Researcher Choices
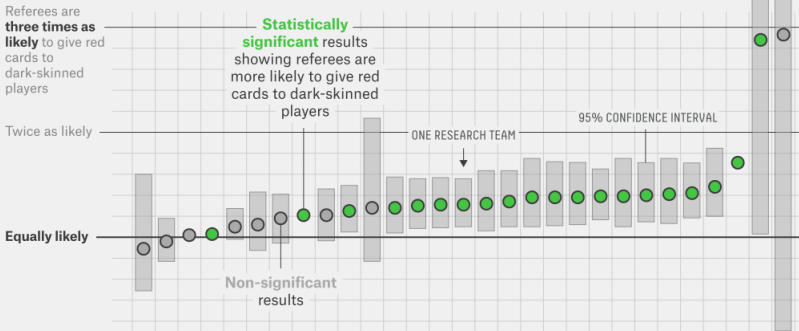
Data Analysis for Journalism and Political Communication
(Spring 2026)

Prof. Bell

Bellringer

Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.

Referees are
three times as
likely to give red
cards to
dark-skinned
players

Statistically
significant results
showing referees are
more likely to give red
cards to dark-skinned
players

Twice as likely

ONE RESEARCH TEAM

95% CONFIDENCE INTERVAL

Equally likely

Non-significant
results

ⓥ FIVETHIRTYEIGHT

SOURCE: BRIAN NOSEK ET AL.

RESEARCHER CHOICES

1. What is my hypothesis?

RESEARCHER CHOICES

1. What is my hypothesis?
2. What do I measure?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?
4. How much data do I collect?

Researcher Choices

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?
4. How much data do I collect?
5. What statistical analyses do I use?

RESEARCHER CHOICES

1. What is my hypothesis?
2. What do I measure?
3. How do I collect my data?
4. How much data do I collect?
5. What statistical analyses do I use?
6. How do I handle outliers, missing data, and other peculiarities?

WHAT IS MY HYPOTHESIS?

Choosing a hypothesis is all about avoiding error:

- **Type I error**: False positives
  - ▸ Sending the innocent to jail - this is bad!

WHAT IS MY HYPOTHESIS?

Choosing a hypothesis is all about avoiding error:

- **Type I error**: False positives
  - ▸ Sending the innocent to jail - this is bad!

- **Type II error**: False negatives
  - ▸ Letting the guilty go free - we can accept this

WHAT IS MY HYPOTHESIS?

Choosing a hypothesis is all about avoiding error:

- **Type I error**: False positives
    - ▸ Sending the innocent to jail - this is bad!

- **Type II error**: False negatives
    - ▸ Letting the guilty go free - we can accept this

Our goal is to reduce Type I error. Assume that the data is innocent (that the hypothesis is false) until it is proven guilty.

Measuring Type I Error

P-value

The chance that we would get a particular result from our test if the true answer is false

MEASURING TYPE I ERROR

P-value

The chance that we would get a particular result from our test if the true answer is false

- The p-value is our chance of committing a Type I error - sending the innocent to jail

Measuring Type I Error

P-value

The chance that we would get a particular result from our test if the true answer is false

- The p-value is our chance of committing a Type I error - sending the innocent to jail
- Common p-value cut-offs in scientific research: .01, .05, and .1 indicate **statistical significance**

MEASURING TYPE I ERROR

Hypothesis: A student cheated on an exam.

P-value (chance that we conclude that student cheated, but they did not): .50

MEASURING TYPE I ERROR

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams

P-value (chance that we conclude that student cheated, but they did not): .25

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams
- The student finished their exam more quickly than other students

P-value (chance that we conclude that student cheated, but they did not): .15

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams
- The student finished their exam more quickly than other students
- The student's roommate saw them up all night studying before the exam

P-value (chance that we conclude that student cheated, but they did not): .40

Measuring Type I Error

Hypothesis: A student cheated on an exam.

- The student performed much better on this exam than on previous exams
- The student finished their exam more quickly than other students
- The student's roommate saw them up all night studying before the exam
- The student missed the same questions as other students

P-value (chance that we conclude that student cheated, but they did not): .75

THE SCIENTIFIC METHOD

- P-values are a product of the data we use and our choices about what we include and exclude from the analysis.

The Scientific Method

- P-values are a product of the data we use and our choices about what we include and exclude from the analysis.
- We follow the scientific method: Theory $\Rightarrow$ Hypothesis $\Rightarrow$ Test $\Rightarrow$ Analyze $\Rightarrow$ Report

The Scientific Method

- P-values are a product of the data we use and our choices about what we include and exclude from the analysis.
- We follow the scientific method: Theory $\Rightarrow$ Hypothesis $\Rightarrow$ Test $\Rightarrow$ Analyze $\Rightarrow$ Report
- But in practice, no analysis plan survives contact with the data

Are Democrats or Republicans Good for the Economy?

Use (a recreation of) FiveThirtyEight's online tool to test what you think is the best approach to answering the question. There are no right or wrong answers - just select the model you think is best, and report your results in the form:

https://bit.ly/smpa2152

(The link to the tool is on the form.)

ARE DEMOCRATS OR REPUBLICANS GOOD FOR THE ECONOMY?

WHAT DO I MEASURE?

Operationalization
The process of defining a measurable version of a concept.

The US maternal mortality rate rose as more states adopted the "pregnancy checkbox"

As more states in the US adopted the "pregnancy checkbox" on death certificates — which asked if the deceased had been pregnant or recently pregnant — the reported maternal mortality rate rose.

Maternal mortality rate, per 100,000 females
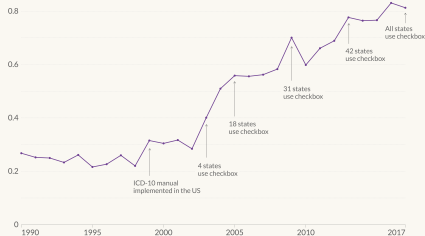
Source: WHO Mortality Database (2022). Adapted from RS Joseph et al. (2021) Maternal mortality in the United States. Data includes "late maternal deaths", which occur up to 1 year after the end of pregnancy.
OurWorldInData.org – Research and data to make progress against the world's largest problems. Licensed under CC-BY by the author Saloni Dattani

A Swedish study illuminated the difference between autism as a diagnosis and autism as a condition

Autistic traits were observed at a relatively stable rate ...

... even as the rate of autism diagnosis more than doubled over a decade.
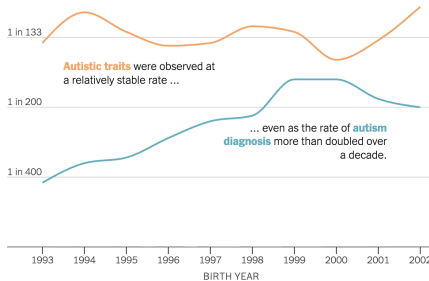
Source: Lundstrom et al., The BMJ (2015)
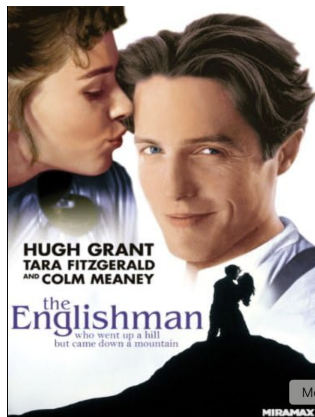
1. Unambiguous

Principles of Good Operationalization

1. Unambiguous
2. Parsimonious

Principles of Good Operationalization

1. Unambiguous
2. Parsimonious
3. Accurate

Principles of Good Operationalization

1. Unambiguous
2. Parsimonious
3. Accurate
4. Reliable

PRINCIPLES OF GOOD OPERATIONALIZATION

1. Unambiguous
2. Parsimonious
3. Accurate
4. Reliable
5. Feasible

PRINCIPLES OF GOOD OPERATIONALIZATION

Index (or Scale) Variable

An index (or scale) variable is a type of proxy measure in which the researcher combines multiple sources of data into a single representation of the concept under investigation using a pre-defined mathematical operation.

Gender Inequality Index, 1993 to 2023

This index covers three dimensions: reproductive health, empowerment, and economic status. Scores are between 0-1 and higher values indicate higher inequalities.
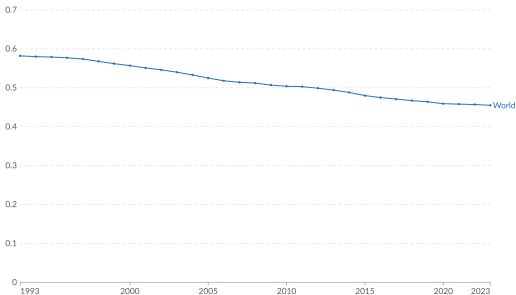
Data source: UNDP, Human Development Report (2025)

OurWorldInData.org/economic-inequality-by-gender | CC BY

EXERCISE: OPERATIONALIZATION

1. You want to measure how happy people are
2. You want to measure people's driving ability
3. You want to measure the political ideology of a member of Congress

HOW DO I COLLECT MY DATA?

Data Generating Process

The rules and procedures that produce the data one is interested in

HOW DO I COLLECT MY DATA?

Data Generating Process

The rules and procedures that produce the data one is interested in

- No amount of statistical wizardry can compensate for bad data
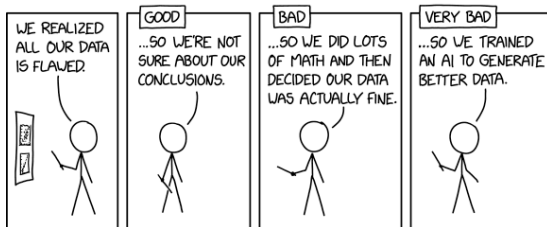
HOW DO I COLLECT MY DATA?

Data Generating Process

The rules and procedures that produce the data one is interested in

- No amount of statistical wizardry can compensate for bad data
- The gold standard of data generating processes is the **random sample**

SAMPLING

- The group we are interested in studying is known as the **population**

SAMPLING

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**

Sampling

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**
- The key to any data analysis project is a quality sample, which is determined by two elements:

Sampling

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**
- The key to any data analysis project is a quality sample, which is determined by two elements:
    1. A **random sample** of the population

Sampling

- The group we are interested in studying is known as the **population**
- Often, we are not able to count every unit in the population, so we take a **sample**
- The key to any data analysis project is a quality sample, which is determined by two elements:
  1. A **random sample** of the population
  2. The **sample size** is sufficiently large

Random Sample

Definition

The probability of any given unit being drawn from the population is uniform (the same)

Random Sample

Definition
The probability of any given unit being drawn from the population is uniform (the same)

- A failure of each unit to have a uniform probability of being drawn from the population is known as **selection bias**

RANDOM SAMPLE

Definition
The probability of any given unit being drawn from the population is uniform (the same)

- A failure of each unit to have a uniform probability of being drawn from the population is known as **selection bias**
- Units are "selecting" into our data because they are more observable than other units

Random Sample

Definition
The probability of any given unit being drawn from the
population is uniform (the same)

- A failure of each unit to have a uniform probability of being
  drawn from the population is known as **selection bias**
- Units are "selecting" into our data because they are more
  observable than other units
- Selection bias reduces our **generalizability** to the
  population because the data is not representative of the
  population

Random Sample

Can I randomly sample 10 students from this class to generalize to:

- the population of GW students?

Random Sample

Can I randomly sample 10 students from this class to generalize to:

- the population of GW students?
- the population of SMPA students?

Random Sample

Can I randomly sample 10 students from this class to generalize to:

- the population of GW students?
- the population of SMPA students?
- the population of Data Analysis students?

EXERCISE: SELECTION BIAS