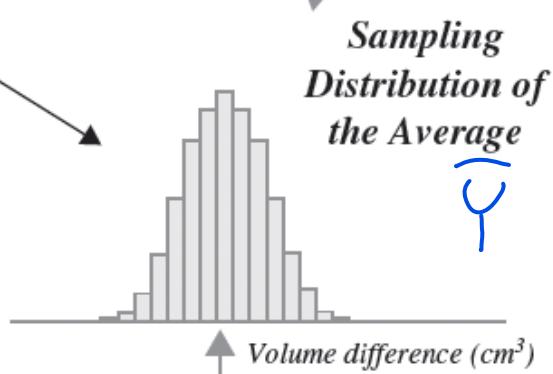
DISPLAY 2.3

The sampling distribution of the sample average



Repeated Sampling from pop'n gives Sampling dist'n of a statistic (e.g. \bar{Y} or t-stat).

Sampling dist'n of Sample mean has std. dev. = $\frac{s}{\sqrt{n}}$



mean of Sampling dist'n of \bar{Y} is Pop'n mean μ .
est. width $\frac{s}{\sqrt{n}} = \text{SE}(\bar{Y})$.

Simulation studies for evaluating statistical procedures

- Statisticians use simulated data to illustrate and evaluate statistical procedures.
 - By simulating many random data sets from a given distribution, we can simulate “repeated sampling.”

R has several functions to simulate data from different dist'n's.

- Assess how well our procedure works over many repetitions of the study.

*Simulate data. Apply procedure (e.g. t-test)
Repeat many times to see long term performance*

- Don't use simulated data to replace real data!

You must tell your audience you're using simulated data.

- Some of the questions simulation can answer:

- If the true mean difference is non-0, what percentage of the time will the t-test detect that?

*This is power. Simulate data where $\mu \neq 0$.
Keep track of how often we get a small p-value.*

- If the true mean difference is 0, what percentage of the time will we erroneously conclude otherwise?

This is type I error. Simulated data where $\mu = 0$. Keep track of how often we get a small p-value.

- If the assumptions are violated, how valid are the t-test results?

Simulate data violating assumptions and keep track of t-test's performance.

Some simulated data: $n = 15$, $\mu = 0.199$, and $\sigma = 0.238$

\bar{Y} from twin study s from twin study

Simulate normal data that looks like observed twin data.

$H_0: \mu = 0$, so this is false.

Sample1	Sample2	Sample3	Sample4	...	Sample100
0.034368712	0.285687424	0.116097648	0.119313798	...	-0.125466602
-0.147501662	0.219390071	0.286322663	0.313450623	...	0.088468446
0.064472217	0.490113750	0.241635142	-0.131699550	...	0.536262267
0.281468713	0.275981920	0.220656241	-0.057754558	...	0.496339000
0.195995701	0.185448629	-0.245751121	0.338914612	...	0.463845305
0.336464624	-0.024340035	0.252574770	0.460485414	...	0.551483270
-0.020218726	0.087148705	0.310385406	0.448660227	...	0.018167965
0.063277938	0.287297140	0.209316425	0.432180127	...	0.240164253
0.357405533	0.311386413	0.247194749	-0.021536929	...	0.453374302
-0.062795791	0.107883615	-0.043621638	0.584253655	...	-0.010039116
0.268787747	0.180521393	-0.291142315	0.057724004	...	0.297346641
0.441313448	0.436914340	-0.068840100	0.253761768	...	0.546447822
-0.072620591	0.018455868	0.135530878	0.341361977	...	0.304928222
0.251458354	-0.305503107	0.337973050	0.085071554	...	0.353577163

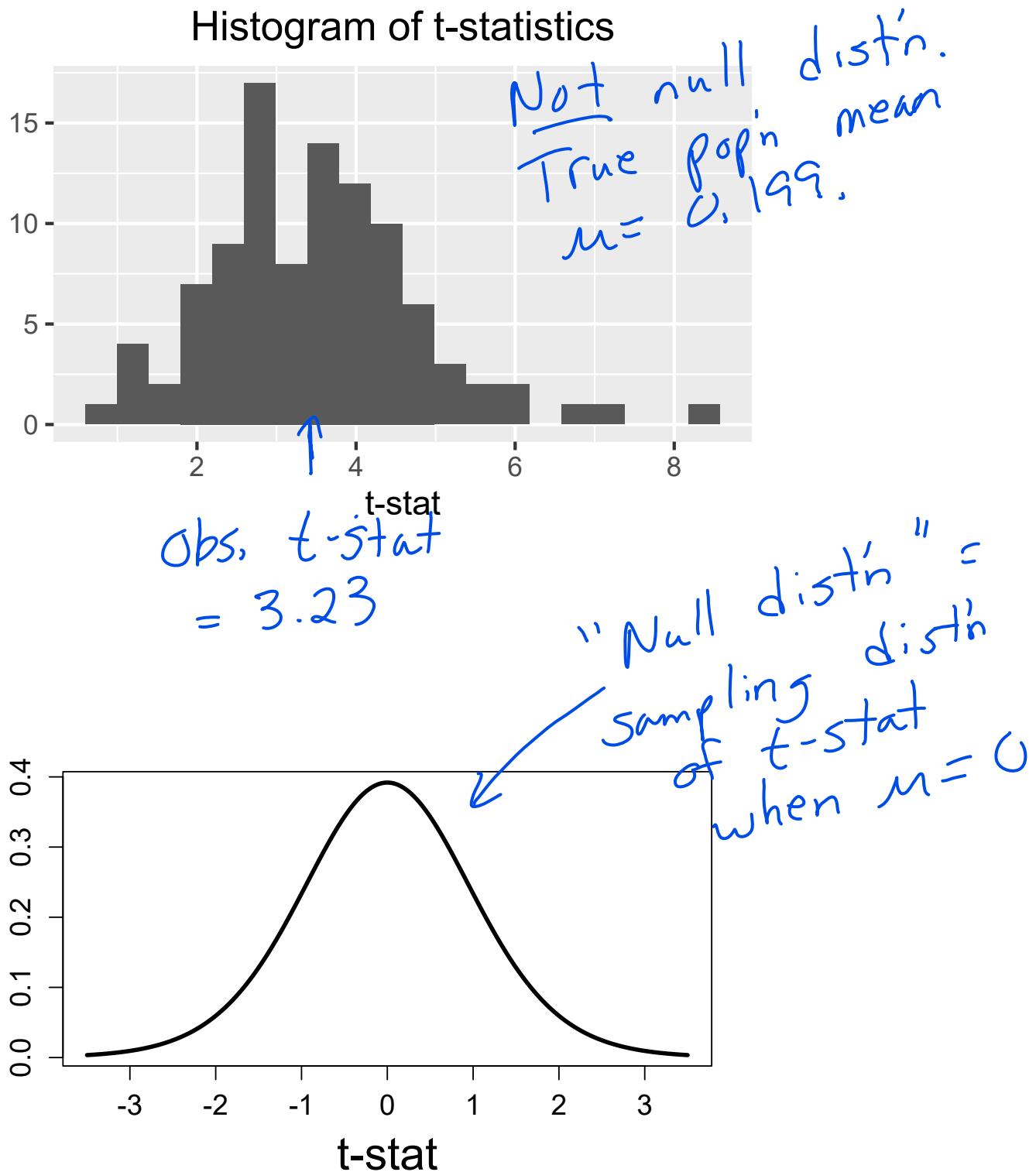
`t.test()` command gives

100 t-stats

100 p-values

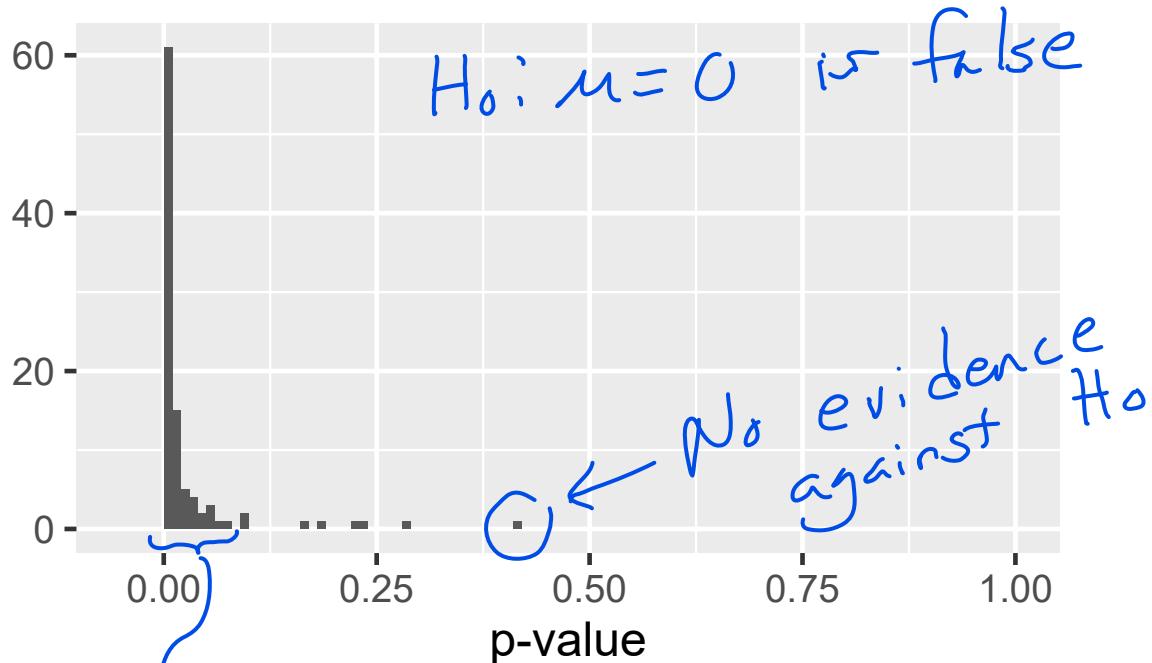
100 CI's

Using simulated data to show distribution of t-statistics:

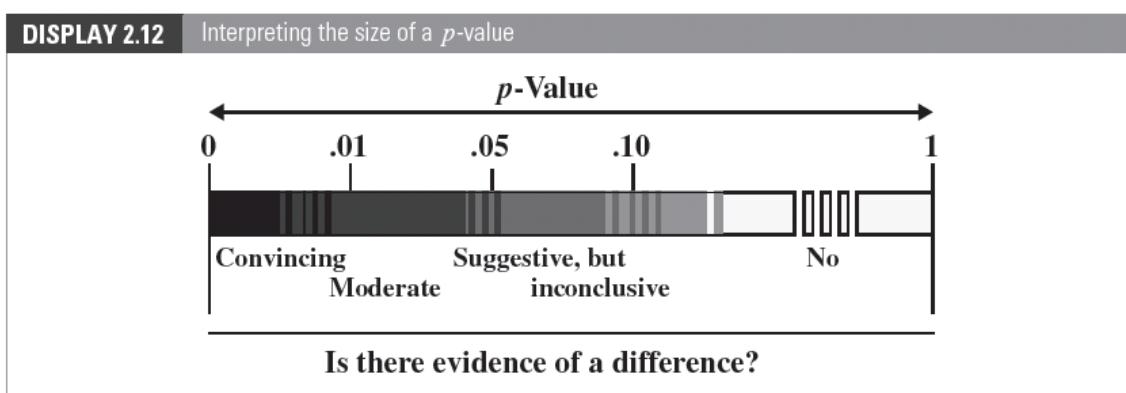


Using simulated data to show distribution of p-values:

Histogram of p-values



"Convincing evidence against H_0 " When $\mu = 0.199 \pm 0$
we usually get a very small p-value.



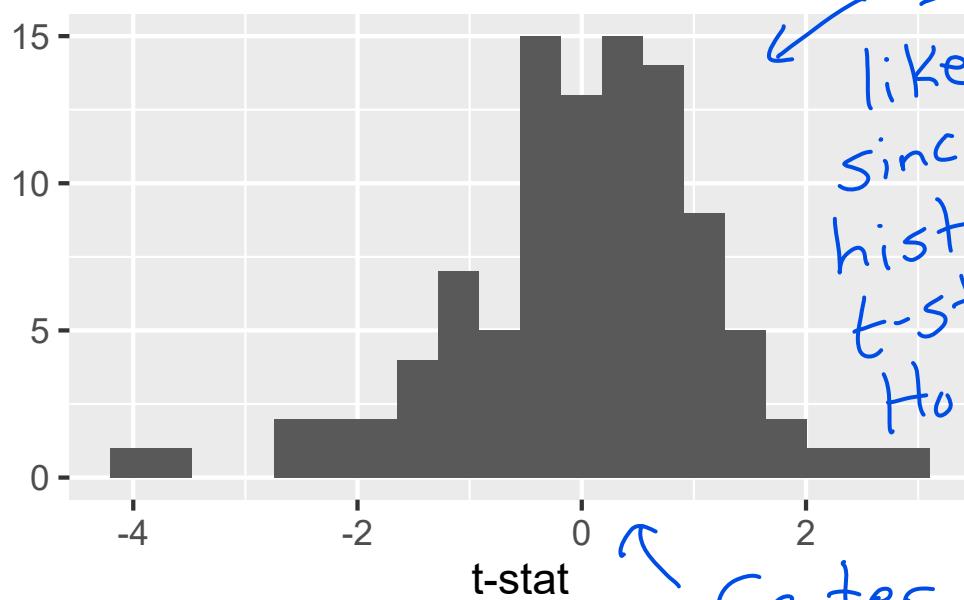
More simulated data: $n = 15$, $\mu = 0$, and $\sigma = 0.238$

Normal popn and $H_0: \mu=0$ is true

Sample1	Sample2	Sample3	Sample4	...	Sample100
-0.129190078	-0.072584935	-0.105937608	0.019780645	...	-0.007479690
0.027901268	-0.538533578	0.309039116	0.319456494	...	-0.049198426
0.189014106	0.499493774	-0.296154515	-0.089905546	...	-0.099440144
0.080470963	-0.043702247	0.281030687	0.036425130	...	0.116033219
-0.200544434	-0.050590742	-0.188859660	0.173931964	...	0.187117528
0.228098587	-0.374522767	-0.116184561	0.638739612	...	-0.211951734
-0.468567701	-0.467341149	-0.227939131	0.245657611	...	0.159991414
0.251888843	-0.245557968	-0.095140089	-0.278385545	...	-0.017804637
0.007354155	-0.396758505	-0.054433588	0.013864734	...	0.005779589
0.081820135	-0.093887720	0.295958511	-0.257709996	...	-0.315667896
-0.013873494	-0.117704040	-0.390590396	-0.047894376	...	-0.028402167
-0.273918112	-0.077778701	-0.552782760	-0.223940700	...	-0.372463192
0.073710668	0.172905812	-0.013511682	0.261786198	...	0.080774950
-0.062238555	-0.358165356	-0.177405679	0.139528159	...	-0.344842570

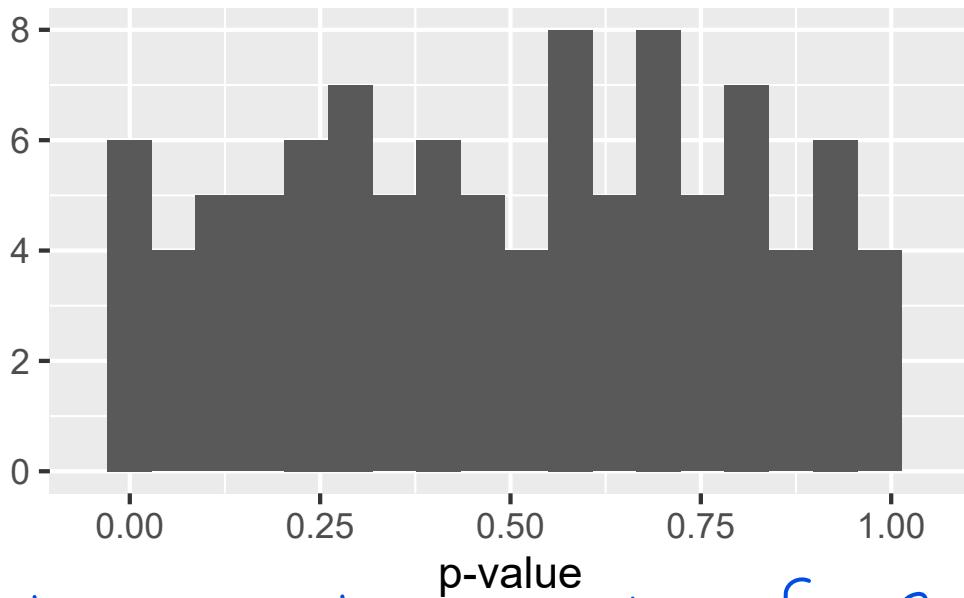
100 t-stats
100 p-values
100 CI's

Histogram of t-statistics



t-stat of 3.23 is very unusual if $H_0: \mu=0$ is true.

Histogram of p-values



Fact: Sampling dist'n of p-value when H_0 true is uniform. So when H_0 true, 5% of the time, p-value < 0.05.

Inferential Errors in Hypothesis Testing

Just by chance, we could get an unusual sample and make an incorrect inference.

Finding evidence against H_0 when H_0 is true.

Type I error. This is considered the worst error (usually). It's making a false claim. Even if H_0 is true, 1% of the time (over repeated sampling), $P < 0.01$.

Not finding evidence against H_0 when H_0 is false.

(Type II error)

If H_0 false can still get a large p-value. This error is related to the power of the test, the ability to detect evidence against H_0 .

Can increase power by taking a larger sample.

A large enough sample can detect a false H_0 even when it's just barely false. This is the issue of practical vs. statistical significance. See section 4.5.1.

Confidence intervals

Need to estimate popn mean diff.

Recall research question. How big is the difference in volumes of left hippocampus where one monozygotic twin has schizophrenia and the other does not?

Hypothesis test measures strength of evidence against H_0 . CI estimates a parameter (e.g. μ) by giving a range of plausible values for it.

Usual formula for a confidence interval of a parameter:

$$\text{point estimate of parameter } \hat{\mu} \pm t_{df}(1 - \alpha/2) \text{ SE(point estimate)}$$

$\hat{\mu}$ margin of error

$$SE(\hat{\mu}) = \frac{s}{\sqrt{n}}$$

Recall: $t\text{-stat} = \frac{\text{pt. est} - \text{value under } H_0}{\text{SE(pt. est.)}}$

$$= \frac{\bar{Y} - 0}{s/\sqrt{n}}$$

Notation

$t_{df}(1 - \alpha/2)$ = quantile from a t dist'n.

df = degrees of freedom

= sample size - # mean parameters

= $15 - 1 = 14$ for twin study

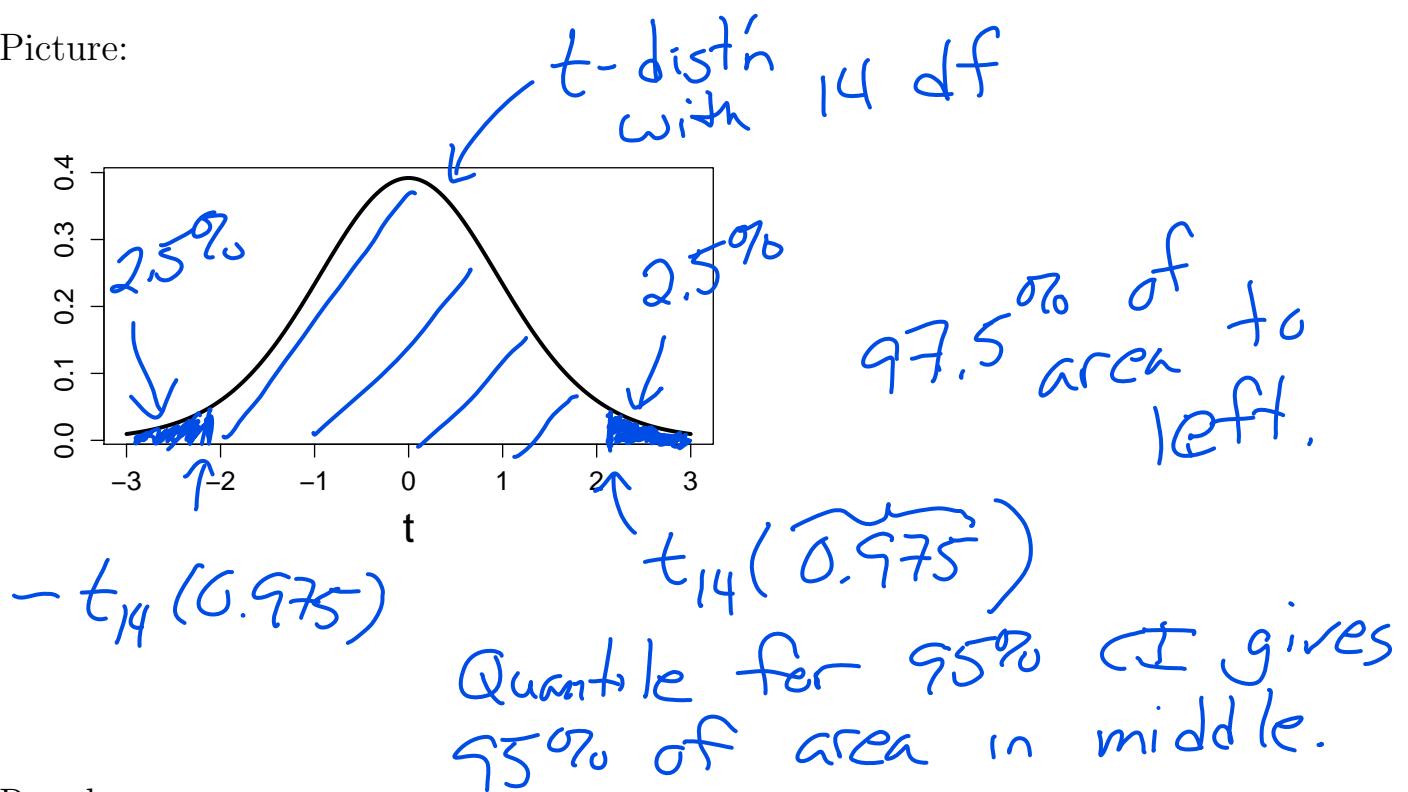
$1 - \alpha$ = confidence level

So if $\alpha = 0.05$, we have a 95% CI.

t-quantile

Notation (same as textbook): $t_{df}(1 - \alpha/2) = t_{14}(0.975)$

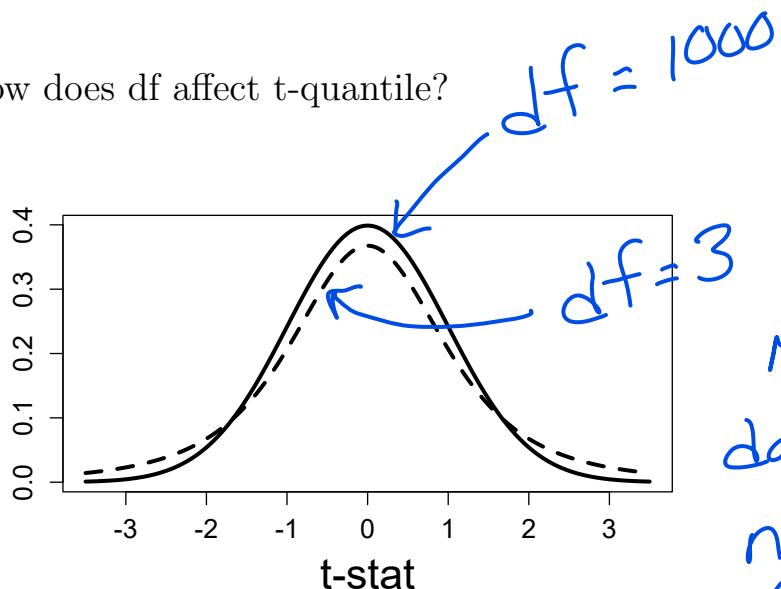
Picture:



R code:

```
> # 0.975 quantile from t distribution with 14 df  
> qt(0.975, 14)  
[1] 2.144787
```

How does df affect t-quantile?



More df (ie. more data) \Rightarrow narrower CI's.
(more precise)

CI for twin study: $\bar{Y} \pm S$ from p. 5 of outline 2

point estimate of parameter $\pm t_{df}(1 - \alpha/2) \text{ SE(point estimate)}$

$$0.199 \pm 2.145 \cdot \frac{0.238}{\sqrt{15}} \quad \begin{matrix} \text{conf.} \\ \text{level} \end{matrix}$$
$$\approx \boxed{(0.06, 0.33)} \quad \leftarrow 95\% \text{ CI}$$

Writing a statistical conclusion for a confidence interval

(from Syllabus)

Elements of a statistical conclusion for a confidence interval:

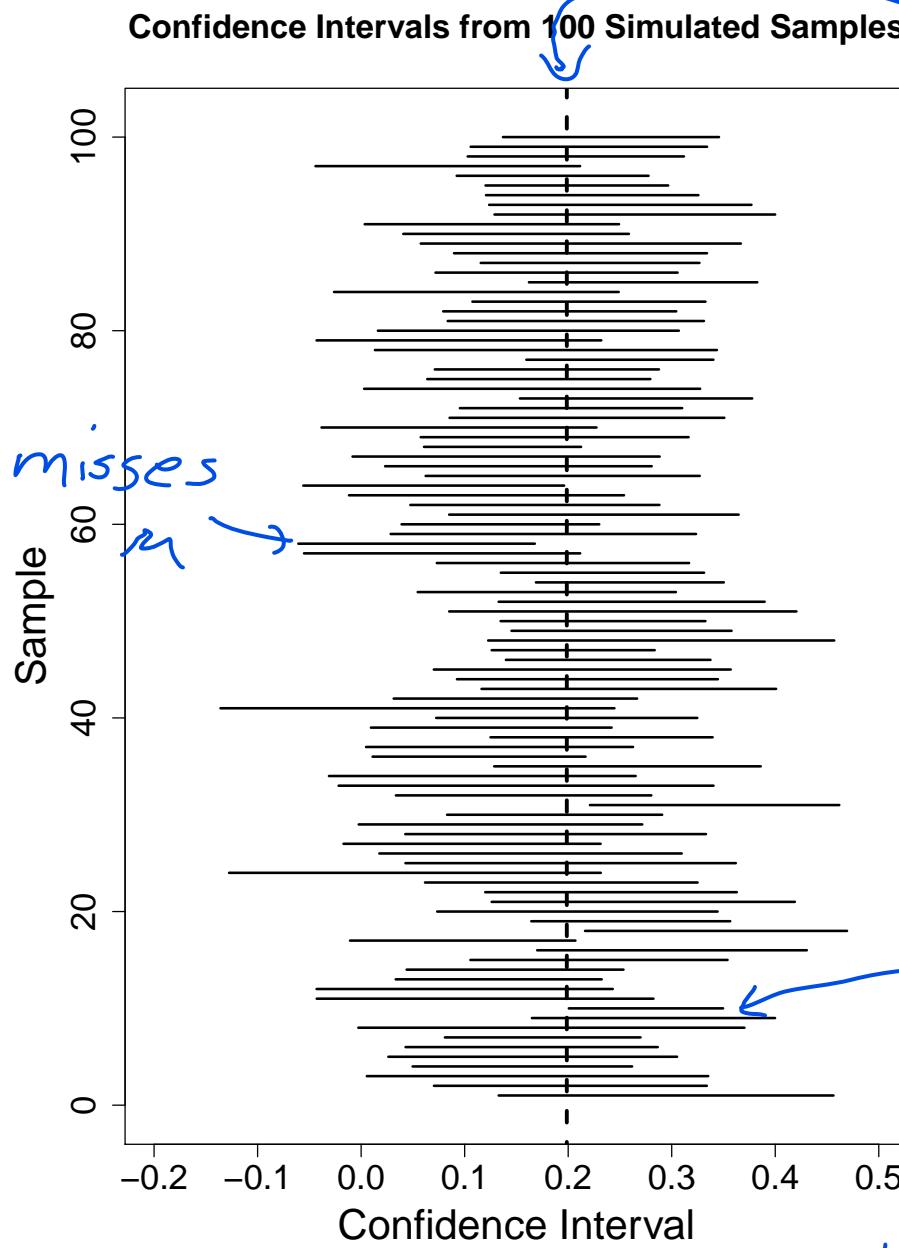
- (A) A statement indicating "estimation" and what quantity was estimated
- (B) The endpoints of the confidence interval (point estimate optional)
- (C) The confidence level

Statistical Conclusion:

A [We estimate the mean diff. in volume between left hippocampi of unaffected vs. affect twins]
B
C is 0.06 to 0.33 cm³ (95% CI)
quantity estimated

Using simulated data to illustrate confidence interval interpretation:

Same data as before with $\mu = 0.199$



Each line segment represents a 95% CI
Most CIs contain true μ

misses μ

About 5% of 95% CI's miss true parameter being estimated. 99% CI's miss about 10% of the time. 99% CIs are wider than 95% CIs (less precise).

Duality between confidence interval and hypothesis test:

Consider $H_0: \mu = \mu_0$ \leftarrow almost always 0
If ^{2-sided} p-value > 0.05 , then μ_0 would
be in the 95% CI. A CI gives a
range of plausible values for parameter
being estimated. If μ_0 is a plausible
value, then p-value will be large.

Note: If you do a 1-sided test in R using `t.test()`, you'll get a 1-sided confidence interval.

```
> # One-sided t-test.  
> t.test(diffs, alternative="greater")
```

$$H_0: \mu = 0$$
$$H_A: \mu > 0$$

One Sample t-test

```
data:  diffs  
t = 3.2289, df = 14, p-value = 0.003031  
alternative hypothesis: true mean is greater than 0  
95 percent confidence interval:
```

0.09029832 Inf $\leftarrow (0.09, \infty)$

sample estimates:
mean of x
0.1986667 we estimate pop'n mean diff
is at least 0.09 cm^3
(95% one-sided CI).

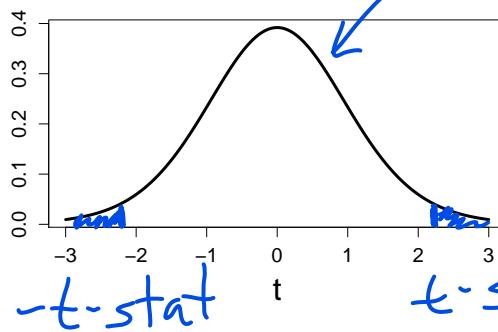
One-sided CIs are unusual.

One-sided vs. two-sided tests Determined by research question, not data.

$$H_0 : \mu = 0 \text{ vs. } H_A : \mu \neq 0$$

two-sided
H_A

$$t\text{-stat} = \frac{\bar{Y}}{S/\sqrt{n}}$$



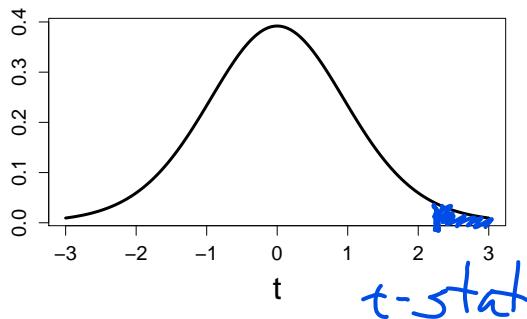
t-distn with 14 df

p-value =
shaded
area

$$H_0 : \mu = 0 \text{ vs. } H_A : \mu > 0$$

one-sided H_A

$$t\text{-stat} = \frac{\bar{Y}}{S/\sqrt{n}}$$

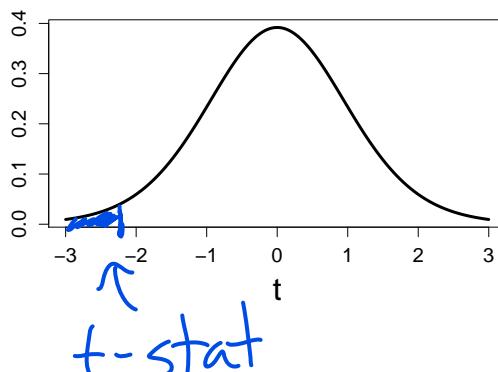


p-value =
area to
the right
of t-stat

$$H_0 : \mu = 0 \text{ vs. } H_A : \mu < 0$$

one-sided H_A

$$t\text{-stat} = \frac{\bar{Y}}{S/\sqrt{n}}$$



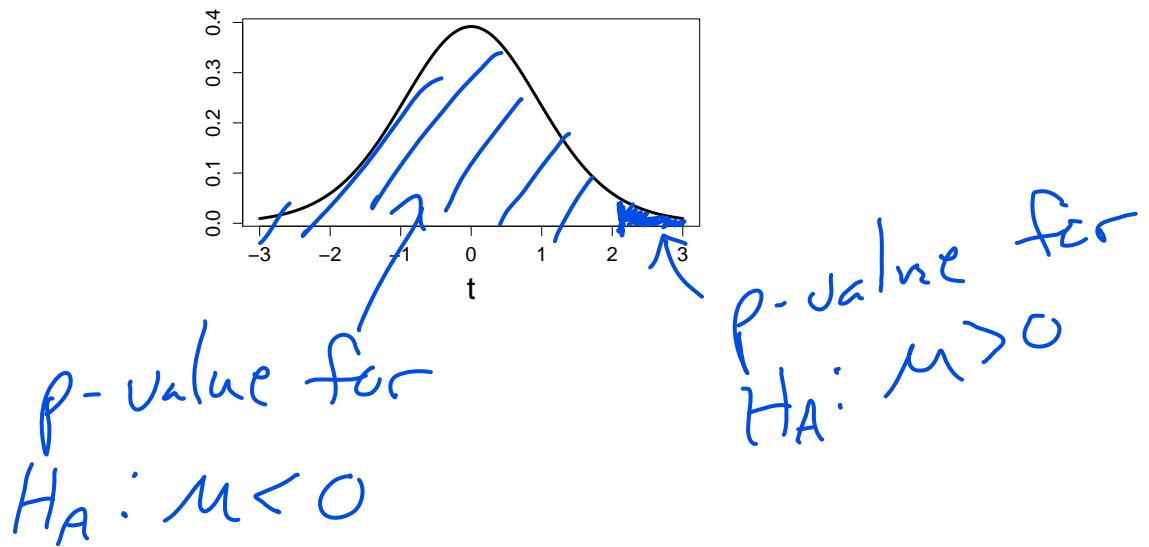
p-value =
area to
left of
t-stat

(picture assumes
 \bar{Y} was negative)

Caution! H_0 and H_A should be chosen before looking at the data.

Looking at data to choose test is called data snooping or p-hacking and it's very bad manners. Can halve p-value by choosing the optimal one-sided H_A . This is dishonest, so don't do it.

Even if you plan a one-sided test, can't use data to decide which side.



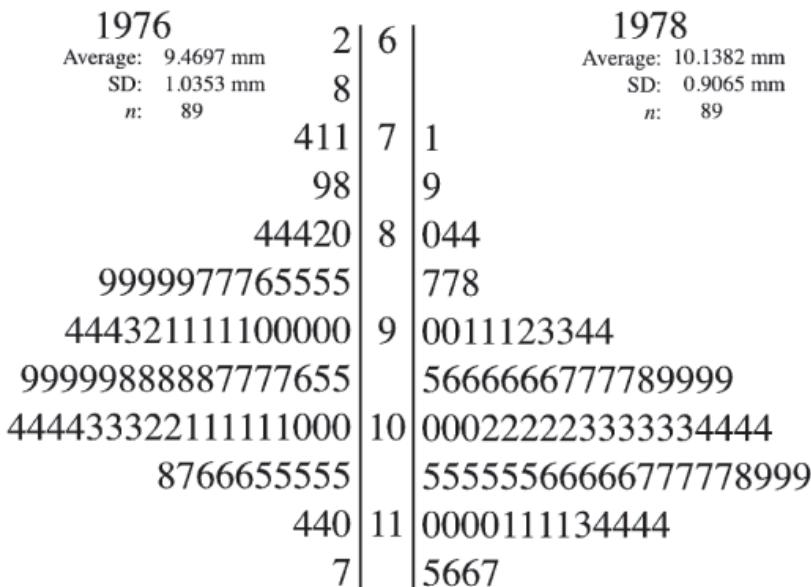
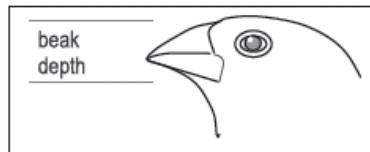
Two-sample t-test ← Data not paired. one-sided HA

Case Study 2.1.1 Is there evidence that beak depth in 1978 is larger than in 1976 for a population of finches?

Drought in 1977. Looking for evidence that finches selected for deeper beaks.

DISPLAY 2.1

Beak depths (mm) of Darwin finches on Daphne Major in 1976, pre-drought, and 1978, post-drought



Legend: | 11 | 0 = 11.0 mm

Data not paired because there is no mechanism to match finch from 1976 to one from 1978.

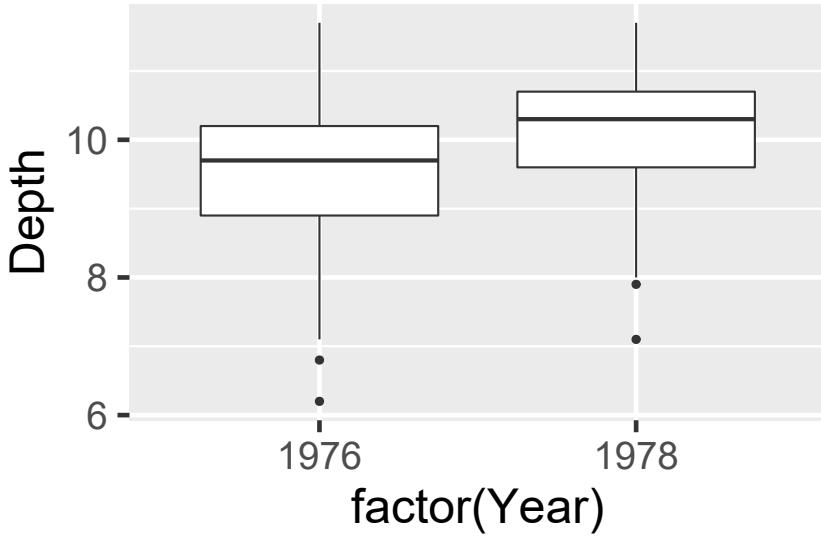
grouping
var response

```
> head(case0201) # Sleuth3 package
```

	Year	Depth
1	1976	6.2
2	1976	6.8
3	1976	7.1
4	1976	7.1
5	1976	7.4
6	1976	7.8

```
# ggplot2 package
```

```
> qplot(factor(Year), Depth, data=case0201, geom="boxplot")
```



1978 depths
are larger
but is this
just due to
randomness?

Need to do a
hypothesis test.

Finch Study Hypothesis Test (both "by hand" and with R)

Null and alternative hypotheses:

$$\begin{aligned} \mu_1 &= \text{pop'n mean in 1976} \\ \mu_2 &= " \quad " \quad 1978 \end{aligned}$$

$$H_0: \mu_1 - \mu_2 = 0 \quad (\text{or } H_0: \mu_1 = \mu_2)$$

$$H_A: \mu_1 - \mu_2 < 0 \quad (\text{or } H_A: \mu_1 < \mu_2)$$

Since research question is asking
if mean in 1978 is larger.

Test statistic:

$$t\text{-stat} = \frac{\text{point estimate of parameter} - \text{null hypothesized value}}{\text{SE of point estimate}}$$

$$= \frac{\bar{Y}_1 - \bar{Y}_2 - 0}{\text{SE}(\bar{Y}_1 - \bar{Y}_2)}$$

$$\text{SE}(\bar{Y}_1 - \bar{Y}_2) = \text{estimate of std. dev. of } \bar{Y}_1 - \bar{Y}_2$$

$$\text{std. dev. of } \bar{Y}_1 - \bar{Y}_2 = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

Estimating σ = pop'n std. dev. for both pop'n
 So use data from both samples to estimate σ .

pooled est. of σ

$$\text{estimate of } \sigma = s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

s_1, s_2 are the sample std. devs.
 s_1^2, s_2^2 are sample variances

Note: the quantity inside square root is a weighted average of sample variances where weights are df from each sample. Sample with more data has more effect on s_p .

$$SE(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Sampling distribution of the test statistic under the null hypothesis:

$$t\text{-stat} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Sampling dist'n is a t-dist'n. with df = sample size - # mean parameters

$$= n_1 + n_2 - 2$$

$$\leftarrow \mu_1 + \mu_2$$

R Code for “by hand” calculation

```
> # Calculate mean and standard deviation for the two groups.  
> # First subset the data.  
> data76<-subset(case0201, Year==1976, select=Depth)  
> head(data76, 5) # Look at the first 5 rows.
```

```
  Depth  
1   6.2  
2   6.8  
3   7.1  
4   7.1  
5   7.4
```

```
> data78<-subset(case0201, Year==1978, select=Depth)  
> head(data78, 5)
```

```
  Depth  
90  7.1  
91  7.9  
92  8.0  
93  8.4  
94  8.4
```

Create 2 data frames, one for each year

```
# Calculate means and standard deviations.
```

```
> mean(data76$Depth)  
[1] 9.469663  
>  
> sd(data76$Depth)  
[1] 1.035331  
>  
> mean(data78$Depth)  
[1] 10.1382  
>  
> sd(data78$Depth)  
[1] 0.9064799
```

shown in Disp. 2.1

```

> # Sample sizes?
> nrow(data76)
[1] 89
>
> nrow(data78)
[1] 89

```

$$\} \quad n_1 = n_2 = 89$$

```

> # Alternative somewhat slicker calculations
> with(case0201, aggregate(Depth~Year, FUN=mean))
  Year      Depth
1 1976  9.469663
2 1978 10.138202

```

```
> with(case0201, aggregate(Depth~Year, FUN=sd))
```

Year	Depth
1 1976	1.0353309
2 1978	0.9064799

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

watch parentheses

```
> # Pooled sd:
```

```
> sqrt(((89-1)*1.0353309^2 + (89-1)*0.9064799^2) / (89+89-2))
[1] 0.9730406
```

$$t\text{-stat} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

```
> # t-stat "by hand"
```

```
> (9.469663 - 10.138202) / (0.9730406 * sqrt(1/89 + 1/89))
[1] -4.583274
```

$$t\text{-stat} = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

```
> # Two-sample t-test in R:  
> t.test(Depth~Year,  
+         data=case0201,  
+         var.equal = TRUE,    # By default, var.equal is FALSE.  
+         alternative="less")
```

Two Sample t-test

data: Depth by Year

t = -4.5833, df = 176, p-value = 4.325e-06

↑ 89 + 89 - 2

≈ 0.0000043

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -0.4273433

one-sided CI.
we'll do a 2-sided
test to get
a 2-sided CI.

sample estimates:

mean in group 1976 mean in group 1978

9.469663

10.138202

Stat Conclusion for hypothesis test:
There is very strong evidence
mean break depth in 1978 is larger
than in 1976 (p-value < 0.0001,
one-sided t-test).

```
> t.test(Depth~Year, data=case0201, var.equal = TRUE)
```

Two Sample t-test

data: Depth by Year

t = -4.5833, df = 176, p-value = 8.65e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.9564088 -0.3806698

sample estimates:

mean in group 1976 mean in group 1978

9.469663

10.138202

Statistical Conclusion:

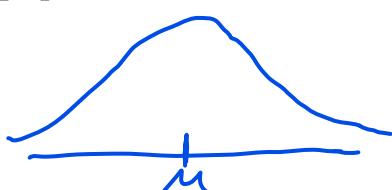
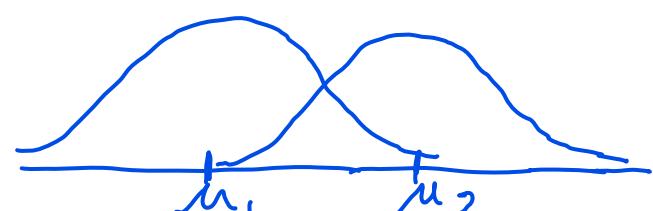
(for CI)

We estimate increase in mean beak depth between 1976 + 1978 is 0.38 to 0.96 mm (95% CI).

2-sided test
↓

0 is not in CI because 0 is not a plausible value for diff. in pop means.

Comparing 1-sample and 2-sample t-tests

	1-sample t-test	2-sample t-test
Populations	One normally-distributed population 	Two normally-distributed populations 
Parameters	Population mean μ Population standard deviation σ	Population means μ_1, μ_2 Population standard deviations σ (until ch. 4)
Samples	Take a random sample of size n from the population	Take a random sample of size n_i from each population ($i = 1, 2$).
Statistics	Sample means: \bar{Y} Sample standard deviations: s	\bar{Y}_1, \bar{Y}_2 s_1, s_2
Parameter Estimates	\bar{Y} estimates μ s estimates σ	\bar{Y}_1 estimates μ_1 , \bar{Y}_2 " μ_2 Both s_1, s_2 est. σ so pool: $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

Comparing 1-sample and 2-sample t-tests—continued

	1-sample t-test	2-sample t-test
Null Hypothesis	$H_0: \mu = 0$	$H_0: \mu_1 - \mu_2 = 0$
Test Statistic	$\frac{\text{pt est} - \text{value under } H_0}{\text{SE(pt est)}}$ $\frac{\bar{Y} - 0}{S/\sqrt{n}}$	$\frac{\text{pt est} - \text{value under } H_0}{\text{SE(pt est)}}$ $\frac{\bar{Y}_1 - \bar{Y}_2 - 0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$
Sampling Distribution of t-stat under H_0	t dist'n with $df = n - 1$ = sample size minus # mean parameters	t dist'n with $df = n_1 + n_2 - 2$
Confidence Interval for parameter of interest	$\text{pt est} \pm t_{df}(1 - \alpha/2)\text{SE(pt est)}$ $\bar{Y} \pm t_{n-1}(1-\alpha/2) \cdot \frac{S/\sqrt{n}}{(1-\alpha/2)}$ $\alpha = 0.05 \leftrightarrow 95\% \text{ CI}$	$\text{pt est} \pm t_{df}(1 - \alpha/2)\text{SE(pt est)}$ $\bar{Y}_1 - \bar{Y}_2 \pm t_{n_1+n_2-2}(1-\alpha/2) \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$