Stat 411/511

Paired T-test & Two Sample model

Oct 5 2015

Charlotte Wickham

Three ways to run code

Code -> Run lines (crtl + enter (windows), ⌘ + return (mac))

    run selected lines in **current** R session

    order matters!

Code -> Source document

    runs all code top to bottom in **current** R session

File -> Compile Notebook

    runs all code top to bottom in a **fresh** R session

    adds code and output to annotated document

My Workflow

Project   a directory I keep all my code and data for this class

a new R script
01-hw1.R

write and run code
line by line

Is the code OK?
Restart R
Source R script

Is documentation OK?
Add annotation
Compile Notebook

Today

Paired t-test

Example, from the Sleuth

Doing the hard work in R

When a one-sample t-test is a paired t-test

A **paired *t*-test** is just a one sample t-test on the **differences** in two responses from paired data.
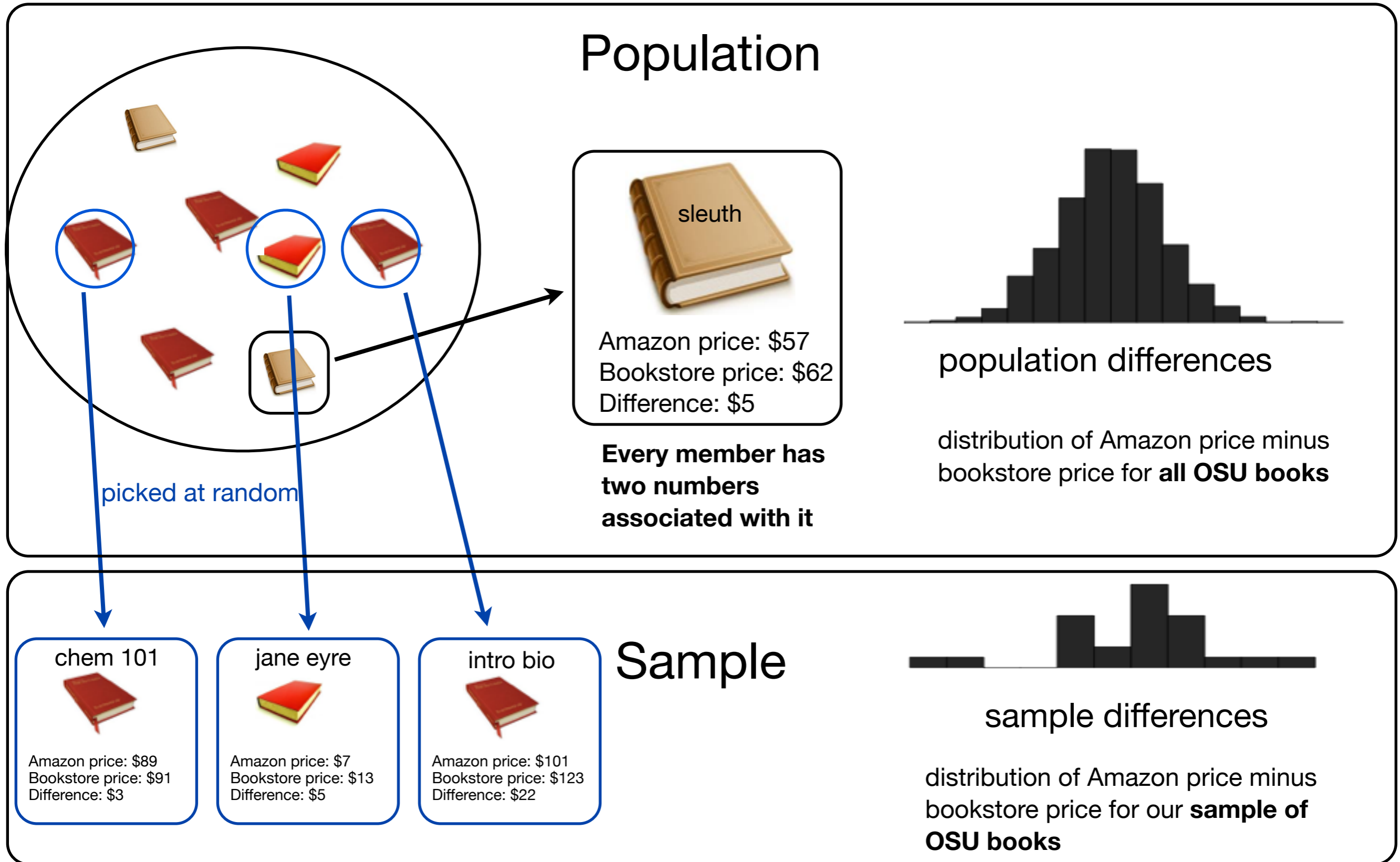
Today,

one response was the price on amazon

the other response was the price at the bookstore

these prices came in pairs (each textbook gave rise to one Amazon price and one bookstore price)

we did a one sample t-test on the differences in the two prices.

We did a **paired** t-test.

Paired data sampling model

Population

picked at random

sleuth

Amazon price: $57
Bookstore price: $62
Difference: $5

Every member has two numbers associated with it

population differences

distribution of Amazon price minus bookstore price for **all OSU books**

Sample

chem 101

Amazon price: $89
Bookstore price: $91
Difference: $3

jane eyre

Amazon price: $7
Bookstore price: $13
Difference: $5

intro bio

Amazon price: $101
Bookstore price: $123
Difference: $22

sample differences

distribution of Amazon price minus bookstore price for our **sample of OSU books**

Identifying paired data

Look for something that links an observation in one group to an observation in the other.

I.e. the textbook links an Amazon price to bookstore price

Ask yourself: if I had one observation, could I easily (and uniquely) identify another in the other group that it "pairs" with?

Your turn

Paired or not?

We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days for Intel's stock and another random sample of 50 days for Southwest's stock.

We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.

A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school.

To evaluate the general fitness level of OSU students, we randomly sample 50 students and get two measurements: their heart rate after 10 minutes of jump rope, and their heart rate again 2 minutes later.

The Schizophrenia case study

Research question: are there physiological differences associated with schizophrenia?

15 identical twins, one affected by schizophrenia.

We have measurements on left hippocampus volume.

Naturally **paired**. Look at the **differences** between twins.

```
library(Sleuth3)
case0202
diffs <- case0202$Unaffected - case0202$Affected
```

Two questions
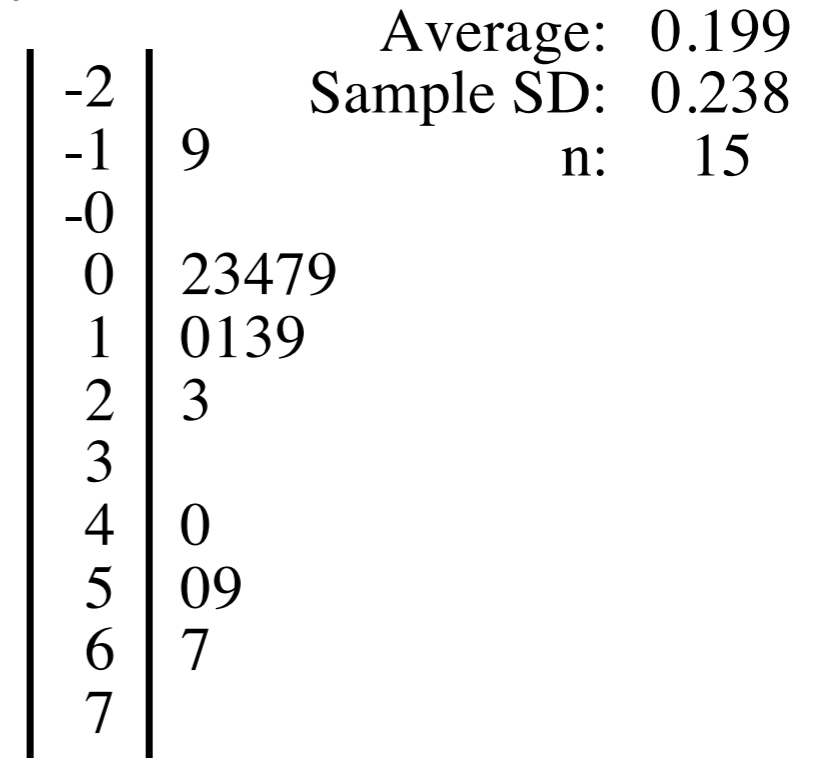
- Is the mean difference in volume, μ, different from zero?

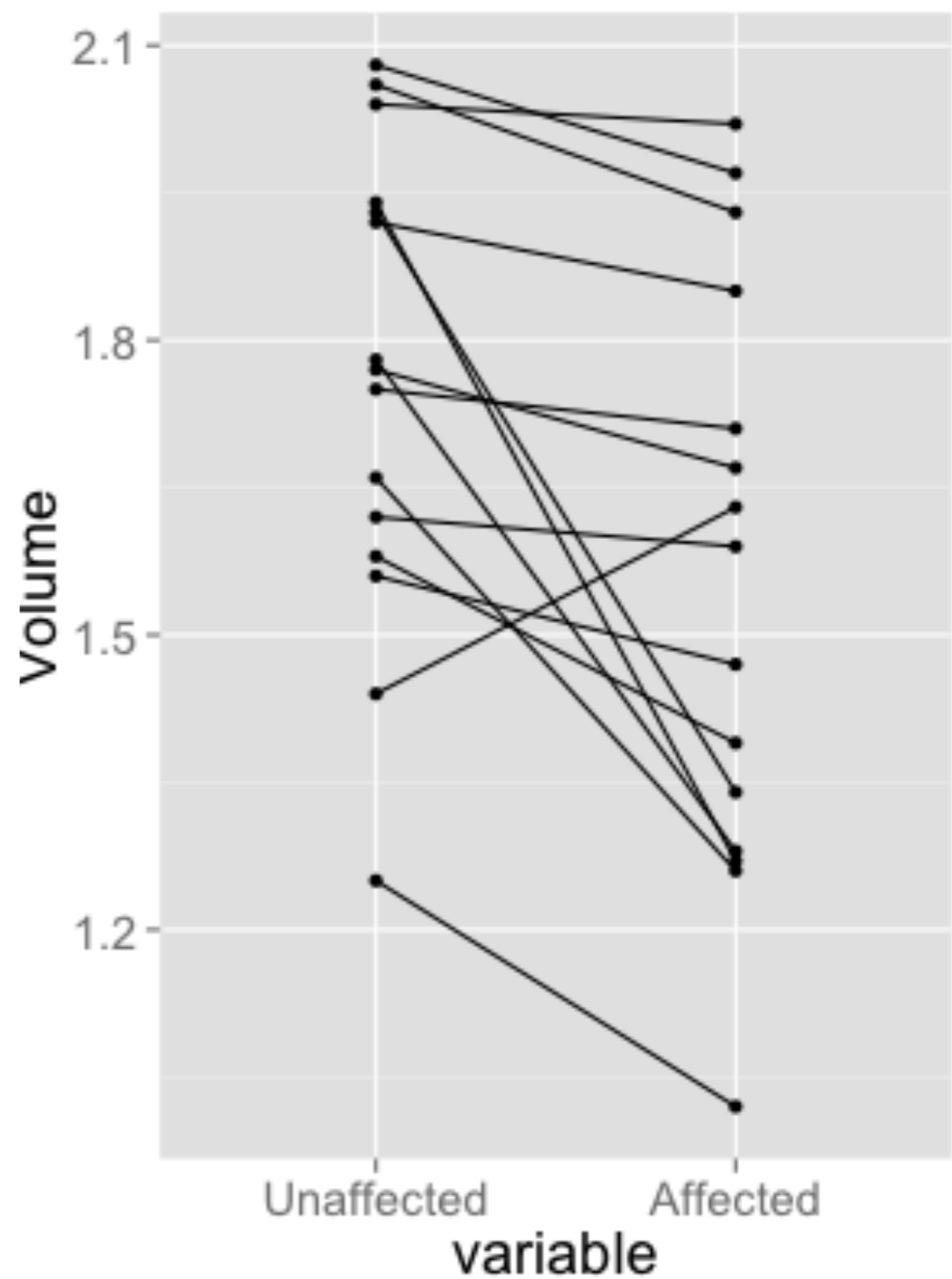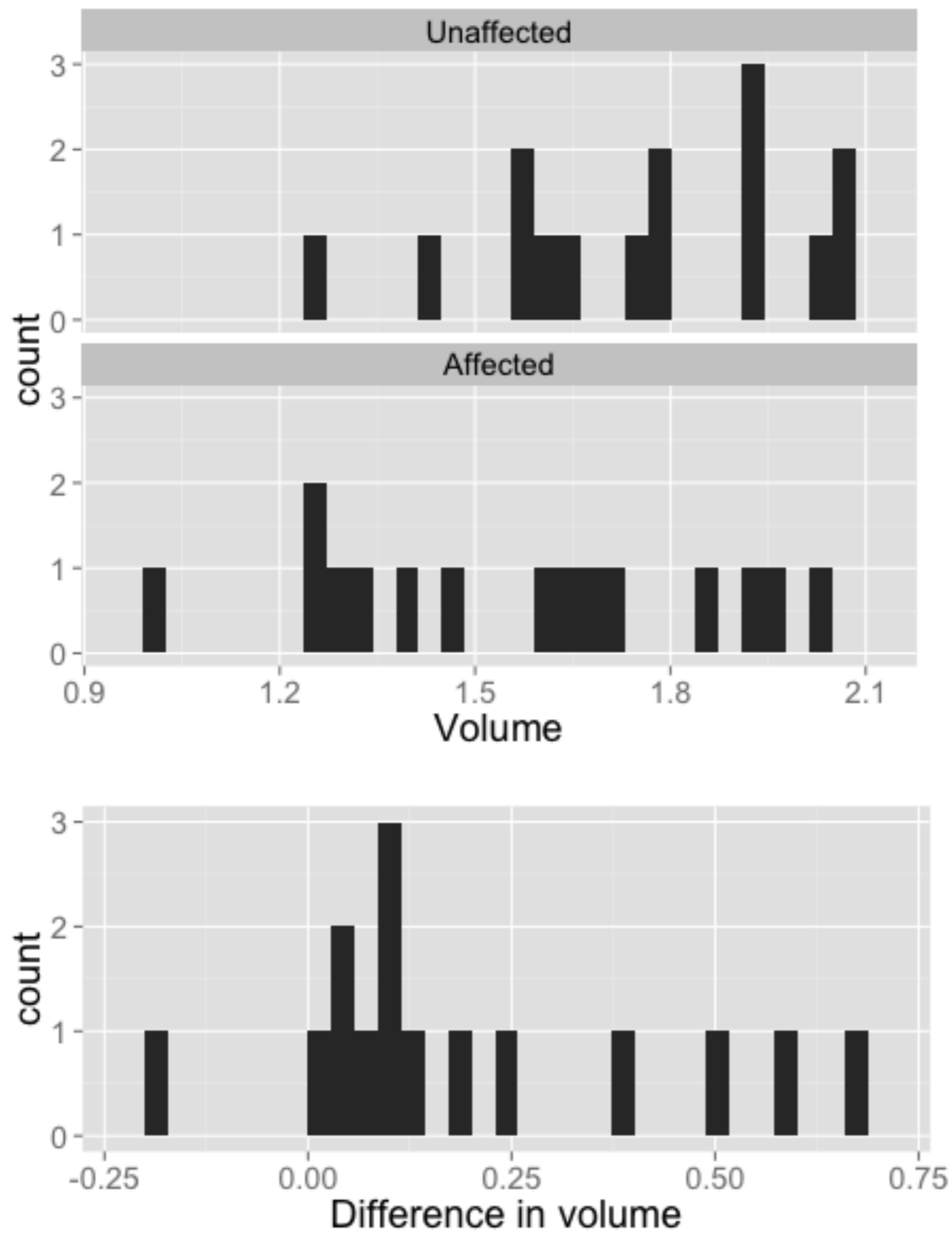- What's a likely range for the mean difference in volume, μ?

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

Pair #	Unaffected	Affected	Difference
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	1.58	1.39	0.19
5	2.06	1.93	0.13
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
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

Differences

Average: 0.199
Sample SD: 0.238
n: 15

```
-2 |
-1 | 9
-0 |
 0 | 23479
 1 | 0139
 2 | 3
 3 |
 4 | 0
 5 | 09
 6 | 7
 7 |
```

Legend: | 6 | 7 represents 0.67 cm^3

Calculations for the paired t-test and 95% confidence interval for the schizophrenia study

1   *Compute differences. Obtain their average, $\overline{Y}$, and standard deviation, s*

Y_i is the volume for the unaffected twin minus the volume for the schizophrenic twin (i=1,...,15)   replace X with Y, in previous slides

Sample average: $\overline{Y}$ = .199 (cm³)
Sample standard deviation: s = .238 (cm³); 14 d.f.

2   *Compute S.E.($\overline{Y}$) = s/$\sqrt{n}$ and d.f. = n-1*

S.E.($\overline{Y}$) = .238/$\sqrt{15}$ = .0615 (cm³)

3   *Paired t-test for the hypothesis that the population mean difference is zero*

Compute the t-statistic for this hypothesis: ($\overline{Y}$-0)/S.E.($\overline{Y}$)

t-statistic = .199/.0615 = 3.236

Find the p-value (2-sided here) as the proportion of values in a t_{n-1} distribution as far or farther from zero than the observed t-statistic

two-sided p-value = .006 (from computer or Table A.2 and interpolation)

4   *95% confidence interval for population mean difference*

Find the 97.5th percentile from the t-distribution on n-1 degrees of freedom

t_{14}(.975) = 2.145 (from Table A.2)

95% confidence interval = $\overline{Y}$ ± $t_{(n-1)}$(.975)✕S.E.($\overline{Y}$)

**.199cm³ ± 2.145 × 0.615cm³ :
0.067cm³ to 0.331cm³**

t.test(diffs)

```
> diffs <- case0202$Unaffected - case0202$Affected
> t.test(diffs)

	One Sample t-test

data:  diffs
t = 3.2289, df = 14, p-value = 0.006062
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.06670409 0.33062922
sample estimates:
mean of x
0.1986667
```

Alternative syntax

```
t.test(case0202$Unaffected,
   case0202$Affected, paired = TRUE)

with(case0202, t.test(Unaffected,
   Affected, paired = TRUE))
```

Your Turn:

Match the symbols to output

t.test(diffs)

$$\overline{X} \qquad \dfrac{\overline{X} - 0}{\mathrm{SE}_{\overline{X}}} \qquad n-1 \qquad \overline{X} \pm t_{n-1}(0.975) \times \mathrm{SE}_{\overline{X}}$$

```
> diffs <- case0202$Unaffected - case0202$Affected
> t.test(diffs)

    One Sample t-test

data:  diffs
t = 3.2289, df = 14, p-value = 0.006062
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.06670409 0.33062922
sample estimates:
mean of x
0.1986667
```

There is **convincing** evidence the **mean difference** in brain volume between the person affected by schizophrenia and their twin unaffected by schizophrenia **is not zero** (two sided p-value < 0.006, paired t-test).

amount of evidence

We estimate the hippocampus in the twin affected by schizophrenia is 0.2 cm^3 smaller **on average** than in the twin unaffected by schizophrenia. point estimate

With 95% confidence, the hippocampus in the twin affected by schizophrenia is between is 0.07cm^3 and 0.33 cm^3 smaller **on average** than in the twin unaffected by schizophrenia. confidence interval
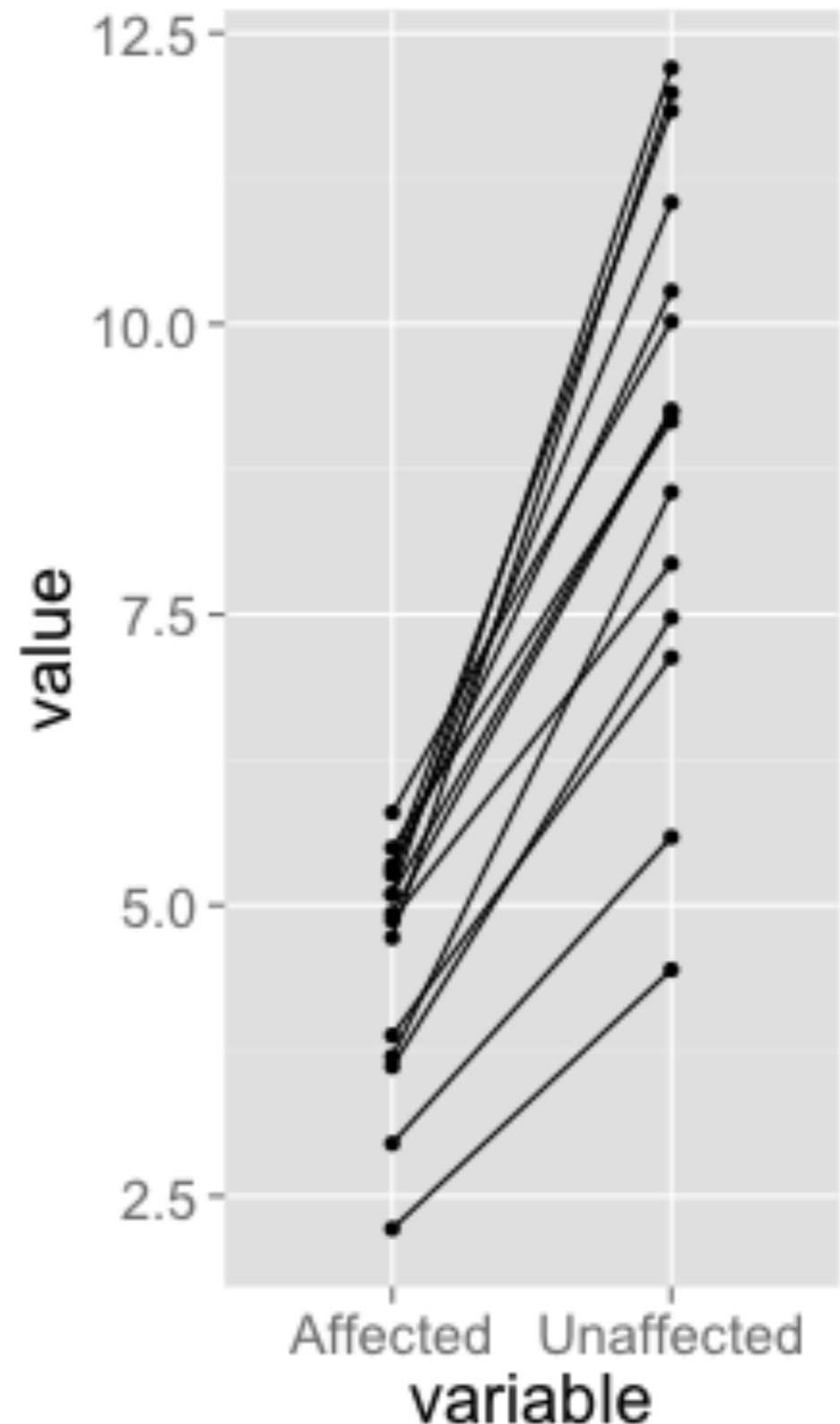
Best practice: always give a confidence interval even if there is no evidence of a difference.

When is taking differences misleading?

If the difference in any individual depends on the magnitude of the responses.

E.g. you get bigger differences in price with more expensive textbooks.

What does that look like?

So far, we have had **one** population of interest.

Next week, we'll talk about having **two** populations of interest.