

Psych 10: Intro to Statistics

Lecture 2 · Sampling & measurement · Mon Sep 28, 2026

Prof. Judith Fan · Fall 2026

First, a word from Morgan



Morgan Tompkins · TA,
Wed 11:30 section

- ✍️ Please write your **full name** legibly on every worksheet and case file. If we can't read your name, we can't give you credit.
- 🆔 Your **SUNet ID** is the part before **@stanford.edu** in your Stanford email. It is not your 8-digit student ID number.
- 🕒 Your **section** is its day and time, like TUES 9:30.

Reminders & Announcements

- If you just joined the class, welcome! Check Canvas for the Week 1 announcements. If the syllabus doesn't answer your question, post on Slack (e.g., in the **#psych10-fall2026-questions** channel).

Week 2 at-a-glance

- Please fill out the **student background survey**, if you haven't done it yet (about 20 minutes, on a laptop). Doing so will make you eligible for 0.5% of extra credit.
- CourseKata **Chapter 3** due this Friday.
- Friday: **Case #1**, *The Sleep Lab Incident*, in your small-group tutorial session with your discussion leader. Time slots go up no later than Thursday morning. Just bring a writing utensil!

Tell us about yourself.

Student Background Survey

We ran out of time for this last week, so here it is now.

The primary purpose of this survey is for the teaching team to get to know you, how you think about learning, and relevant aspects of your circumstances that may affect your learning experience in this course.

The second purpose is to contribute to educational research on how people learn statistics and programming. Before you decide to participate, you should know why the research is being done, how your data will be used, and what your options are.

Meet Erik Brockbank



- 🧪 Erik studies how people explain how we get to know people, such as by interacting with them in conversation.
- 🎓 Postdoctoral researcher here in Psychology.

Department of Psychology

You are being invited to participate in a research study

to understand how people learn statistics / computer programming.

Dr. Erik Brockbank, a researcher in the Psychology department who is not involved with this course, will tell you more about this study and your options for participating.

The research study

In this course, we will administer various surveys and/or assignments to gain insight into your learning experience in this course, as well as what you are learning. By consenting to participate in this study, insights gained from the information you provide in these surveys/assignments might be used to improve your learning experience as well as that of students in similar courses. We are only recruiting students 18 years of age or older to participate in this study.



info sheet

Time Involvement: This academic quarter; your participation will **not** involve any additional work beyond what is expected to be completed as part of your enrollment in this introductory statistics course.

Payments: There is no monetary compensation for participating in the study.

Risks, benefits, and your rights

Risks and Benefits: Your participation is completely voluntary. There are neither specific benefits nor anticipated risks associated with participation in this study. You can withdraw at any time by completing the opt-out form (on next slide).

Data will be securely stored and de-identified. We cannot and do not guarantee or promise that you will receive any benefits from this study. While there is no direct benefit to you for participating in the study, the surveys and other learning materials that are being used as part of this study were designed with benefits to your learning experience in mind. Based on the results of this study, the principal investigator may learn more about how to improve student learning, and society may benefit from this knowledge.

Participant's Rights: Your decision whether or not to participate in this study will **not** affect your experience or grade in this class (the professor/teaching team will **not** know who is or is not participating in the study until *after* final grades are assigned).

Opting out

If you agree to participate in this research, no action is needed. If you DO NOT consent to participate in this study, or you choose to opt out at any time during the quarter, please submit this form online (QR code or <https://tinyurl.com/p10optout>).

Again, your instructor(s) will not have access to the list of students who have opted out until after final grades are posted.



opt-out form

The survey

If you haven't filled it out yet, please do so before lecture on Wednesday. You can expect it to take about 20 minutes, in one sitting, on a laptop. The link is on the syllabus page at **psych10.github.io**.

A good-faith response earns 0.5% of extra credit. You'll receive that extra credit whether or not you take part in the research study.

Today

Sampling & measurement

1. Exploring data: variation, and where data come from
2. Measuring by sampling
3. Measurement & different kinds of variables
4. A surprising survey result: what could have produced it?

1 Exploring data

Variation is everywhere

Look around the room: heights, how much sleep people got last night, commute times, phone habits, how people feel about this class. No two of you have the same numbers.

Our job as scientists is to ask questions about where that variation comes from!

Where do data come from?

- To study any of that, we have to **measure** it — attach a number or a category to each person.
- And we have to **sample** — we can't measure everyone, so who ends up in the data matters.

Every dataset is the result of two decisions someone made: what to measure, and whom to include.

Four rows of a class survey (example data)

ID	sleep_hours	chronotype	phone_checks
S01	6.5	owl	41
S02	8.0	lark	17
S03	7.2	owl	29
S04	5.9	owl	55

- Each **row** is one person (a “case”). Each **column** is a variable. Each **cell** is one value.
- Numbers that are amounts (hours, counts) are **quantitative**. Labels are **categorical** — even when they’re coded as numbers.

With your neighbor

Introduce yourself, then look at the four rows again.

- How many cases are there, and how many variables?
- Which columns are quantitative, and which are categorical?



What is a model?

- A **model** is a simplified description of the data that helps us figure out where the variation came from.
- Whatever produced the data is called the **data generating process**, or DGP.
- Lots of different processes can produce the same pattern. Our job is to list them, then think about what data would tell them apart.

Try it with your neighbor

Suppose a class survey finds that **far more students call themselves “night owls” than “morning people.”**

List **three different processes** that could have produced this observation. Make them genuinely different — not three versions of the same story. Then circle the one you find most plausible.

Let's hear some

Six kinds of explanation

1. The effect as described
2. The reverse direction
3. A third variable
4. Who ended up in the sample
5. How it was measured
6. Plain chance

2 Measuring by sampling

How much of the Earth's surface is covered by water?



How do we know??

how much of the earth's surface is water



Images

Videos

News

Shopping

Maps

Books

Flights

Finance

About 586,000,000 results (0.46 seconds)

71 percent

About **71 percent** of the Earth's surface is water-covered, and the oceans hold about 96.5 percent of all Earth's water. Nov 13, 2019



United States Geological Survey (.gov)

<https://www.usgs.gov> › water-science-school › science

[How Much Water is There on Earth? Completed - USGS.gov](#)



Imagine you
had a globe and
some process
for poking it in
random places
many, many
times.

**How often would you
expect to poke it in a
spot where there's
water?**



Worksheet - Part 1

We're about to toss this globe around the room. If you catch it, look at where your right thumb landed, toss the globe on to someone else, and come to the front of the room: the corner on your left if your thumb was on water, the corner on your right if it was on land.

- a.** Your prediction: what fraction of catches will be "water"?
- b.** Name one thing that could make this process *biased*, meaning it would give us the wrong answer even after hundreds of tosses.

Spoiler alert

Random processes can generate data that reveal underlying patterns!

Worksheet • Part 1, continued

c. Our class result: _____ water out of _____ catches. Is that closer to your prediction or to the real answer (about 71%)? Write one sentence on why our class result might differ from the real answer.

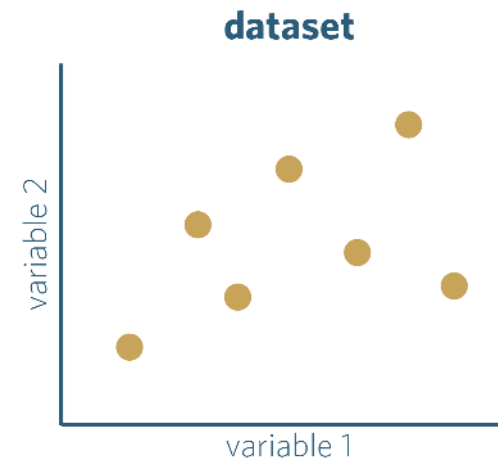
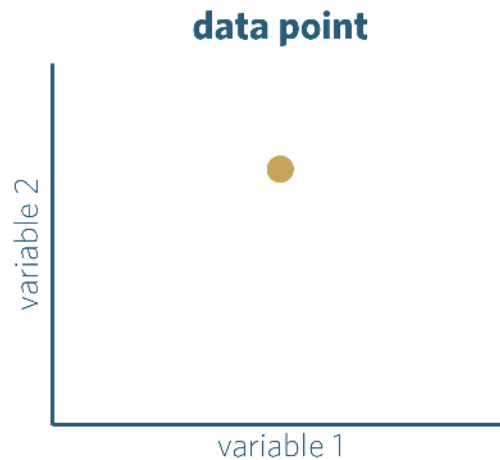
Population, sample, and sampling variation

- The **population** is every point on the surface of this globe. (Not the Earth — the globe's map projection is already one source of bias.)
- A **random sample** means every point is equally likely to end up under your thumb. Is it? Where do people tend to catch a ball?
- **Independent** means one catch doesn't affect the next. (Is the thrower aiming?)
- **Sampling variation:** Our tosses won't give you exactly 71%. Do it again and you'll get a different number. That's not a mistake — it's what samples do.

3 How do we measure things in psychology?

What are data?

- “Data” is a plural term. (“Datum” is the singular form. “Data point” is a good way to refer to a single observation.)
- A **dataset** is composed of **variables**. For each variable, there are usually multiple **values** that reflect measurements of some specific thing.



Types of data

Categorical

- **nominal** — e.g., what language do you speak at home?
- **ordinal** — e.g., rate your mood from 1 to 5

Quantitative (a.k.a. Numerical)

- **discrete** — e.g., how many marbles are in the jar?
- **continuous** — e.g., how much time has passed?
 - **interval:** equal steps, but no true zero (temperature in °F — 40° isn't "twice as hot" as 20°)
 - **ratio:** equal steps *and* a true zero (height, duration — 40 really is twice 20)

Notice the progression

	nominal	ordinal	interval	ratio
countable: number of times observed	✓	✓	✓	✓
ordered values, but the gap between them is not necessarily equal	✗	✓	✓	✓
quantitative values: can be added or subtracted	✗	✗	✓	✓
has a “true zero”: can be multiplied and divided	✗	✗	✗	✓

Two kinds are enough most of the time

For most of what you'll do, two kinds cover it: **quantitative** (the numbers are amounts, and differences between them mean something) and **categorical** (the numbers, if there are any, are just labels).

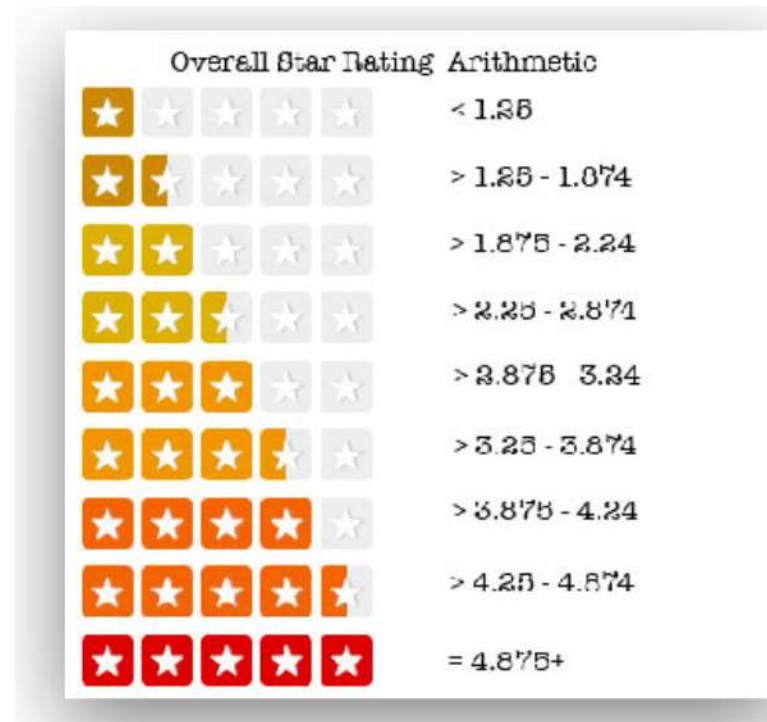
The one that causes trouble is **ordinal**. The values are written as numbers, but the gaps between them aren't equal, so they work more like ordered labels.

Beware of treating ordinal data like interval or ratio data

...without thinking about whether it is appropriate!



"My average iPhone model is 6.33!"



"This restaurant has a 4.183 rating!"

Worksheet - Part 2

For each variable below, decide whether it is **quantitative** or **categorical**, then which of the four levels it is: **nominal**, **ordinal**, **interval**, or **ratio**. Say why in a few words.

1. Bedtime last night (for example, 11:30 pm)
2. Finishing place in a 5K (1st, 2nd, 3rd, ...)
3. Course number (PSYCH 10, PSYCH 250)
4. Earthquake magnitude (for example, 4.5)
5. Birth year

Worksheet · Part 2, continued

Suppose the class survey from earlier recorded chronotype as 1 = lark, 2 = owl, and a friend averages the column and gets 1.6. “*So the class is more owl than lark,*” they say. **True or false?** Justify your answer in one or two sentences, using what the numbers *mean*. Would your answer change if the survey had also offered 3 = neither?

Takeaway

- Two easy mistakes: “it’s a number, so it must be quantitative,” and “it’s just a label.”
- Ask what the values mean and what you’re going to do with them. The same column can be either kind. What you want to think about is what feature of this variable matters for your question.
- Every measurement involves thinking about the units being used and the scale it’s on.

Missing Data

- Sometimes a dataset does not contain all of the values it could (or that you would expect it to).
- Pay attention to & think about why those values might be missing, and how different choices of how to handle those missing values could affect your ability to answer the scientific question at hand.

4 A surprising survey result

Worksheet - Part 3

Suppose a campus-wide survey run by *The Stanford Daily* found **far fewer first-generation students than the admissions office reports.**

- a.** List **three different processes** that could have produced this result. Make them different from one another, not variations on one explanation.
- b.** Circle the one you find most plausible.
- c.** Pick one of the others. What data, existing or new, would help you tell that process apart from the one you circled?

Let's hear some

FAQ: Why does lecture cover stuff other than what is in CourseKata?

Good question!

My goal in lectures is to provide fuller context for the ideas you're learning about in the assigned CourseKata chapters, and the skills you're putting into practice in section and on Fridays.

Rather than spending all our time re-hashing what you can find elsewhere, I think our time here is better spent drawing broader connections. We'll review CK material some of the time, just not all of the time. We're going for something else here!

Before Friday

- Tonight: the **student background survey**, if you haven't done it yet. Link on the syllabus page at psych10.github.io.
- CourseKata **Chapter 3** (distributions & histograms) before Friday. Wednesday's lecture assumes you've at least started it.
- Friday: **The Sleep Lab Incident**. Check Thursday's post for your time slot.

Hand in your worksheet on the way out

On the last two lines, write down something new you learned today, and let us know if there's anything you'd like us to go over again. Thanks!