Working with Experimental Data

22 January 2020

Modern Research Methods

Course Website: https://cumulativescience.netlify.com/

Last Time: Intro to R

Interacting with R with an Rmarkdown notebook in Rstudio (Cloud)

```
---
title: "Assignment 0: Intro to R and RStudio"
subtitle: "Modern Research Methods"
author: "Molly Lewis"
date: "`r format(Sys.time(), '%d %B %Y')`"
output:
  html_document:
    highlight: kate
    theme: cosmo
---
```

Header (YAML)

Reproducible Research

At the click of a button, or the type of a command, you can rerun the code in an R Markdown file to reproduce your work and export the results as a finished report.

```
```{r}
365 + 112
```
```

R chunks

```
The mean number of boy baptisms in a given year is `r my_mean_boys`.
```

Plain text (with inline R)

.Rmd -> “Knit” -> .html

Rmarkdown – formatting text

```
Plain text

End a line with two spaces
to start a new paragraph.

*italics* and **bold**

`verbatim code`

sub/superscript^2^~2~

~~strikethrough~~

escaped: \* \_ \\

endash: --, emdash: ---

equation: $A = \pi*r^{2}$

equation block:

$$E = mc^{2}$$

> block quote

# Header1  {#anchor}

## Header 2  {#css_id}

### Header 3  {.css_class}

#### Header 4
```

Plain text

End a line with two spaces to start a new paragraph.

italics and **bold**

`verbatim code`

sub/superscript$^2_2$

~~strikethrough~~

escaped: * _ \

endash: –, emdash: —

equation: $A = \pi * r^2$

equation block:

$$E = mc^2$$

> block quote

Header1

Header 2

```
<http://www.rstudio.com>

[link](www.rstudio.com)

Jump to [Header 1](#anchor)

image:

![Caption](smallorb.png)

* unordered list
    + sub-item 1
    + sub-item 2
        - sub-sub-item 1
```

http://www.rstudio.com

link

Jump to Header 1

image:

Caption

- unordered list
    - sub-item 1
    - sub-item 2

```
:   Definition 1

| Right | Left | Default | Center |
|------:|:-----|---------|:------:|
|   12  |  12  |    12   |    12  |
|  123  |  123 |   123   |   123  |
|    1  |    1 |     1   |     1  |

- slide bullet 1
- slide bullet 2

(>- to have bullets appear on click)

horizontal rule/slide break:

***

A footnote [^1]
```

Right	Left	Default	Center
12	12	12	12
123	123	123	123
1	1	1	1

- slide bullet 1
- slide bullet 2

(>- to have bullets appear on click)

horizontal rule/slide break:

A footnote [1]

[Cheatsheat]

Rmarkdown – setting "options" for R chunks

Useful options:

- `message` – display messages in knitted document?
- `warning` – display warnings?
- `include` – include chunk after running?
- `eval` – run code?

````
**Load Arbuthnot data**
```{r, message = TRUE}
arbuthnot <- read_csv("data/arbuthnot.csv")
```
````

Load Arbuthnot data

```
arbuthnot <- read_csv("data/arbuthnot.csv")
```

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   boys = col_double(),
##   girls = col_double()
## )
```

````
**Load Arbuthnot data**
```{r, message = FALSE}
arbuthnot <- read_csv("data/arbuthnot.csv")
```
````

Load Arbuthnot data

```
arbuthnot <- read_csv("data/arbuthnot.csv")
```

Can specify "global" option (all chunks) - `knitr::opts_chunk$set()`

```
knitr::opts_chunk$set(eval = TRUE, message = FALSE)
```

Useful to label chunks

````
```{r calculate_sum}
365 + 112
```
````

R Syntax

Variable assignment: = or <-

Data types:

- Numeric – e.g., 1, 2, 3, 6.2
- Character string – e.g., "apple"
- Logical - e.g., TRUE/FALSE

Data Structures:

- Vector
- Data frame/tibble

Functions

```
function_name <- function(var){

   Do something

   return(new_variable)
}
```

R for Python programmers: http://ramnathv.github.io/pycon2014-r/

Data

- **Variable** – unique measurement or quantity
  - e.g., temperature, mood, attendance, # of books owned, reaction time, color
- Types of variables:
  - **Qualitative** – describe quality (no intrinsic ordering)
  - **Quantitative** – describe quantity
    - **Binary** – 1 or 0 (or, TRUE or FALSE)
    - **Integers** – whole numbers
    - **Real numbers** – have fractional/decimal part
- **Observation**– Smallest unit you have data about (person, trial in an experiment, city, school)
- **Value** – Quantity/quality associated with a particular variable and observation

Variable – Jan. high temp. in PGH
Observation – day
Value – 38 degrees

Variable – age of students in MRM
Observation – student
Value – 19.3 years

Variable – Condition
Observation – person
Value – tennis class or chess class

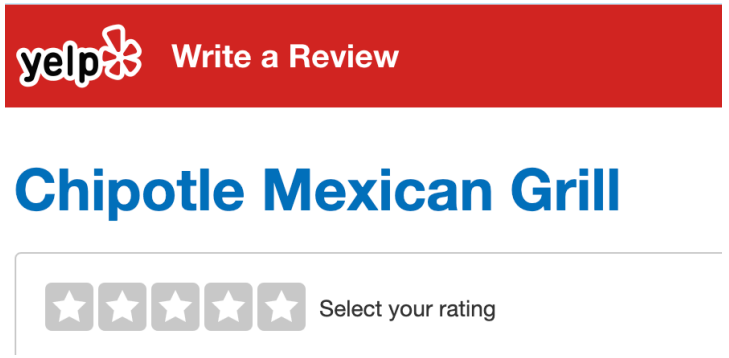
Discrete vs. continuous measurement

Discrete – measurement can only take one of a set of values

- Days of the week, dog breeds, # of children, # of Twitter followers, # of Yelp stars
- No "middle ground"

Continuous – measurement that is real number, and could take one of any range of values

- Most quantitative variables (reaction time, judgement on slider scale)
- Limited by precision of instrument



CONTINUOUS

measured data, can have ∞ values within possible range.

I AM 3.1" TALL

I WEIGH 34.16 grams

DISCRETE

OBSERVATIONS CAN ONLY EXIST AT LIMITED VALUES, OFTEN COUNTS.

I HAVE 8 LEGS and 4 SPOTS!

@allison_horst

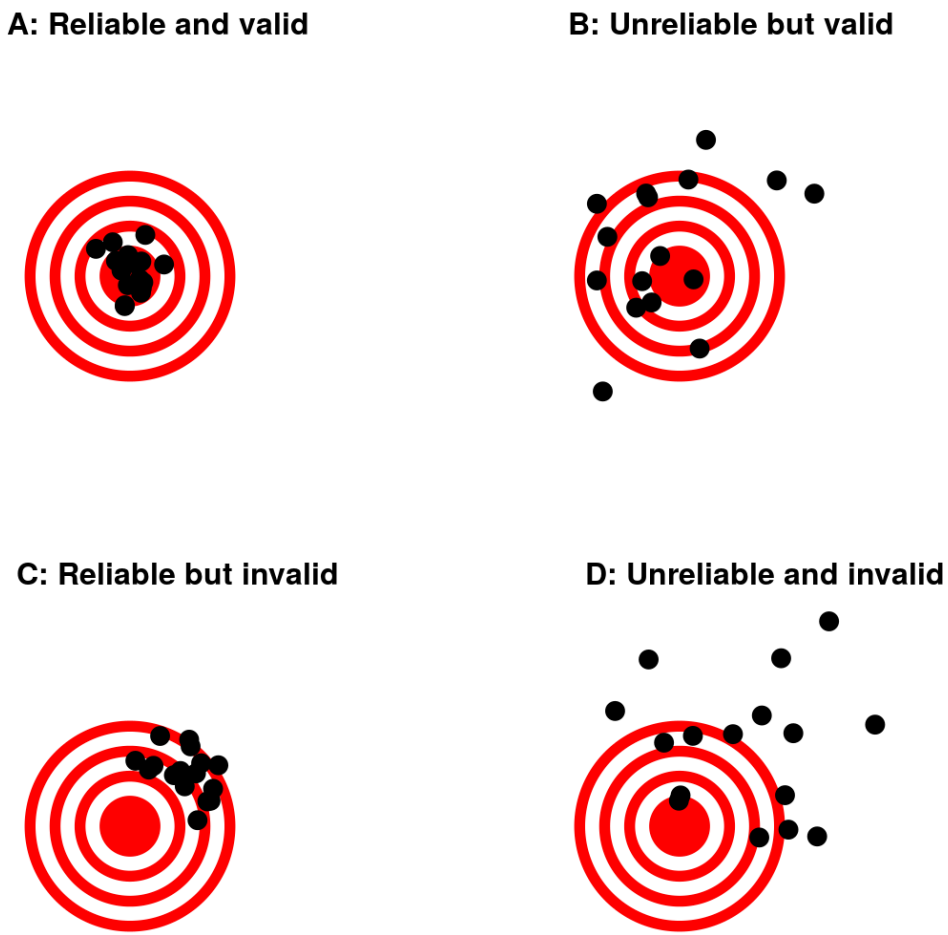
Artwork by @allison_horst

What makes a good measurement?

Reliability – Consistency of measurement (Do you say the same thing today as yesterday?)

Validity – Are we measuring the construct we want to measure?

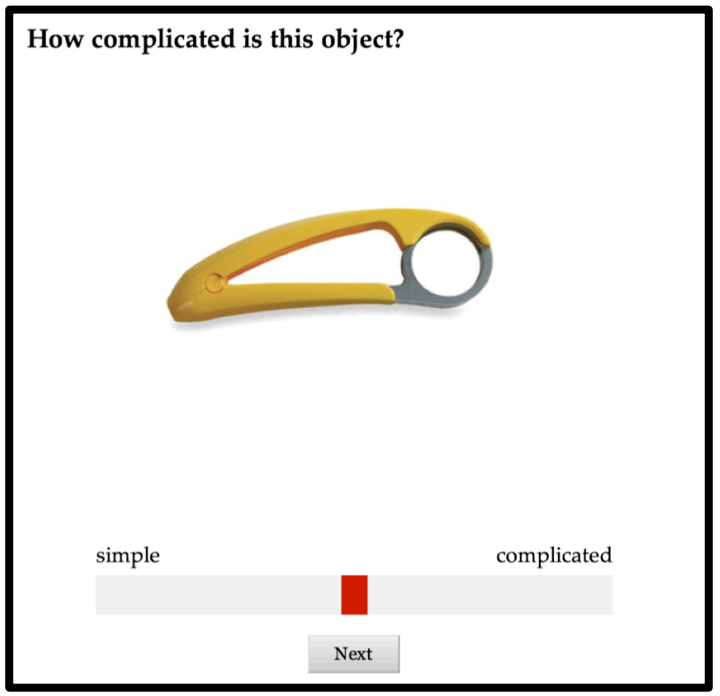
Suppose the thing we're trying to measure is the center of the bullseye…



(Poldrack, 2019)

Example: Measuring "Conceptual Complexity"

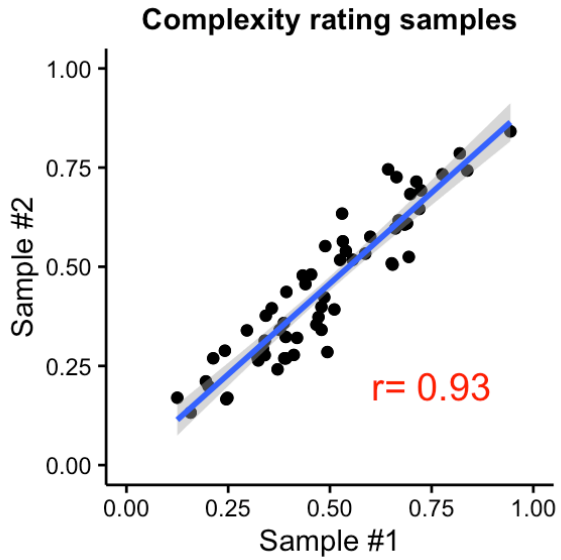
table vs. democracy



0 1

N Sample 1 = 60 participants

N Sample 2 = 60 participants



Evaluate the reliability and validity of this measure.

(Lewis & Frank, 2016)