Statistical Foundations: Sampling

17 February 2020

Modern Research Methods

The Single Experiment

Population

Question

Hypothesis

Exp. Design

Experimenter

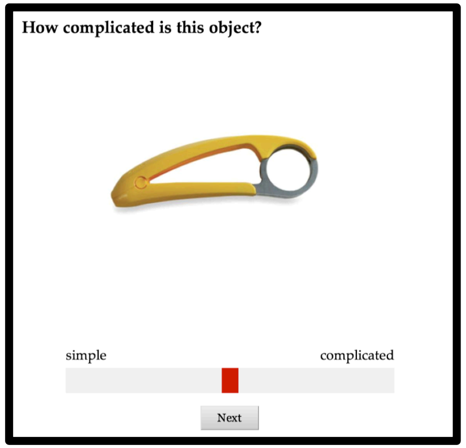
Data

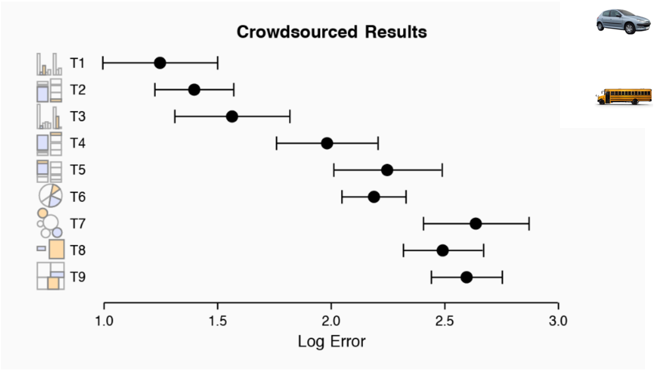
Analyst

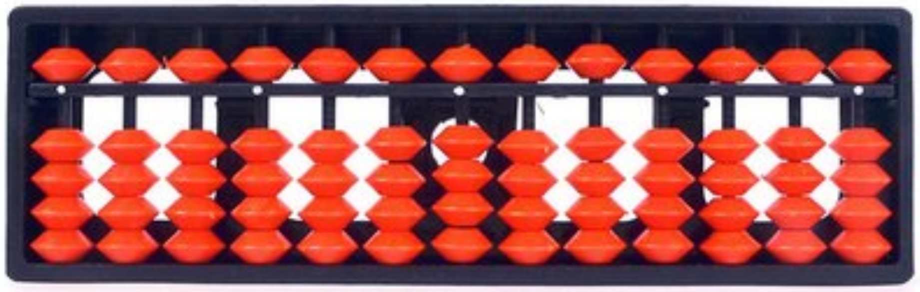
Code

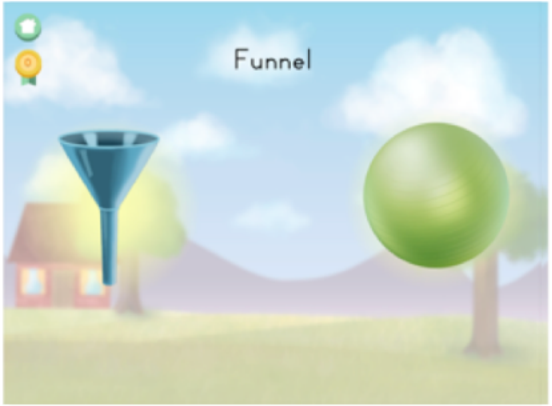
Estimate

Claim









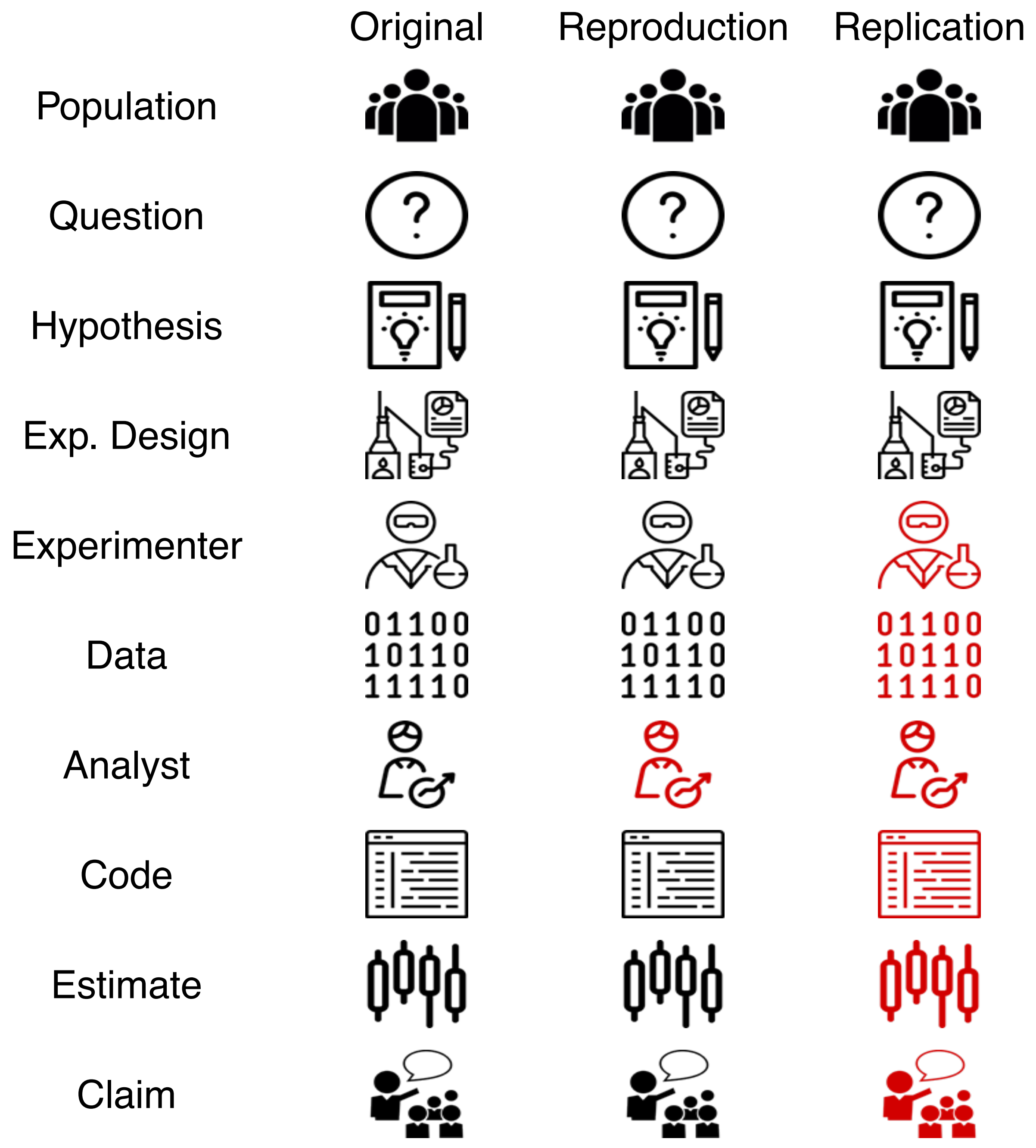
Overview of course

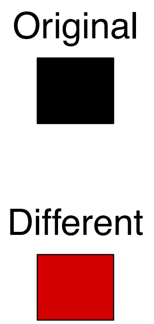
1) **Philosophy of Cumulative Science**

2) **The Single Experiment** – Experimental data, tools in R for working with data and plotting data, reproducibility

3) **Repeating an Experiment** – Intro to statistical concepts, replication of experiments

4) **Aggregating Many Experiments** – Meta-analysis





REPRODUCE = Get same result from same dataset.

REPLICATE = Get same result with a new dataset

* Sometimes people are sloppy with these terms and use them interchangeably.

(Patil, Peng, & Leek, 2019)

Contents lists available at ScienceDirect

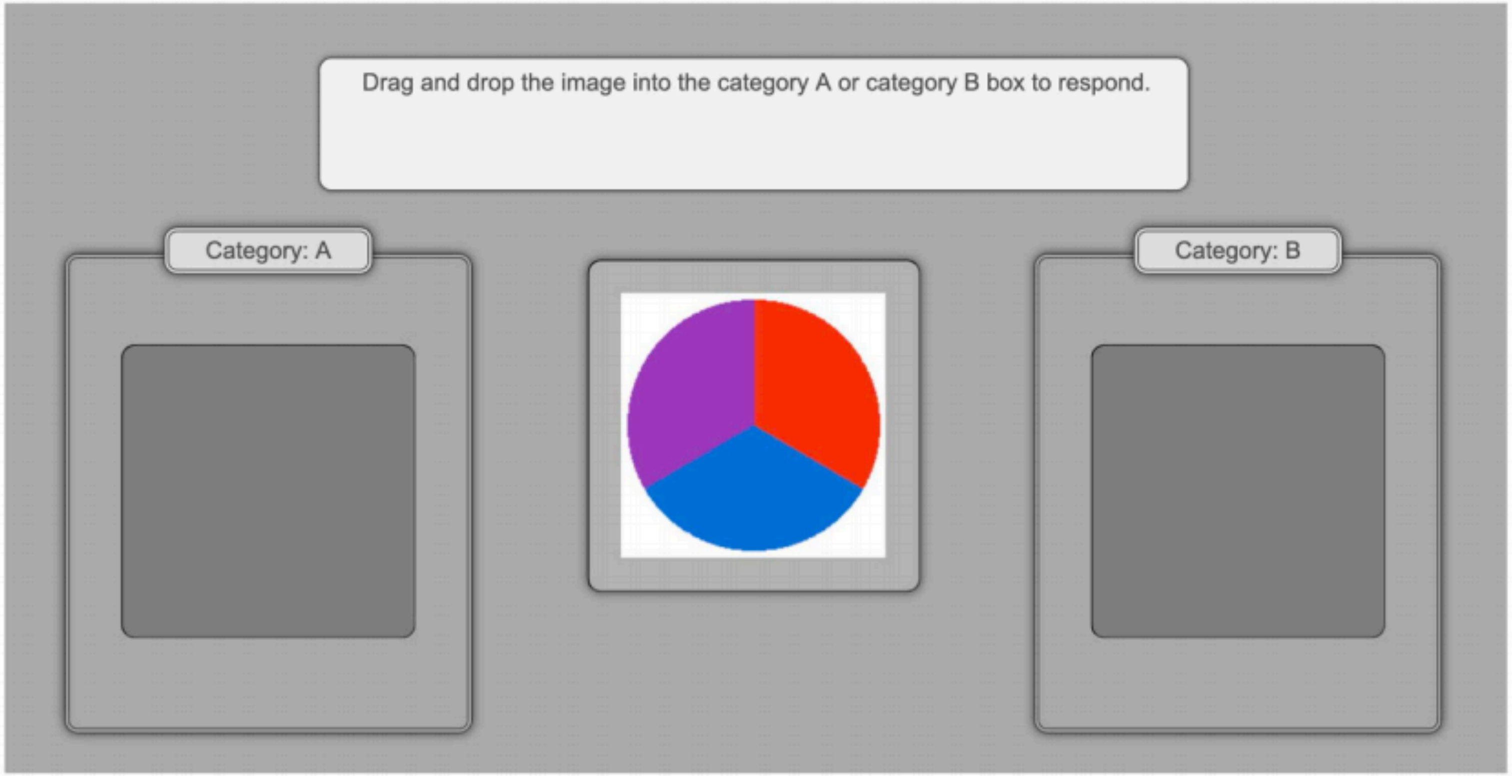
Cognition

journal homepage: www.elsevier.com/locate/cognit

Finding categories through words: More nameable features improve category learning

Martin Zettersten*, Gary Lupyan

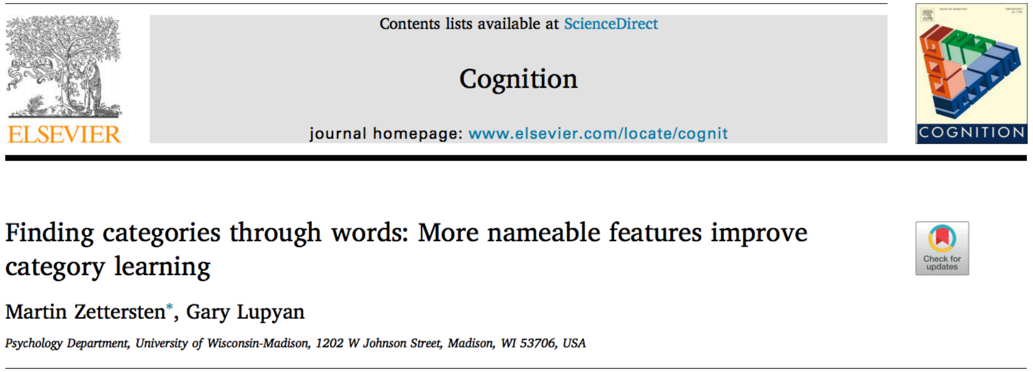
Psychology Department, University of Wisconsin-Madison, 1202 W Johnson Street, Madison, WI 53706, USA



High nameability 

Low nameability 

Replication vs. Reproduction


Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

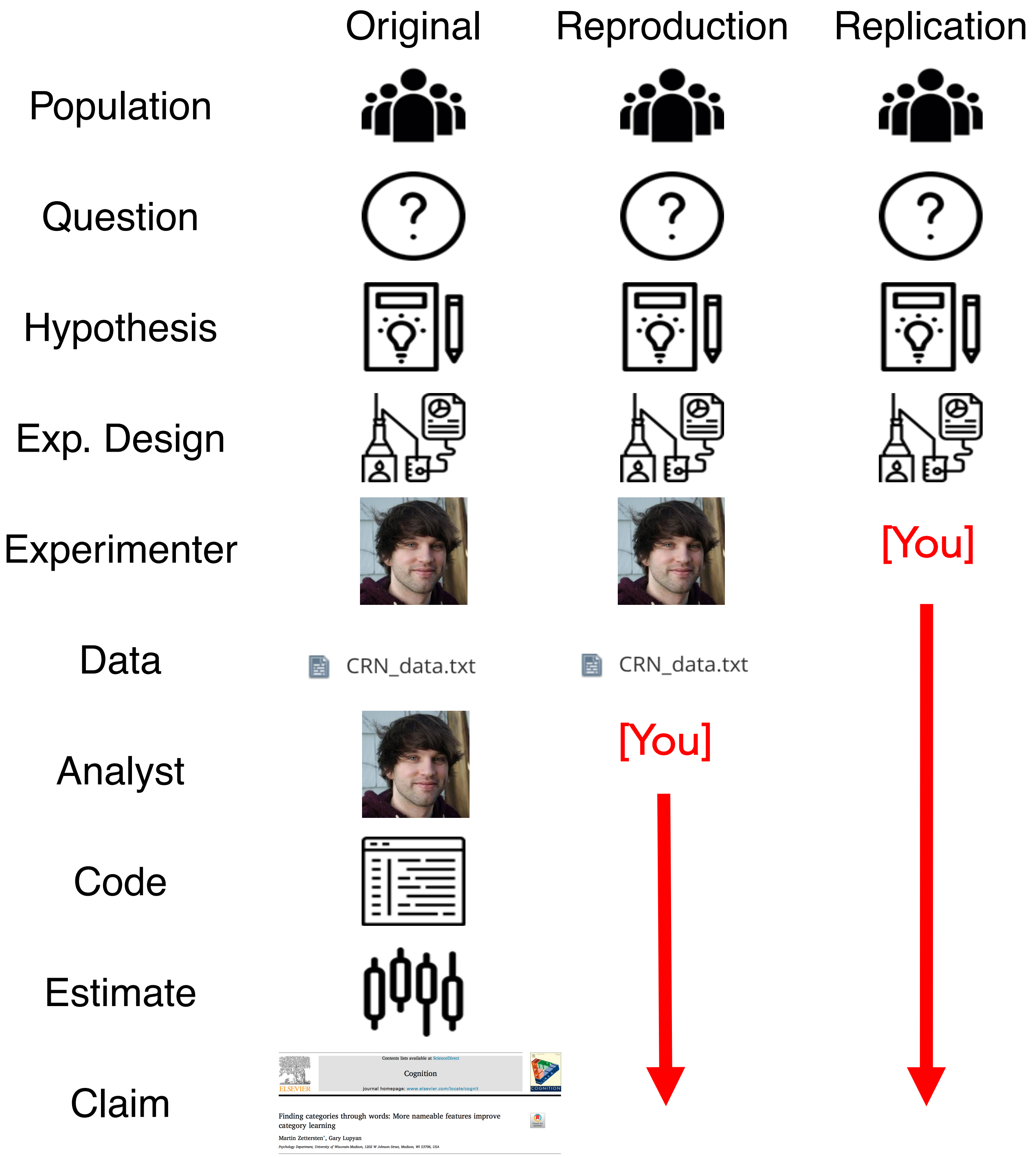
ELSEVIER

COGNITION

Finding categories through words: More nameable features improve category learning

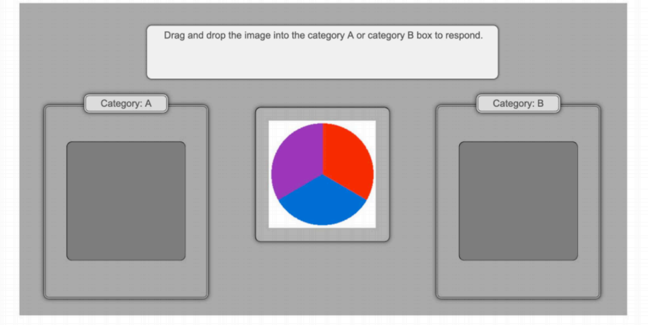
Martin Zettersten*, Gary Lupyan

Psychology Department, University of Wisconsin-Madison, 1202 W Johnson Street, Madison, WI 53706, USA



Replicating Zettersten and Lupyan (2020)

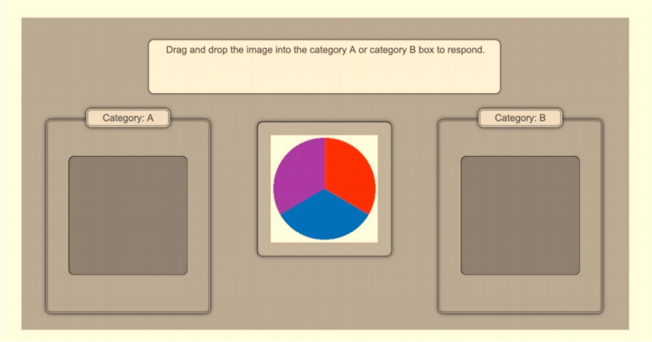
Original



predicting participants' trial-by-trial accuracy on training trials from condition, including a by-subject random intercept.[3] We used the lme4 package version 1.1-21 in R (version 3.6.1) to fit all models (D. Bates & Maechler, 2009; R Development Core Team, 2019). Participants in the High Nameability condition (M = 84.0%, 95% CI = [78.6%, 89.4%]) were more accurate than participants in the Low Nameability Condition (M = 67.7%, 95% CI = [59.9%, 75.4%]), b = 1.02, 95% Wald

Replication

[You]



High Nameability Condition = 75%
Low Nameability Condition = 69%

Should you expect to replicate the original finding? Did you replicate it? What would convince you? Discuss with a partner.

predicting participants' trial-by-trial accuracy on training trials from condition, including a by-subject random intercept.[3] We used the lme4 package version 1.1-21 in R (version 3.6.1) to fit all models (D. Bates & Maechler, 2009; R Development Core Team, 2019). Participants in the High Nameability condition (M = 84.0%, 95% CI = [78.6%, 89.4%]) were more accurate than participants in the Low Nameability Condition (M = 67.7%, 95% CI = [59.9%, 75.4%]), b = 1.02, 95% Wald

?=

High Nameability Condition = 75%
Low Nameability Condition = 69%

In order to evaluate this replication, we need think about *sampling*.

In the next few classes, we're going to discuss sampling in order to reason about the replicability of psychological effects.

Reading for today

Chapter 12 Sampling

One of the foundational ideas in statistics is that we can make inferences about an entire population based on a relatively small sample of individuals from that population. In this chapter we will introduce the concept of statistical sampling and discuss why it works.

Anyone living in the United States will be familiar with the concept of sampling from the political polls that have become a central part of our electoral process. In some cases, these polls can be incredibly accurate at predicting the outcomes of elections. The best known example comes from the 2008 and 2012 US Presidential elections, when the pollster Nate Silver correctly predicted electoral outcomes for 49/50 states in 2008 and for all 50 states in 2012. Silver did this by combining data from 21 different polls, which vary in the degree to which they tend to lean towards either the Republican or Democratic side. Each of these polls included data from about 1000 likely voters – meaning that Silver was able to almost perfectly predict the pattern of votes of more than 125 million voters using data from only 21,000 people, along with other knowledge (such as how those states have voted in the past).

Distributions

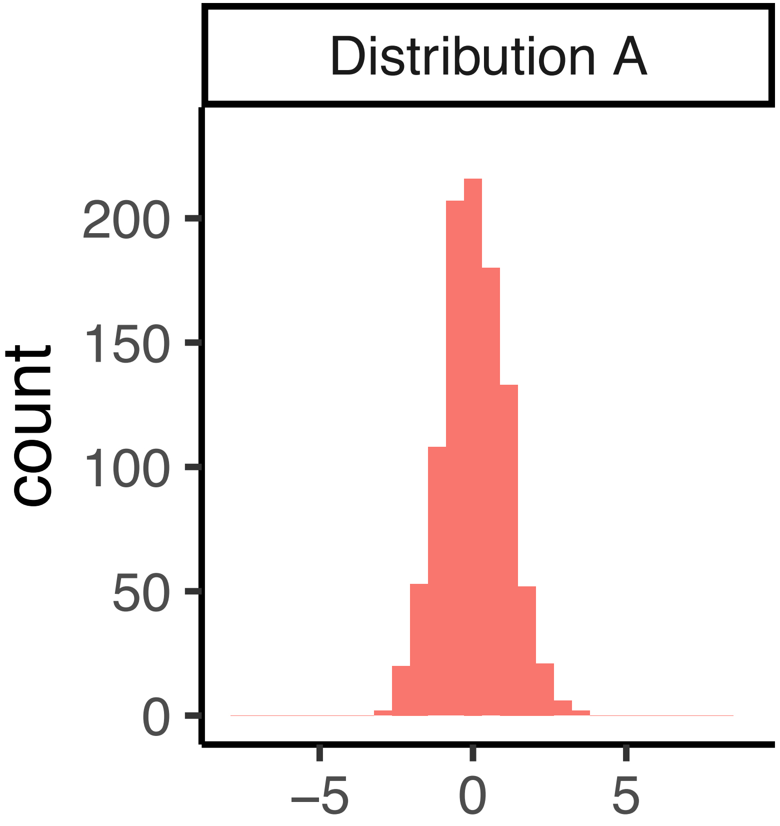
Distributions = counts of a variable

Plot with histograms

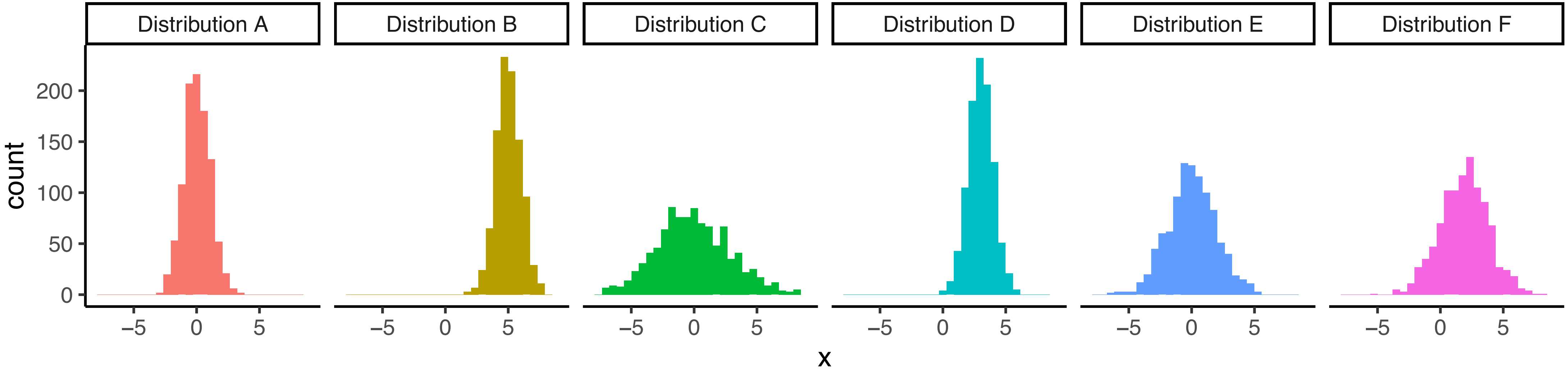
Two measures:

- **Mean** measures center ("central tendency")
- **Variance** measures dispersion.

(There are other measures of the center and dispersion of a distribution, but these are the measures we're going to focus on here)

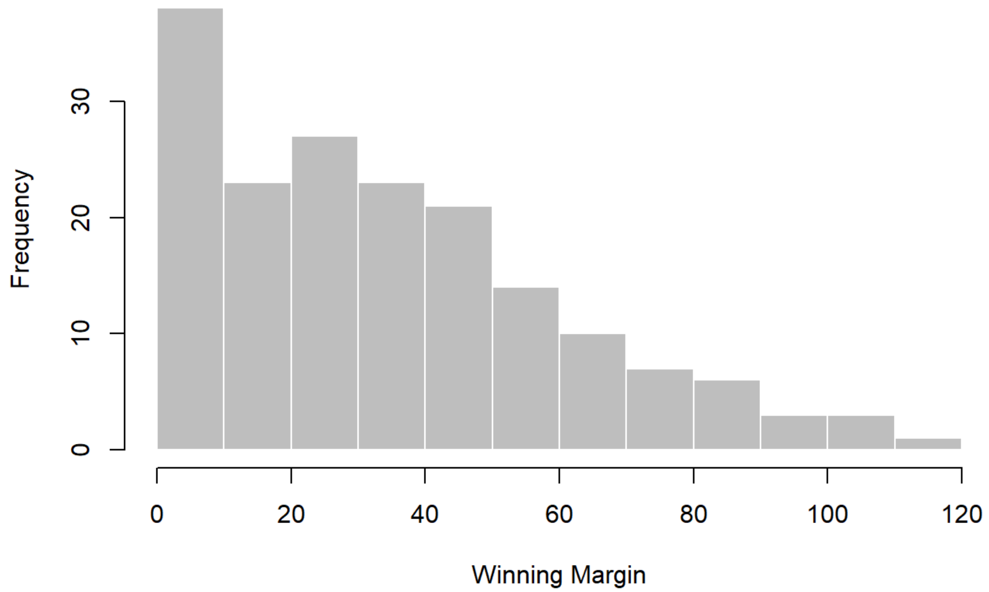


What is the mean of these distributions? Which ones have low vs. high variance?



Mean = 0 Low variance	Mean = 5 Low variance	Mean = 0 V. High variance	Mean = 3 Low variance	Mean = 0 High variance	Mean = 2 High variance

Calculating mean



the observation	its symbol	the observed value
winning margin, game 1	X_1	56 points
winning margin, game 2	X_2	31 points
winning margin, game 3	X_3	56 points
winning margin, game 4	X_4	8 points
winning margin, game 5	X_5	32 points

$$\frac{56 + 31 + 56 + 8 + 32}{5} = \frac{183}{5} = 36.60$$

```
(56 + 31 + 56 + 8 + 32) / 5
```

```
## [1] 36.6
```

```
mean( x = afl.margins )
```

$$\boxed{\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i}$$

(Thanks to Danielle Navarro,
LSR https://learningstatisticswithr.com/)

Calculating variance

$$\mathrm{Var}(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

Variance is the average squared deviation from the mean of a dataset.

Standard deviation is the square root of variance.

i [which game]	X_i [value]	$X_i - \bar{X}$ [deviation from mean]	$(X_i - \bar{X})^2$ [absolute deviation]
1	56	19.4	376.36
2	31	-5.6	31.36
3	56	19.4	376.36
4	8	-28.6	817.96
5	32	-4.6	21.16

```
( 376.36 + 31.36 + 376.36 + 817.96 + 21.16 ) / 5
```

```
## [1] 324.64
```

```
var( afl.margins )
```

Our goal as scientists

- As scientists, we want to **estimate parameters** about the world.
- One of the most common parameters is the mean.
- For example: What is the mean accuracy in the high nameability condition? What is the mean accuracy in the low nameability condition? (Zettersten & Lupyan, 2020)
  - Are the two means different from each other?
- As psychologists we're interested in the population of ALL PEOPLE if they had done our experiment.
- But, to save time and effort, we only measure a **sample.**

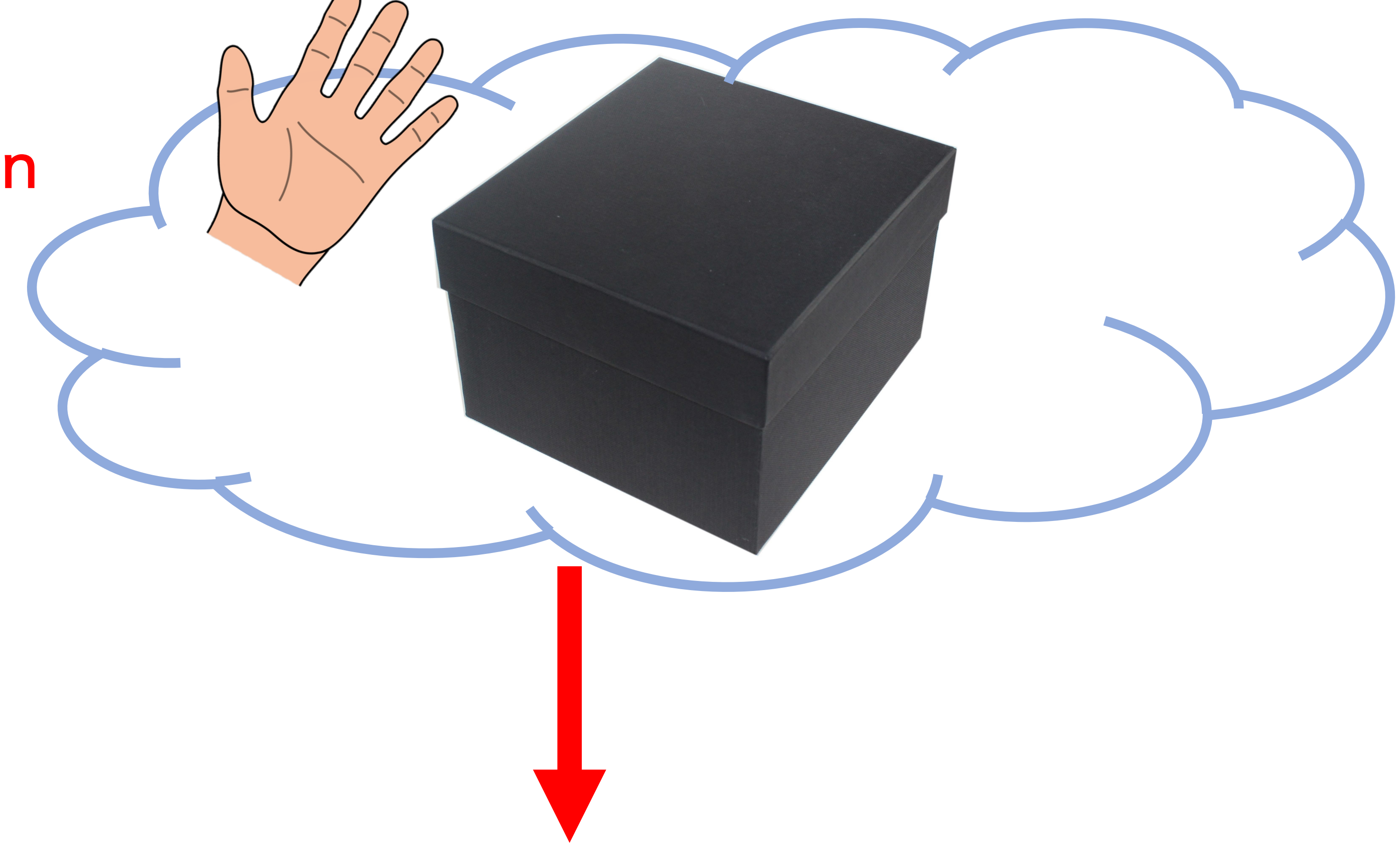
Population vs. sample

- A sample is a random subset of the population.
- That means there are really two distributions.
- **Population**: The distribution of all people (7.53 billion), or maybe all people who speak English (1.5 billion), or maybe all people at UW-Madisoin (44k)
- **Sample**: Zettersten and Lupyan only tested 50 participants.
- Unlike the Zorbia example, we don't know what the population looks like (and we usually don't).

Challenge: Make (good) inferences about the population from the sample.

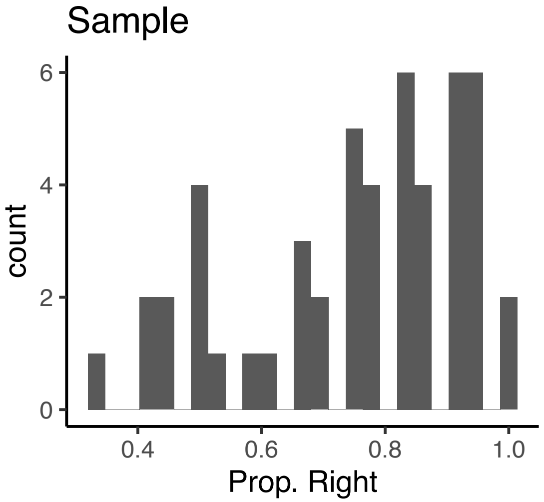
Population

N = a lot

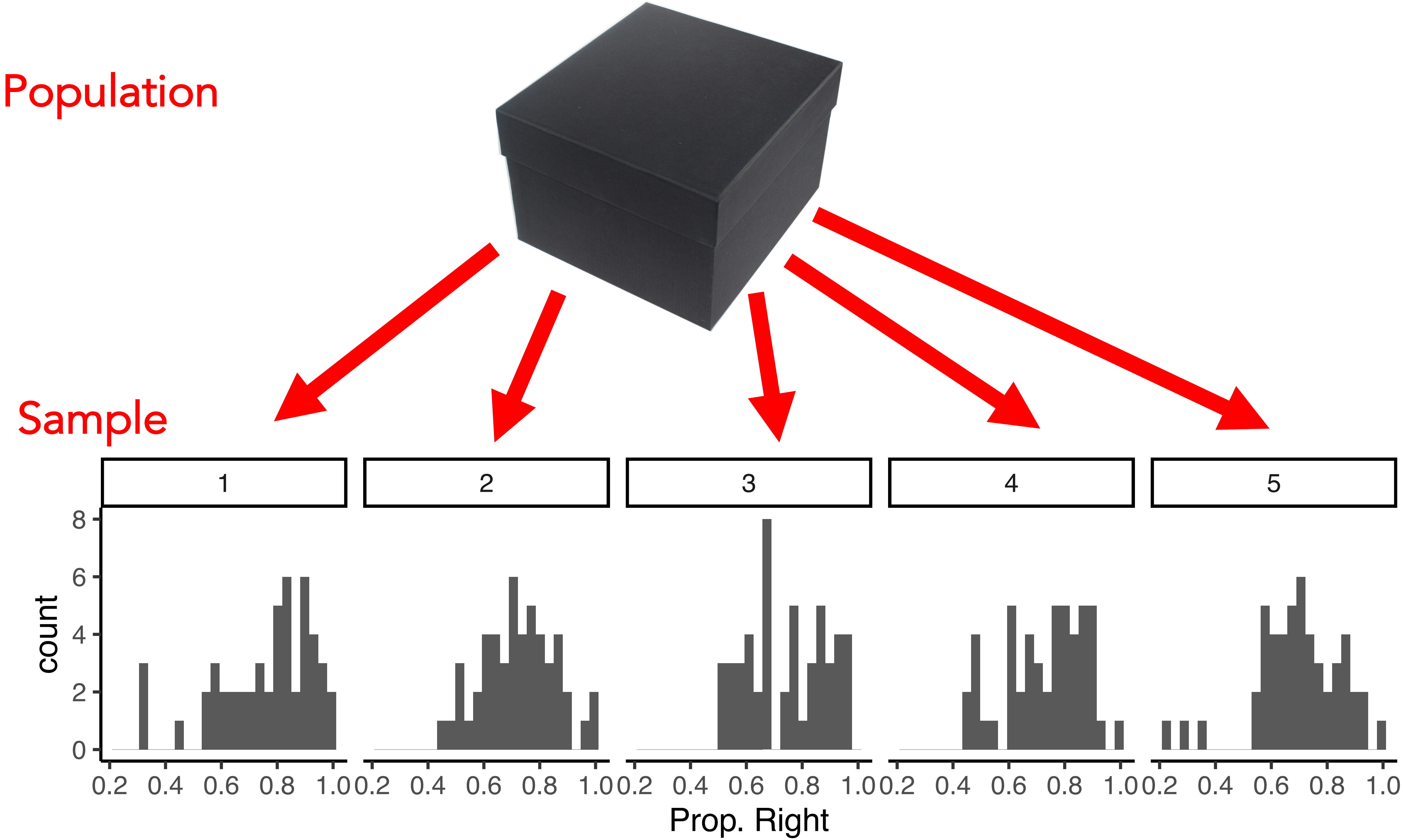


Sample

N = 50

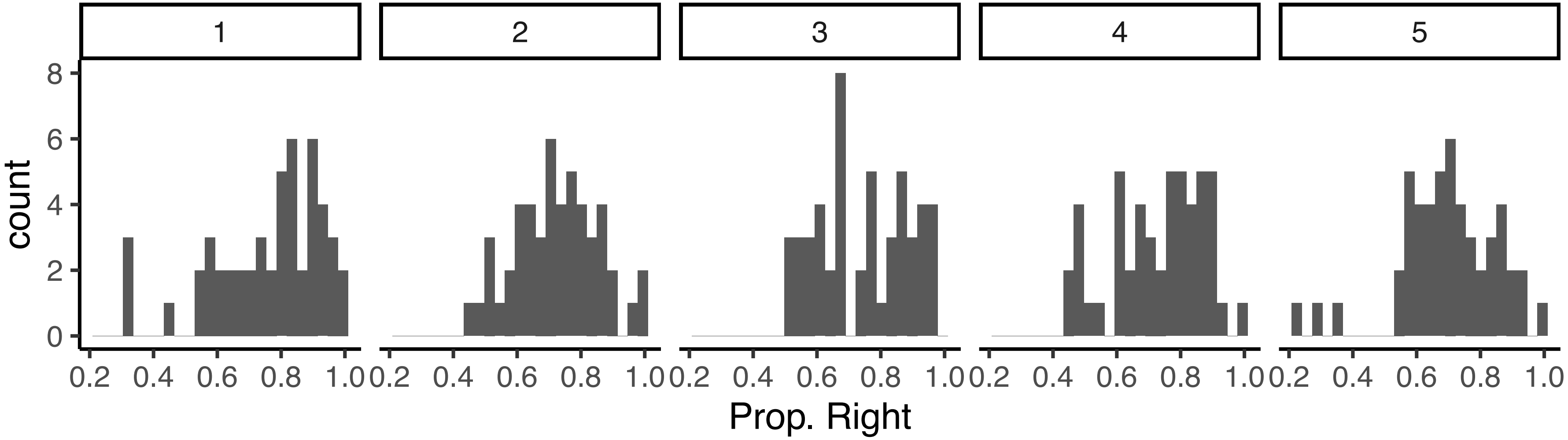


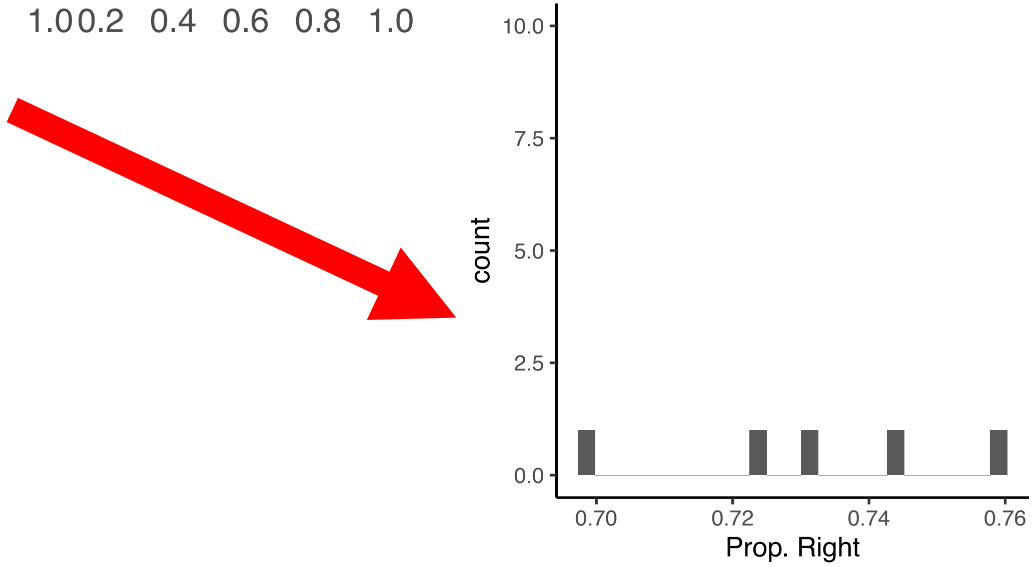
Use mean of sample to estimate mean of population.



Population
Sample
1
2
3
4
5
count
0
2
4
6
8
0.2
0.4
0.6
0.8
1.0
Prop. Right

Sampling distribution of the mean



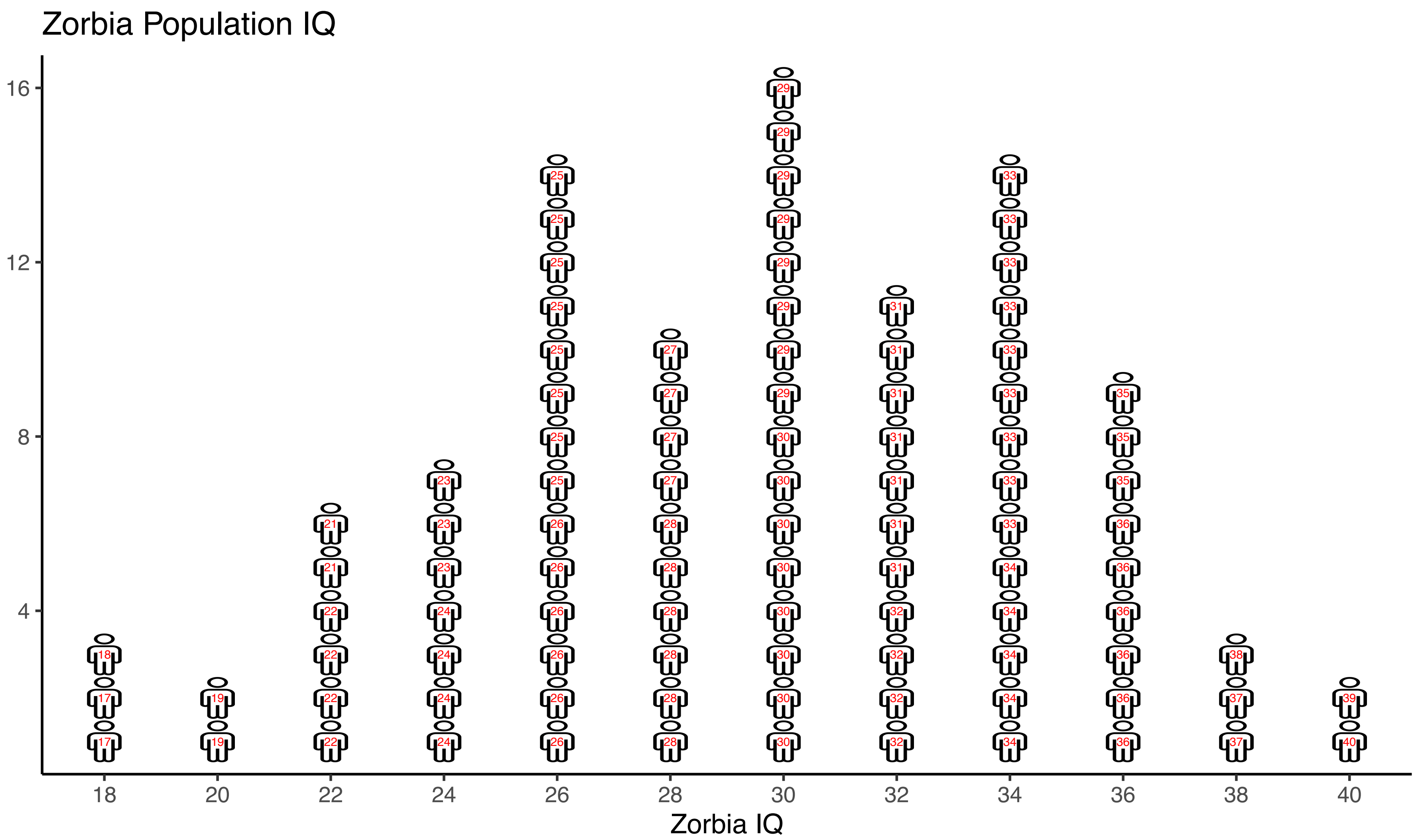


Two things to know about the sampling distribution of the mean

1. The mean of the sampling distribution is the same as the mean of the population.

2. The variance of the sampling distribution of means gets smaller as the sample size increases. (i.e. we get better at estimating the population with more data)

What’s the mean IQ of Zorbia?



N = 97
Mean = 29

In class simulation

What can we learn from sampling the population?

In groups of ~5:

1. Cut the people of Zorbia out.
2. Put them in the envelope.
3. **Each person in the group** should take a sample of **three**.
4. Calculate the average.
5. Write it on a stick note, and add it to the class plot
6. Do steps 3-5 once more.

Key points from Zorbia Simulation

- Mores samples give better estimate of population mean
- Two samples from the same population will tend to have somewhat different means
  - Conversely, two different samples means does NOT mean that they come from different populations

Next Time: Distributions and probability

Explore this Shiny app: https://gallery.shinyapps.io/CLT_mean/

Chapter 16 Hypothesis testing

In the first chapter we discussed the three major goals of statistics:

- Describe
- Decide
- Predict

In this chapter we will introduce the ideas behind the use of statistics to make decisions – in particular, decisions about whether a particular hypothesis is supported by the data.

Acknowledgements

- Slides 12-13 have content adapted from Danielle Navarro, Learning Statistics with R (https://learningstatisticswithr.com/)