Methods Camp

Session 1

Instructor: Casey F. Breen TA: Leng Seong Che

Department of Sociology
University of Texas at Austin

August 17th, 2026

Welcome and objectives

- We are excited to welcome you to UT and to help you prep for your first year!
- **Goal** for this week: build comfort and exposure with the basic language of quantitative work you will see throughout methods courses and papers.
- Ask lots of questions

Methods Bootcamp Schedule

Day 1	Monday, August 17	2:00–4:30 PM	Math
Day 2	Tuesday, August 18	9:00 AM–12:00 PM	Math
Day 3	Wednesday, August 19	9:00 AM–12:00 PM	Coding
Day 4	Thursday, August 20	9:00 AM–12:00 PM	Coding
Day 5	Friday, August 21	9:00–10:00 AM	Workflow

Why Do We Need Functions?

- Social Scientists make claims about how one thing is related to another:
  - "Poverty is negatively associated with high school graduation."
  - "Adult mortality varies by educational attainment."

Why Do We Need Functions?

- Social Scientists make claims about how one thing is related to another:
  - "Poverty is negatively associated with high school graduation."
  - "Adult mortality varies by educational attainment."
- A **function** describes **how an outcome changes as some other quantity changes**.

Why Learn Functions?

Learning **functions** will help you understand:

- How researchers model relationships between variables
- What regression coefficients and equations mean
- How to interpret tables and figures in research articles

What is a Function?

A **function** $f(x)$ is a rule that takes an input and produces exactly one output.

What is a Function?

A **function** $f(x)$ is a rule that takes an input and produces exactly one output.

- In social sciences, we may write things like:

$$\text{Income} = f(\text{Education}, \ldots)$$

What is a Function?

A **function** $f(x)$ is a rule that takes an input and produces exactly one output.

- In social sciences, we may write things like:

$$\text{Income} = f(\text{Education}, \dots)$$

- Example in words: "Take a respondent's years of schooling, and the function tells you their predicted income."

Function Notation

- We write a function as $f(x)$, read aloud as "f of x."

Function Notation

- We write a function as $f(x)$, read aloud as "f of x."
- x is the **input** (also called the **independent variable**).
- $f(x)$ is the **output** (also called the **dependent variable**) that depends on x.

Function Notation

- We write a function as $f(x)$, read aloud as "f of x."
- x is the **input** (also called the **independent variable**).
- $f(x)$ is the **output** (also called the **dependent variable**) that depends on x.
- The letter f is just a label for the rule; it could be f, g, or any name.

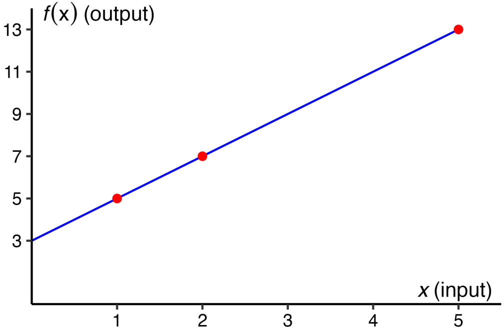
Evaluating a Function

Let $f(x) = 2x + 3$.

- To evaluate the function at a specific input, substitute that number in for x.

x	Substitute	$f(x)$
1	$2(1) + 3$	5
2	$2(2) + 3$	7
5	$2(5) + 3$	13

Graphing a Function



Note: a straight line means the output increases by the same amount for every one-unit increase in x — this is what makes $f(x) = 2x + 3$ a ***linear*** *function.*

Functions With More Than One Input

- A function can take **more than one input**.

Functions With More Than One Input

- A function can take **more than one input**.
- $f(x, y)$ — a rule with two inputs, x and y.

Functions With More Than One Input

- A function can take **more than one input**.
- $f(x, y)$ — a rule with two inputs, x and y.
- Example in words: "a respondent's reported life satisfaction depends on both their income *and* their health."

Functions With More Than One Input

$$
\begin{aligned}
f(x, y) &= x + 2y \\
f(1, 4) &= 1 + 2(4) = 9 \\
f(3, 2) &= 3 + 2(2) = 7
\end{aligned}
$$

Exponents and Logs

- **Growth is often multiplicative, not additive.**
  - Populations, prices, and epidemics tend to grow by a *percentage* each period, not by a fixed amount.

Exponents and Logs

- **Growth is often multiplicative, not additive.**
  - Populations, prices, and epidemics tend to grow by a *percentage* each period, not by a fixed amount.
- **Some social variables are skewed.**
  - Income and wealth may have a few very large values.

Exponents and Logs

- **Growth is often multiplicative, not additive.**
  - Populations, prices, and epidemics tend to grow by a *percentage* each period, not by a fixed amount.
- **Some social variables are skewed.**
  - Income and wealth may have a few very large values.
- **Logs are a shortcut for percentage change.**
  - Comes up constantly when describing change over time.

What Is an Exponent?

b^n means multiplying b by itself n times.

What Is an Exponent?

b^n means multiplying b by itself n times.

- $2^3 = 2 \times 2 \times 2 = 8$
- $10^2 = 10 \times 10 = 100$

What Is an Exponent?

b^n means multiplying b by itself n times.

- $2^3 = 2 \times 2 \times 2 = 8$
- $10^2 = 10 \times 10 = 100$
- b is the **base**; n is the **exponent**, or power.

Exponents: Key rules

$$b^m \times b^n = b^{m+n}$$

$$\frac{b^m}{b^n} = b^{m-n}$$

$$(b^m)^n = b^{mn}$$

$$b^0 = 1, \qquad b^{-n} = \frac{1}{b^n}$$

What is a Logarithm?

The **logarithm** is the **inverse** of exponentiation. It answers: “To what power must I raise the base to get this number?”

$$\log_b(x) = y \quad \Longleftrightarrow \quad b^y = x$$

What is a Logarithm?

The **logarithm** is the **inverse** of exponentiation. It answers: "To what power must I raise the base to get this number?"

$$\log_b(x) = y \quad \Longleftrightarrow \quad b^y = x$$

- Example: $\log_{10}(100) = 2$ because $10^2 = 100$.
- Example: $\log_2(8) = 3$ because $2^3 = 8$.

Natural Log

In most statistical software, the default logarithm is the **natural logarithm**,

$$\ln(x) = \log_e(x), \qquad e \approx 2.718.$$

Log Rules

$$\ln(a \times b) = \ln(a) + \ln(b)$$
$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$$
$$\ln(a^n) = n\ln(a)$$
$$\ln(1) = 0, \qquad \ln(e) = 1$$

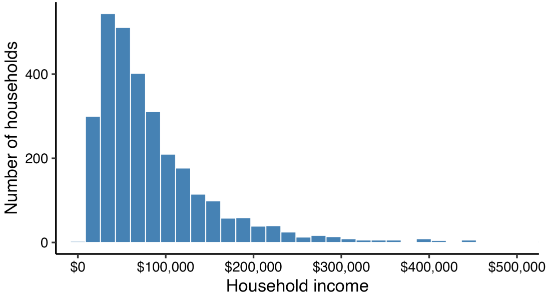
Why Log-Transform Variables?

Household income is a classic example of a **right-skewed** variable (a few very large values).

Raw income ($)	30,000	45,000	60,000	90,000	900,000
$\ln$(income)	10.31	10.71	11.00	11.41	13.71

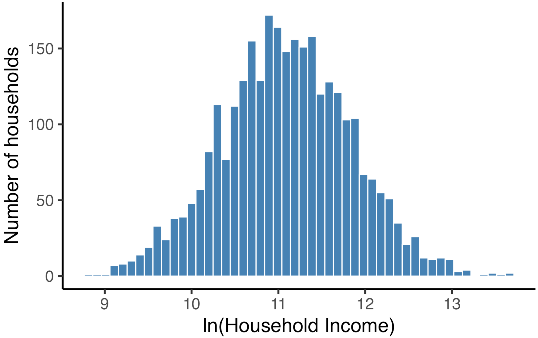
Why Sociologists Log-Transform Variables

- On the raw scale, the last value is **wildly far** from the rest.



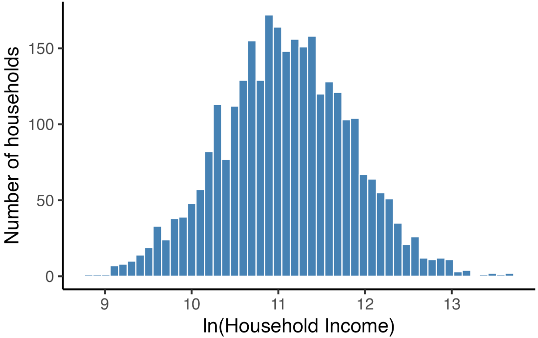
Why Log-Transform Variables?

- On the log scale, the gap shrinks dramatically.



Why Log-Transform Variables?

- On the log scale, the gap shrinks dramatically.
- This is why many social science variables are reported or modeled "in logs"—the transformation makes extreme values far less dominant.



Summation

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

Summation

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

- $\sum$ (the Greek letter sigma) means "add up what follows."

Summation

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

- $\sum$ (the Greek letter sigma) means "add up what follows."
- i is the **index**: indicates i^{th} observation (respondent, case, year, ...).

Summation

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

- $\sum$ (the Greek letter sigma) means "add up what follows."
- i is the **index**: indicates i^{th} observation (respondent, case, year, ...).
- The number below $\sum$ $(i = 1)$ is where the counting **starts**; the number above (n) is where it **ends**.

Summation

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

- $\sum$ (the Greek letter sigma) means "add up what follows."
- i is the **index**: indicates i^{th} observation (respondent, case, year, ...).
- The number below $\sum$ $(i = 1)$ is where the counting **starts**; the number above (n) is where it **ends**.
- x_i is the value being added for observation i.

Summation: Example

Suppose we survey 5 respondents about hours of TV watched per day: $\{2, 3, 1, 4, 5\}$.

$$\sum_{i=1}^{5} x_i = 2 + 3 + 1 + 4 + 5 = 15$$

Summation: Example

Suppose we survey 5 respondents about hours of TV watched per day: $\{2, 3, 1, 4, 5\}$.

$$\sum_{i=1}^{5} x_i = 2 + 3 + 1 + 4 + 5 = 15$$

To build the **sample mean**:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{15}{5} = 3$$

Useful Summation Properties

$$\sum_{i=1}^{n} c = n \cdot c \qquad \text{(sum of a constant)}$$

$$\sum_{i=1}^{n} c\, x_i = c \sum_{i=1}^{n} x_i \qquad \text{(constants pull out)}$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \qquad \text{(sum of sums)}$$

The Product Operator (Π)

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times x_3 \times \cdots \times x_n$$

The Product Operator (Π)

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times x_3 \times \cdots \times x_n$$

- Same idea as Σ, but **multiplying** instead of adding.

The Product Operator (Π)

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times x_3 \times \cdots \times x_n$$

- Same idea as Σ, but **multiplying** instead of adding.
- Example: $\prod_{i=1}^{4} i = 1 \times 2 \times 3 \times 4 = 24$

Exercise 1: Functions and Operations

Work in pairs.

1. Let $f(x) = 5x - 4$. Compute $f(2)$ and $f(0)$.
2. Let $f(x, y) = 3x + y^2$. Compute $f(2, 3)$.
3. Simplify: $\ln(50) - \ln(10)$. (Hint: use a log rule.)
4. Solve for x: $\ln(x) = 2$. (Hint: rewrite in exponential form.)
5. A class survey of 6 students records number of siblings: $\{1, 0, 2, 3, 1, 2\}$.
   - Write the total number of siblings using summation notation and calculate it.
   - Compute the mean $\bar{x}$.

Solution: Exercise 1

1. $f(2) = 5(2) - 4 = 6; \quad f(0) = -4$
2. $f(2,3) = 3(2) + 3^2 = 6 + 9 = 15$
3. $\ln(50) - \ln(10) = \ln(50/10) = \ln(5) \approx 1.609$
4. $\ln(x) = 2 \;\Rightarrow\; x = e^2 \approx 7.389$
5. $\sum_{i=1}^{6} x_i = 1 + 0 + 2 + 3 + 1 + 2 = 9; \quad \bar{x} = 9/6 = 1.5$

Ratios and Proportions

What Is a Ratio?

A **ratio** compares the size of one quantity to another. It can be written $a : b$ or a/b.

What Is a Ratio?

A **ratio** compares the size of one quantity to another. It can be written $a : b$ or a/b.

- Example: a classroom has 18 women and 12 men.

$$\text{Ratio of women to men} = 18 : 12 = 3 : 2$$

What Is a Ratio?

A **ratio** compares the size of one quantity to another. It can be written $a : b$ or a/b.

- Example: a classroom has 18 women and 12 men.

$$\text{Ratio of women to men} = 18 : 12 = 3 : 2$$

- A ratio compares **part to part** — it does not need to reference the whole group at all.

What Is a Proportion?

A **proportion** is a special kind of ratio: a **part** compared to the **whole**. It always falls between 0 and 1.

$$\text{Proportion} = \frac{\text{part}}{\text{whole}}$$

What Is a Proportion?

A **proportion** is a special kind of ratio: a **part** compared to the **whole**. It always falls between 0 and 1.

$$\text{Proportion} = \frac{\text{part}}{\text{whole}}$$

- Example: 18 women out of 30 total students $\Rightarrow$ proportion female $= 18/30 = 0.60.$

What Is a Proportion?

A **proportion** is a special kind of ratio: a **part** compared to the **whole**. It always falls between 0 and 1.

$$\text{Proportion} = \frac{\text{part}}{\text{whole}}$$

- Example: 18 women out of 30 total students $\Rightarrow$ proportion female $= 18/30 = 0.60$.
- Unlike a ratio, a proportion always references the entire group.

What Is a Proportion?

A **proportion** is a special kind of ratio: a **part** compared to the **whole**. It always falls between 0 and 1.

$$\text{Proportion} = \frac{\text{part}}{\text{whole}}$$

- Example: 18 women out of 30 total students $\Rightarrow$ proportion female $= 18/30 = 0.60$.
- Unlike a ratio, a proportion always references the entire group.
- Multiply a proportion by 100 to get a **percentage**: $0.60 \times 100 = 60\%$.

Rates, Relative Change, and Percentages

Rates

A **rate** is a ratio that relates the number of events to a population, time period, or another measure of exposure, allowing meaningful comparisons across groups.

Rates

A **rate** is a ratio that relates the number of events to a population, time period, or another measure of exposure, allowing meaningful comparisons across groups.

- **Example: Crime rate per 100,000 people**

$$\text{Crime rate} = \frac{\#\ \text{crimes during a given period}}{\text{population at risk}} \times 100{,}000$$

Why not just report raw counts?

A city of 8 million will almost always have more murders than a town of 8,000 —

Why not just report raw counts?

A city of 8 million will almost always have more murders than a town of 8,000 — **rates let us compare *risk*, not just volume.**

Percentages

$$\text{Percentage} = \frac{\text{part}}{\text{whole}} \times 100$$

Percentages

$$\text{Percentage} = \frac{\text{part}}{\text{whole}} \times 100$$

- Example: 45 out of 60 survey respondents approve of a policy.

$$\frac{45}{60} \times 100 = 75\%$$

Percentages

$$\text{Percentage} = \frac{\text{part}}{\text{whole}} \times 100$$

- Example: 45 out of 60 survey respondents approve of a policy.

$$\frac{45}{60} \times 100 = 75\%$$

A **percentage** is simply a proportion expressed on a 0–100 scale instead of a 0–1 scale.

Absolute Change

$$\text{Absolute change} = \text{New value} - \text{Old value}$$

Absolute Change

$$\text{Absolute change} = \text{New value} - \text{Old value}$$

- Example: a town's population goes from 10,000 to 11,500.

$$\text{Absolute change} = 11{,}500 - 10{,}000 = 1{,}500 \text{ people}$$

Absolute Change

$$\text{Absolute change} = \text{New value} - \text{Old value}$$

- Example: a town's population goes from 10,000 to 11,500.

$$\text{Absolute change} = 11{,}500 - 10{,}000 = 1{,}500 \text{ people}$$

This tells the raw size of the change, but not whether that change is large or small **relative to where you started**.

Relative (Percentage) Change

$$\% \, \text{Change} = \frac{\text{New} - \text{Old}}{\text{Old}} \times 100$$

Relative (Percentage) Change

$$\%\,\text{Change} = \frac{\text{New} - \text{Old}}{\text{Old}} \times 100$$

- Same example: population goes from 10,000 to 11,500.

$$\frac{11{,}500 - 10{,}000}{10{,}000} \times 100 = 15\%$$

Relative (Percentage) Change

$$\%\,\text{Change} = \frac{\text{New} - \text{Old}}{\text{Old}} \times 100$$

- Same example: population goes from 10,000 to 11,500.

$$\frac{11{,}500 - 10{,}000}{10{,}000} \times 100 = 15\%$$

This tells the size of the change **relative to the starting point** — usually the more meaningful number for comparing across groups.

Percentage Points vs. Percent Change

Example: The unemployment rate rises from **5%** to **7%**.

- **Percentage point change**:

$$7 - 5 = 2 \text{ percentage points}$$

Percentage Points vs. Percent Change

Example: The unemployment rate rises from **5%** to **7%**.

- **Percentage point change**:

$$7 - 5 = 2 \text{ percentage points}$$

- **Percent change**:

$$\frac{7-5}{5} \times 100 = 40\%$$

Percentage Points vs. Percent

- A **percentage point** is the arithmetic difference between two percentages.

Percentage Points vs. Percent

- A **percentage point** is the arithmetic difference between two percentages.
- A **percent change** is the relative difference, expressed as a percentage of the starting value.

Percentage Points vs. Percent

- A **percentage point** is the arithmetic difference between two percentages.
- A **percent change** is the relative difference, expressed as a percentage of the starting value.
- These are **not** the same number, and mixing them up **changes the meaning** of a claim.

Recap: Rates, Change, and Percentages

- **Rate**: a ratio standardized by a base (often population and/or time) so groups can be compared fairly.
- **Percentage**: a proportion $\times 100$.
- **Absolute change**: New − Old.
- **Percent (relative) change**: $\dfrac{\text{New} - \text{Old}}{\text{Old}} \times 100.$
- **Percentage point**: the raw arithmetic gap between two percentages — **not** the same as percent change.

Exercise 2: Ratios, Proportions, Rates & Percentages

A town has 5,400 women and 4,600 men.

1. Compute the sex ratio (males per 100 females).
2. Compute the proportion of the town that is female.
3. Last year the town recorded 27 burglaries. Compute the burglary rate per 100,000 residents. (Total population = 10,000.)
4. The town's unemployment rate fell from 8% to 6%.
   - What is the change in **percentage points**?
   - What is the **percent (relative) change**?
5. A city's child poverty rate falls from 20% to 15%. An abstract reads: "Poverty drops 5%." Is this headline accurate? Why or why not?

Solution: Exercise 2

1. Sex ratio $= \dfrac{4{,}600}{5{,}400} \times 100 \approx 85.2$ males per 100 females.
2. Proportion female $= \dfrac{5{,}400}{10{,}000} = 0.54$ (54%).
3. Burglary rate $= \dfrac{27}{10{,}000} \times 100{,}000 = 270$ per 100,000 residents.
4. Percentage-point change: $6 - 8 = -2$ p.p. Percent change: $\dfrac{6-8}{8} \times 100 = -25\%$.
5. The headline is **imprecise**: the rate fell by 5 *percentage points*, not 5 percent. The true relative decline is 25%.

Exercise 3

A researcher studies how income relates to years of education (X), using a model of **log income**:

$$\ln(Y) = 9.00 + 0.08\,X$$

where Y is annual income in dollars.

1. For a respondent with $X = 12$ years of education, compute $\ln(Y)$.
2. Convert $\ln(Y)$ into Y (dollars) by exponentiating. (Hint: use the inverse of $\ln$.)
3. Now consider a respondent with 13 years of education. Compute their income.
4. The exponentiated coefficient on X is $e^{0.08} \approx 1.083$. Interpret this number in one sentence.

Solution: Exercise 3

1. $\ln(Y) = 9.00 + 0.08(12) = 9.00 + 0.96 = 9.96$
2. $Y = e^{9.96} \approx \$21{,}163$
3. $\ln(Y) = 9.00 + 0.08(13) = 10.04; \quad Y = e^{10.04} \approx \$22{,}926$
4. Each additional year of education is associated with income being multiplied by about **1.08** — roughly an **8% increase** in income, *on average*.