Statistical Foundations: Sampling

4 October 2021

Modern Research Methods

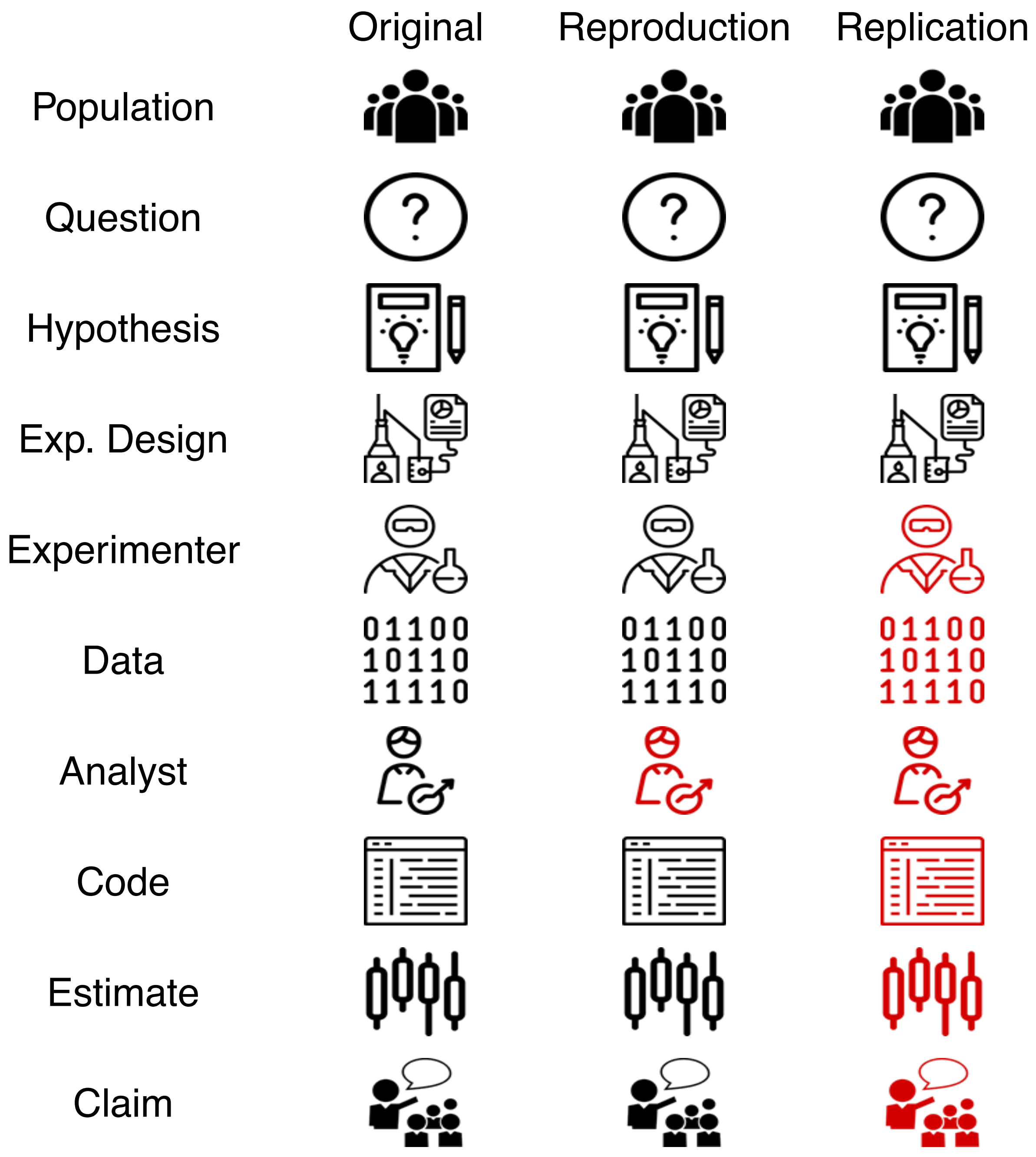
Overview of course

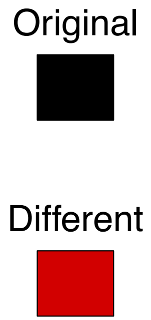
1) **Philosophy of Cumulative Science**

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

3) **Repeating an Experiment** – Intro to statistical inference, 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

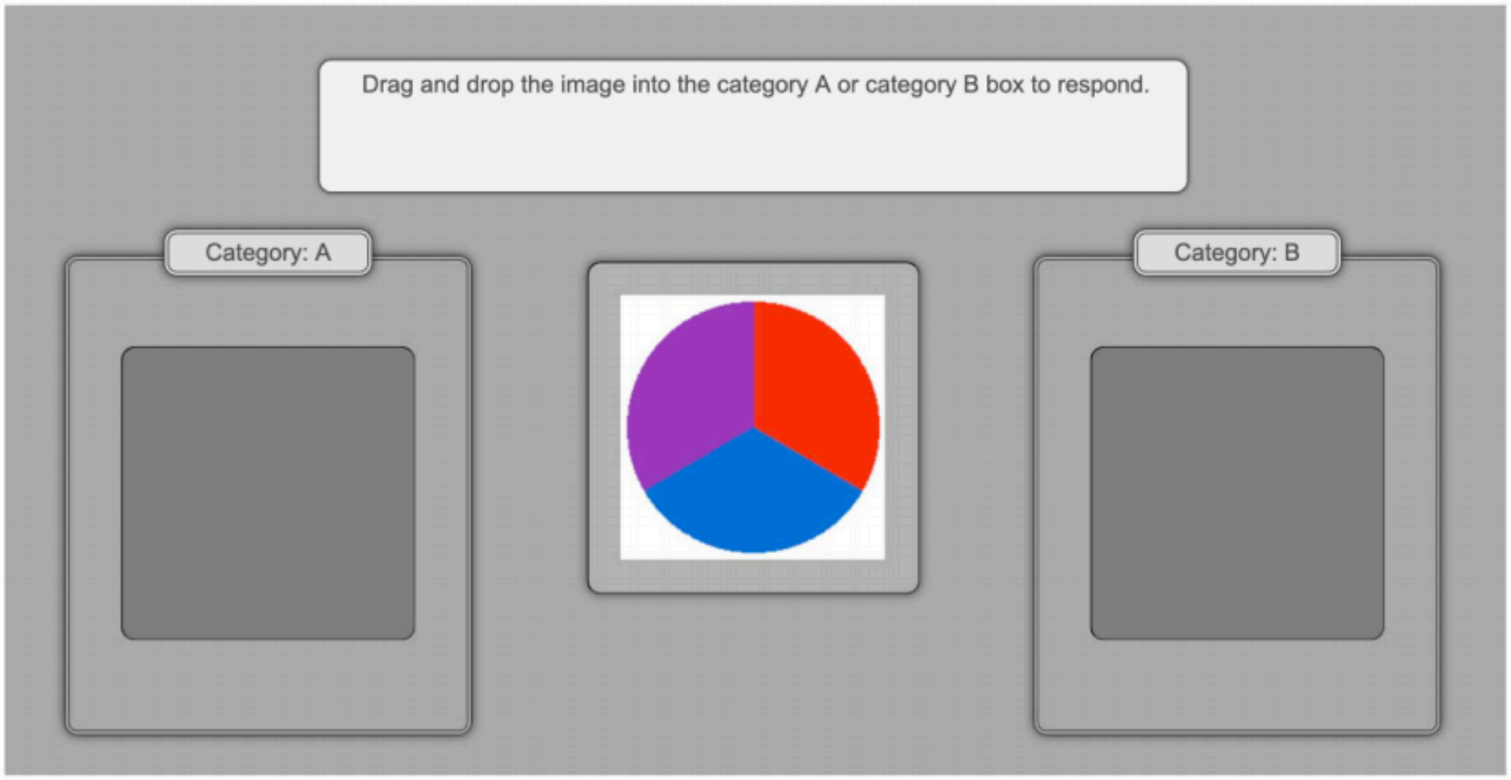
Cognition

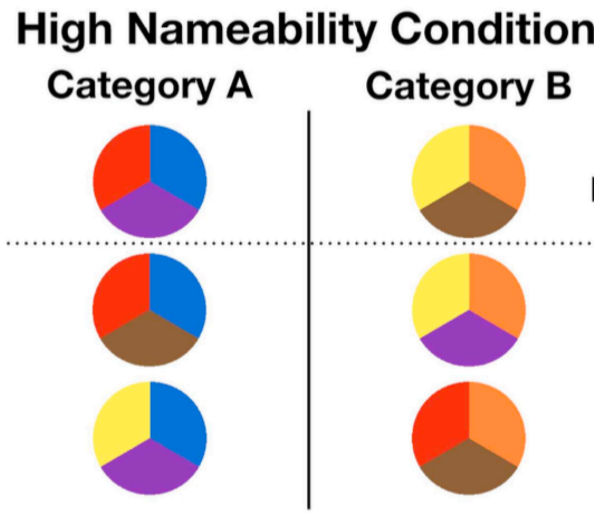
journal homepage: www.elsevier.com/locate/cognit

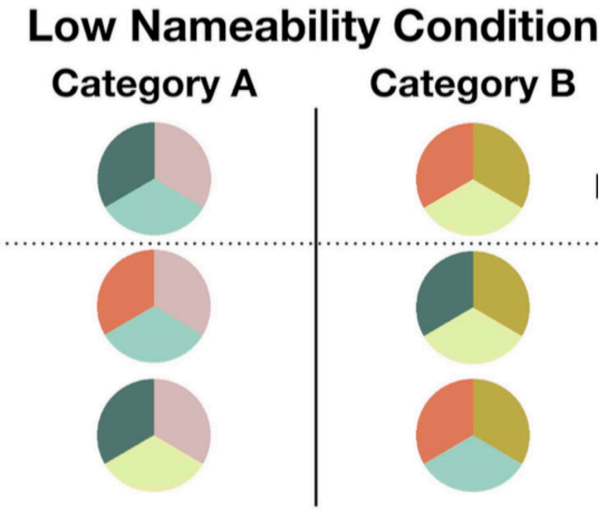
Finding categories through words: More nameable features improve category learning

Martin Zettersten*, Gary Lupyan

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

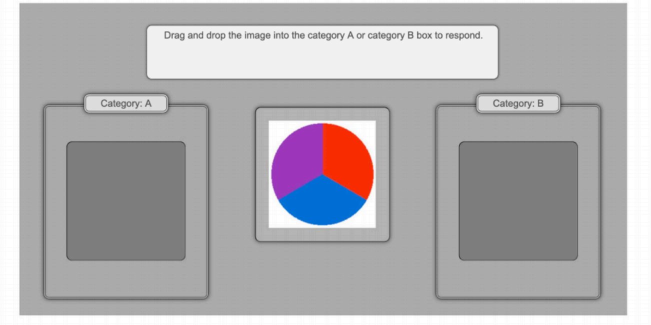






Replicating Zettersten and Lupyan (2020)

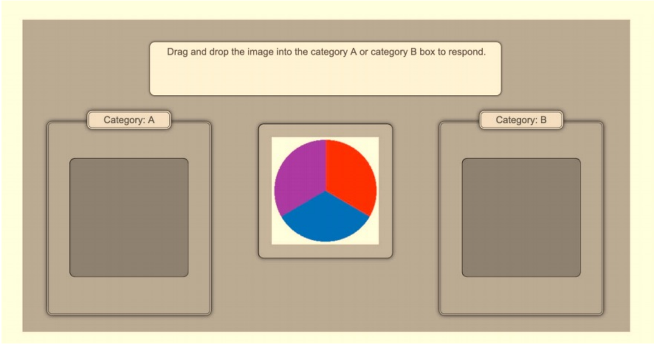
Original



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?

> 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 *statistical inference.*

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

Frameworks of statistical inference

- Does 75 differ from 69?
- Null hypothesis testing
  - Do a hypothesis test, get a p-value
  - If p-value is less that .05 -> difference is “significant”
- Estimation
  - The difference between 75 and 69 is 7 +/- 3

- Estimation is a much more productive framework
  - Contains more information (not black or white)
  - NHST certainty is an illusion – satisfies human “preference for black or white over nuance” (Cumming, 2015)

Is the temperature less than 60 degrees at your vacation destination?

- Null hypothesis testing -> yes
- Estimation -> 13 +/ 3 degrees

Distributions

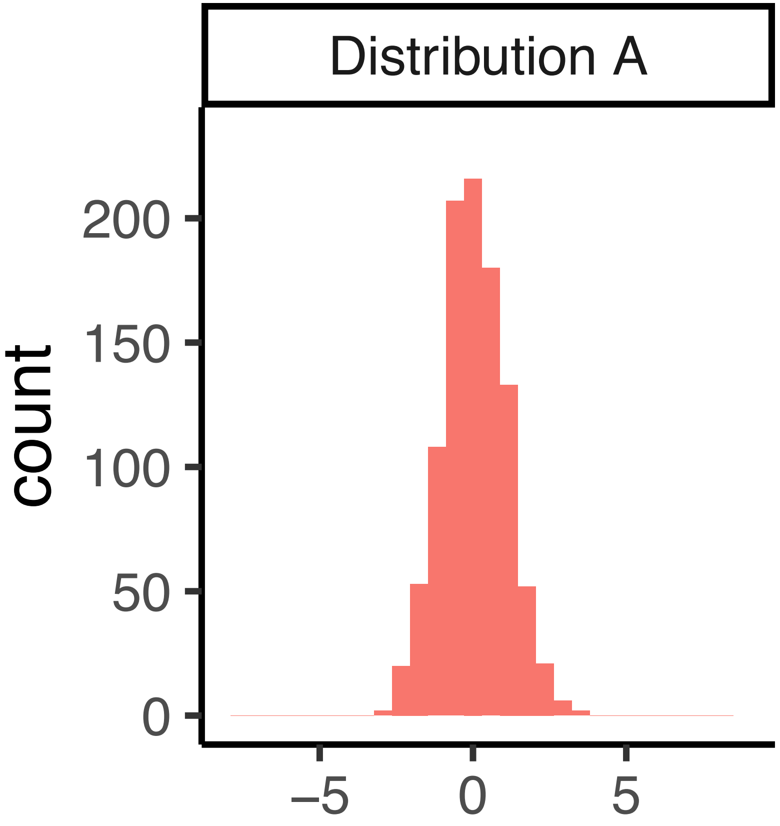
Distributions = counts of a variable

Plot with histograms

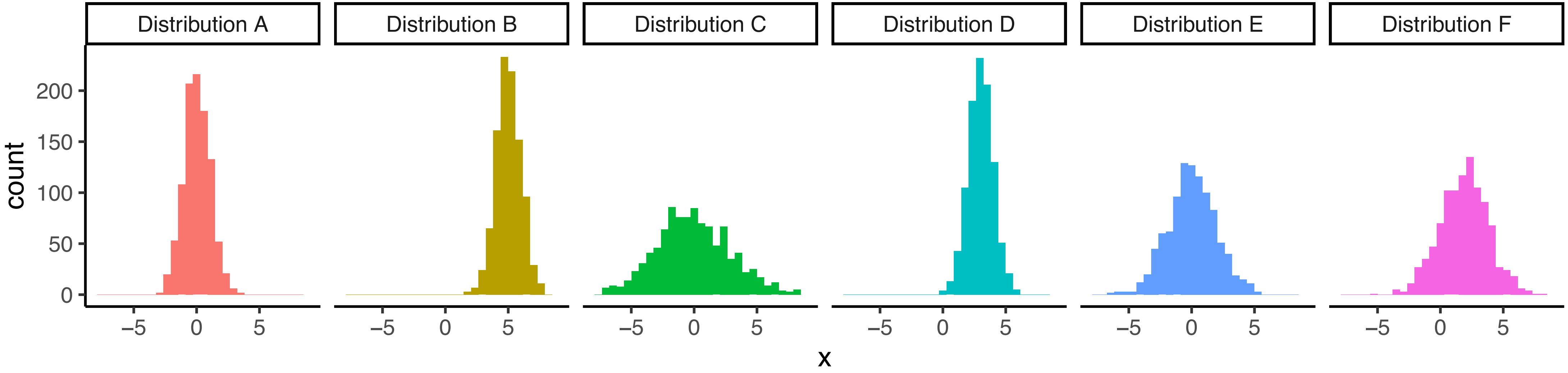
Two measures:

- **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

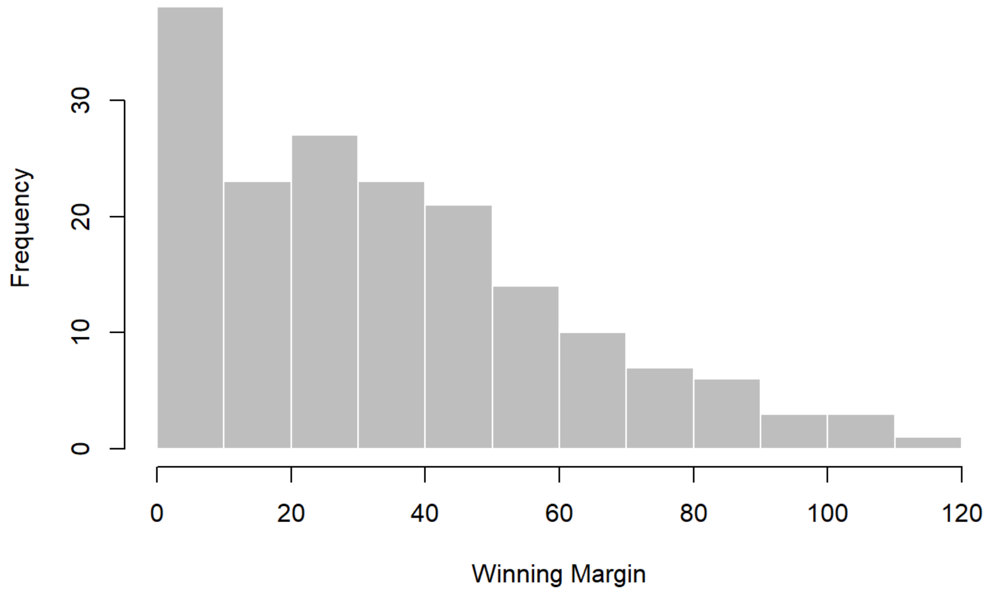
Mean = 3
Low variance

Mean = 0
High variance

Mean = 2
High variance

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

Calculating mean and 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

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

Mean

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

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

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

Standard deviation is the square root of variance.

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)
- 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.**

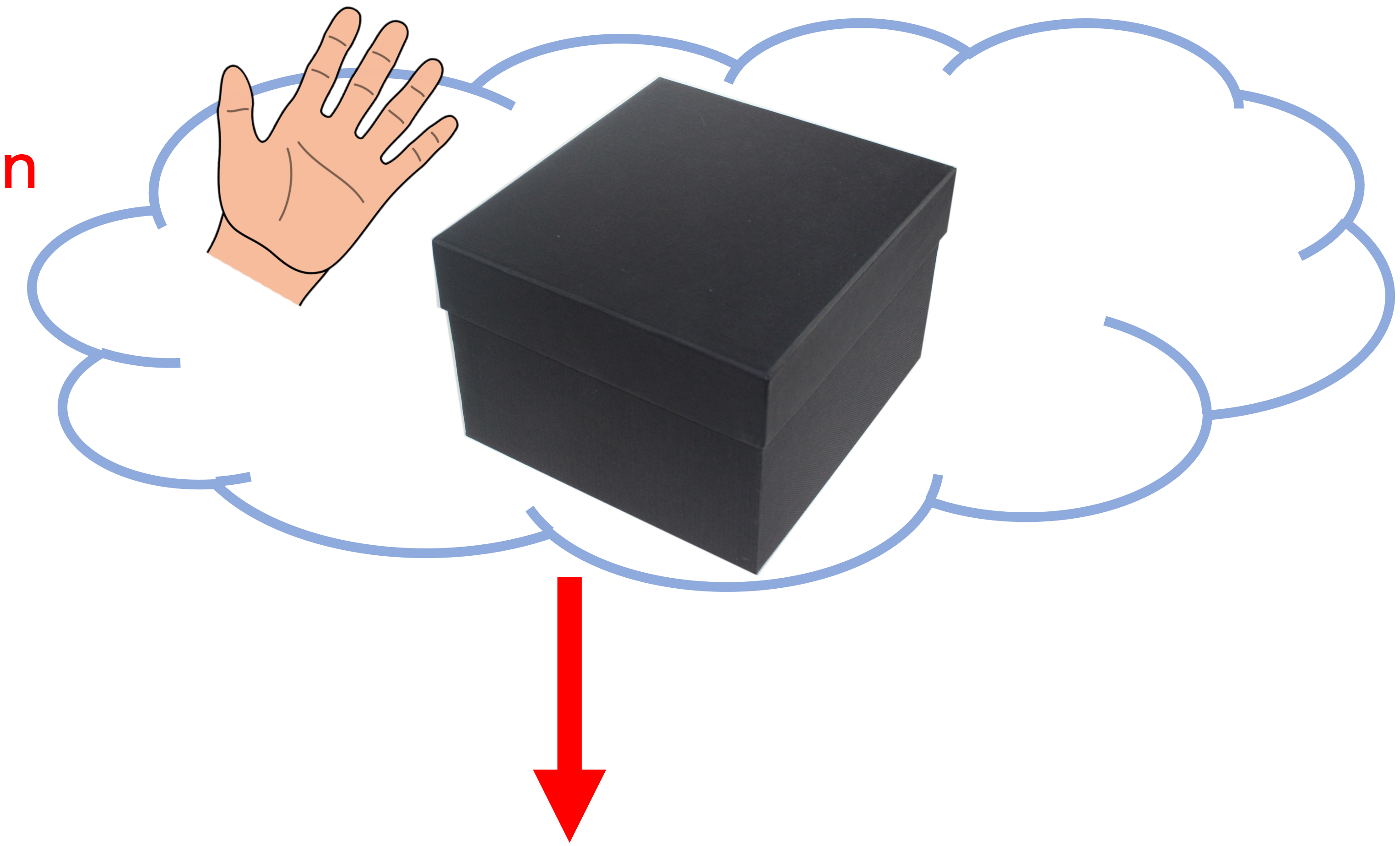
Population vs. sample

- 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-Madison (44k)
- **Sample**: Zettersten and Lupyan only tested 50 participants.
- We don't know what the population looks like (and we usually don't).

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

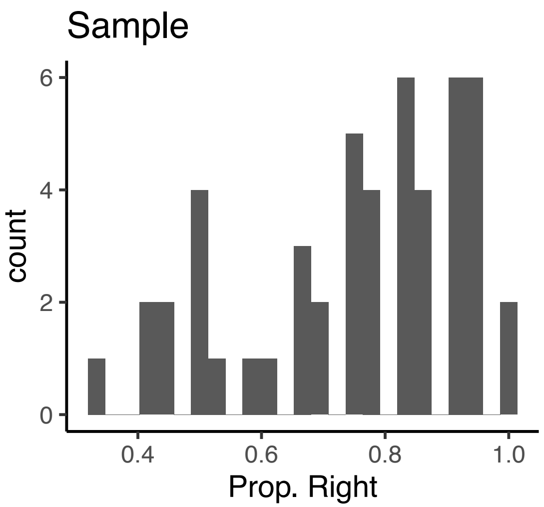
Population

N = a lot

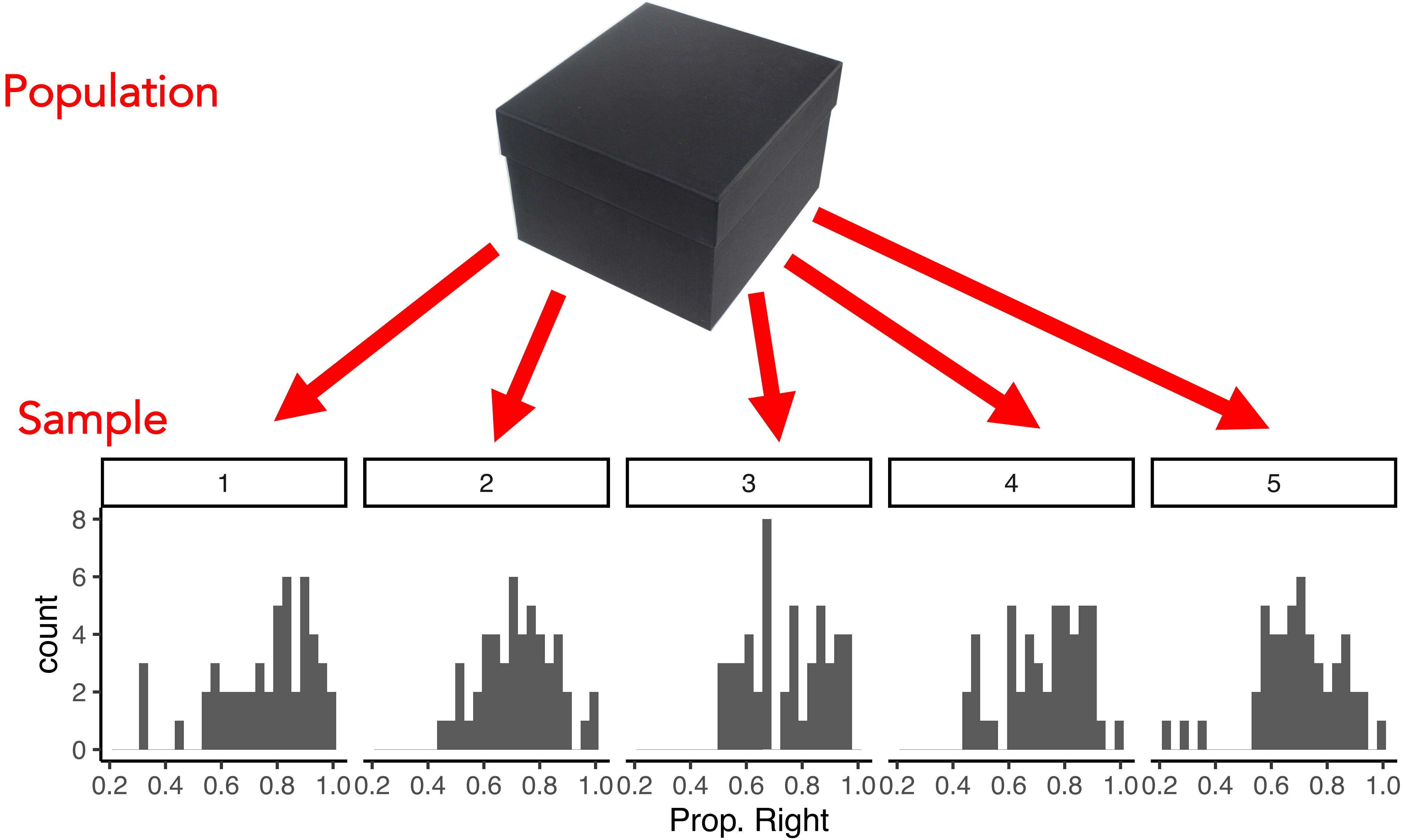


Sample

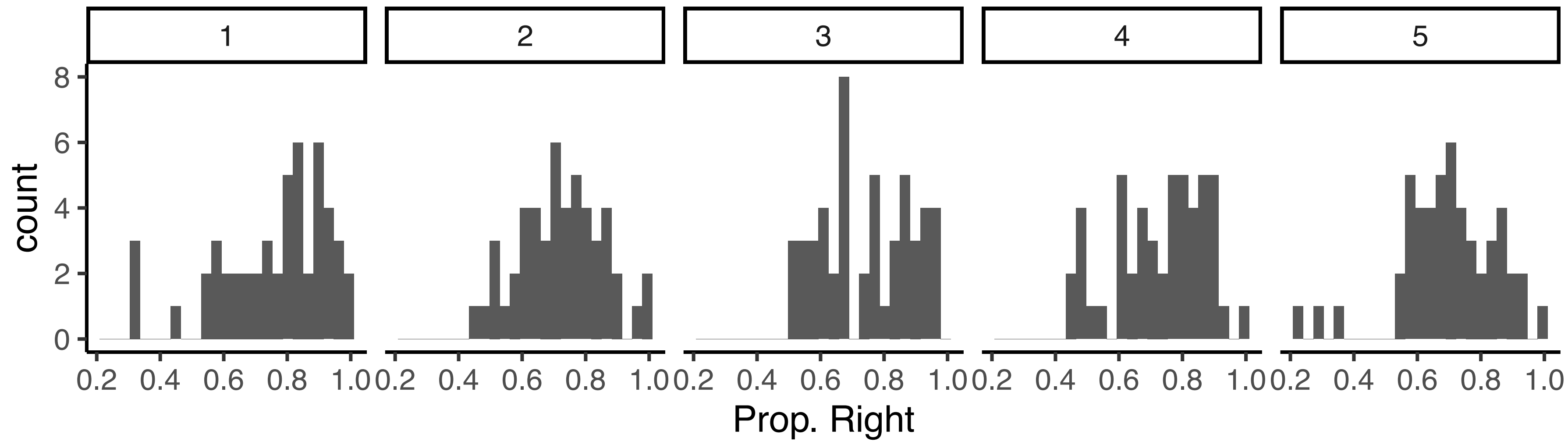
N = 50

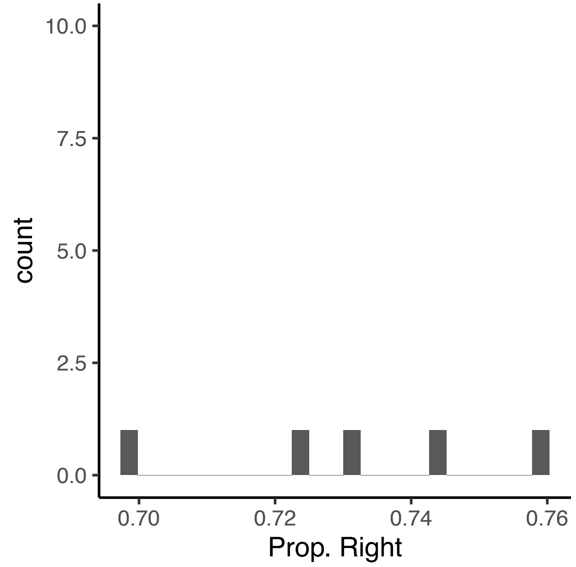


Use mean of sample to estimate mean of population.

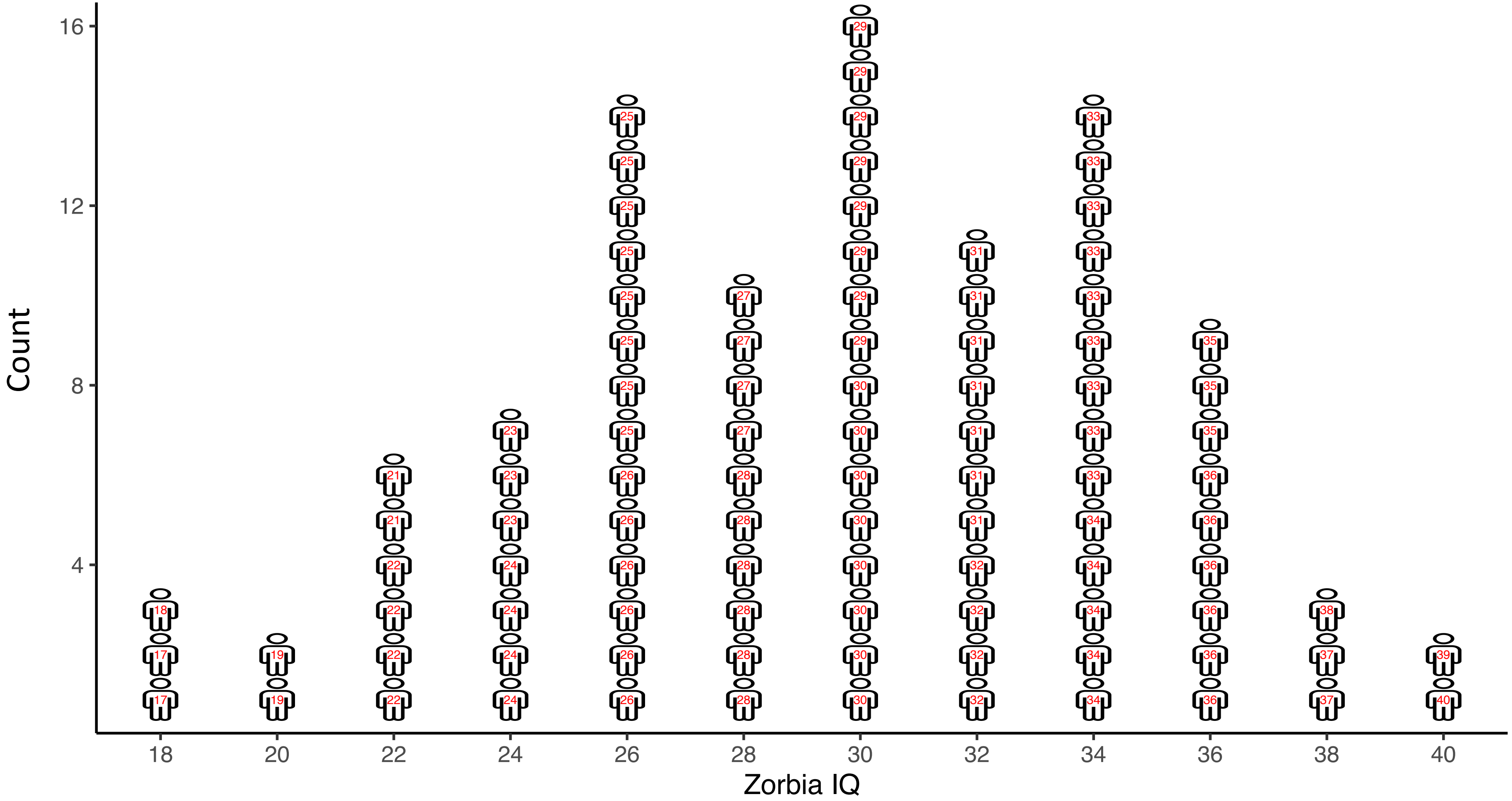

Population
Sample
1
2
3
4
5
count
0
2
4
6
8
0.2
0.4
0.6
0.8
1.0
Prop. Right

Sampling distribution of the mean





Zorbia Population IQ



N = 97
Mean = 29

In class simulation

What can we learn from a sample of this population?

In groups of ~5:

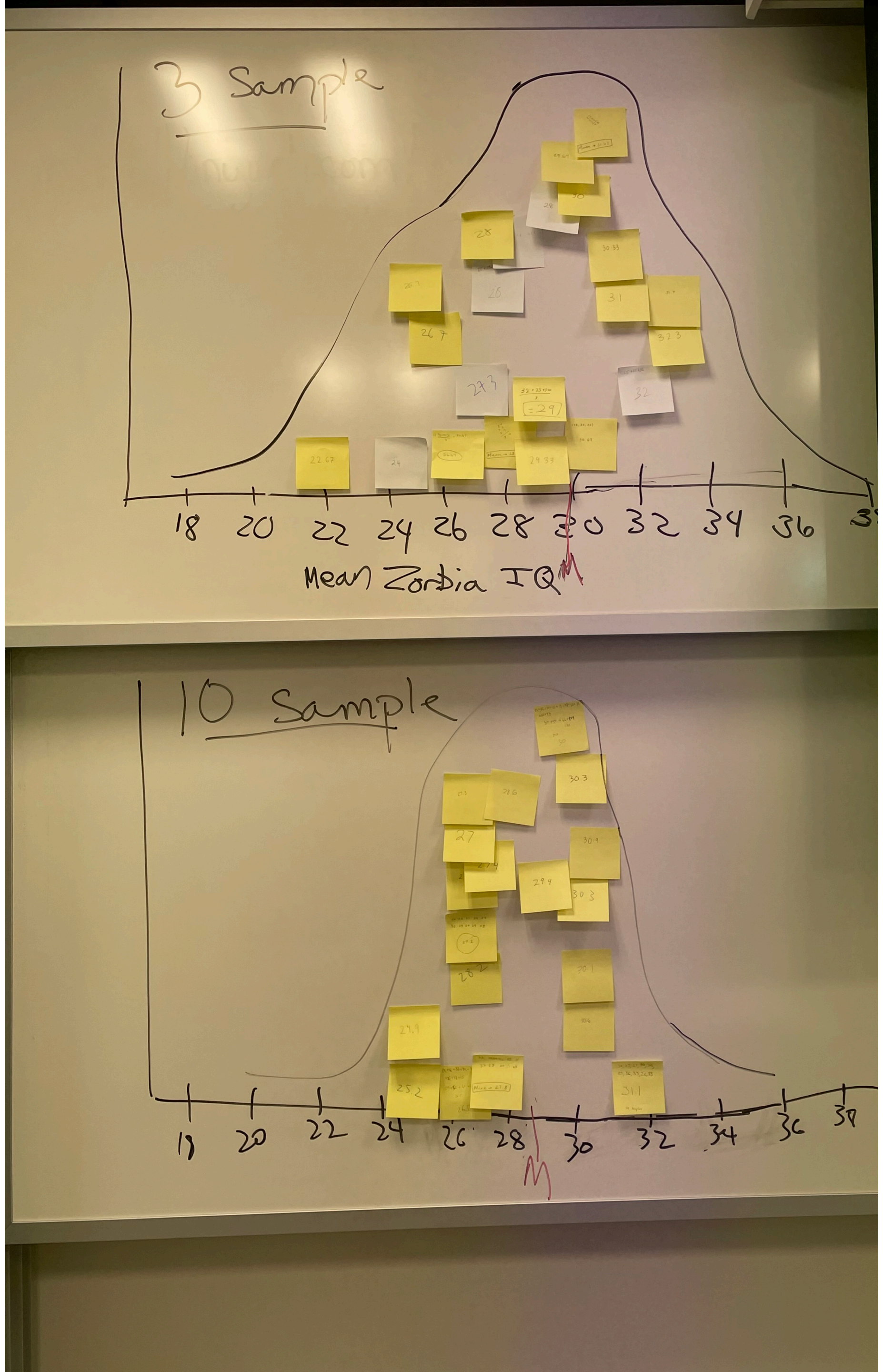
1. Cut the people of Zorbia out.
2. Put them in the envelope.
3. **Each person in the group** should take a sample of **three**.
4. Calculate the average.
5. Write it on a stick note, and add it to the class plot
6. Do steps 3-5 once more.

In class simulation

What can we learn from a sample of this 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 **ten**.
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.


3 Sample
18 20 22 24 26 28 30 32 34 36
Mean Zorbia IQ
M
10 Sample
20 22 24 26 28 30 32 34 36
M

Key points from Zorbia simulation

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

- The variance of the sampling distribution of means gets smaller as the sample size increases
  - Mores samples give better estimate of population mean

Next Time: Confidence Intervals

- Guest lecture from Roderick
- Reading:

Inference by Eye

Confidence Intervals and How to Read Pictures of Data

Geoff Cumming and Sue Finch
La Trobe University

Acknowledgements

- Slide 12 adapted from Danielle Navarro, Learning Statistics with R (https://learningstatisticswithr.com/)