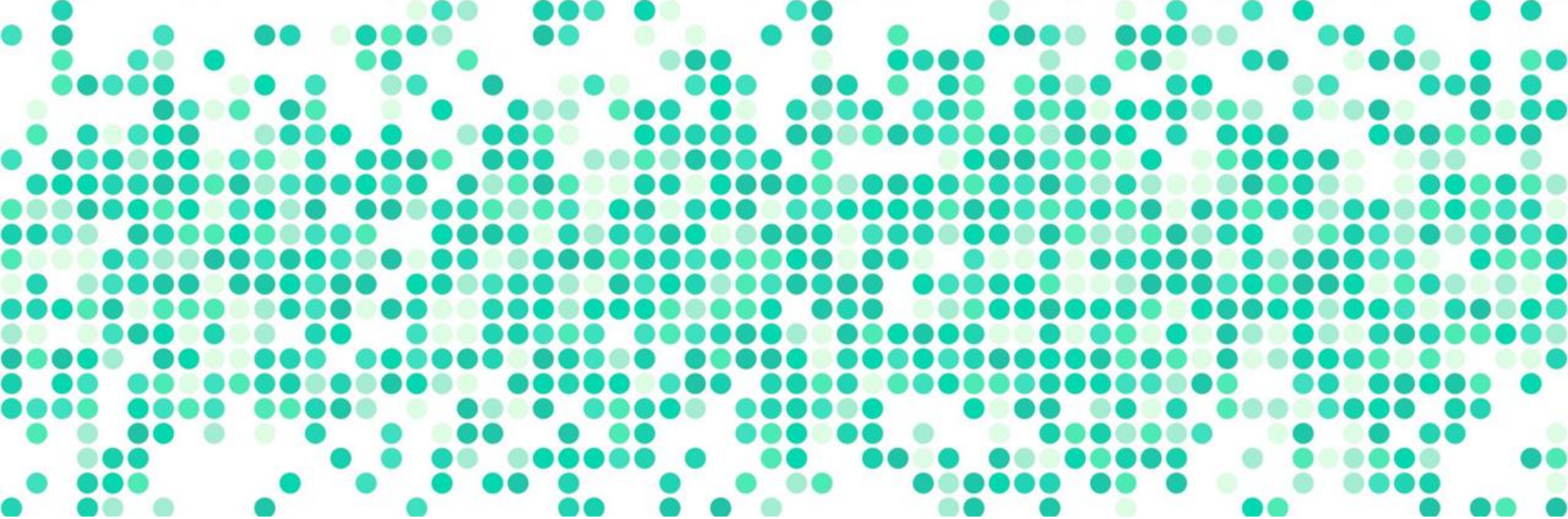


DATA ANALYSIS

Week 1: Statistical Thinking / What are data?

logistics

Before/End of Tuesday

- Read the syllabus for this course. If you have questions about the syllabus, then please ask them in class or on the pre-class survey.
- Complete the Pre-class survey!

Before Thursday

- Watch: Introduction to Google Sheets. Leave an annotation to complete participation.
- Read Chapter 1 from the Gravetter & Wallnau (2017) textbook. Leave an annotation for any questions you have.
- Take the Skills Assessment Exam from the textbook (Appendix A) and submit a reflection. This will help me support you better in the course.

Here are the to-do's for this week:

- Submit Week 1 Quiz (due Sunday midnight)
- Start working on Problem Set 1 (Chapter 1 problems)
- If you haven't already, take the Skills Assessment Exam from the textbook (Appendix A) and submit a reflection. This will help me support you better in the course.
- Submit any lingering questions here!
- Extra credit opportunities:
  - Submit Pre-class survey (1 point)
  - Submit Exra Credit Questions (1 point for 8 submissions)
  - Submit Optional Meme Submission (1 point for winners!)

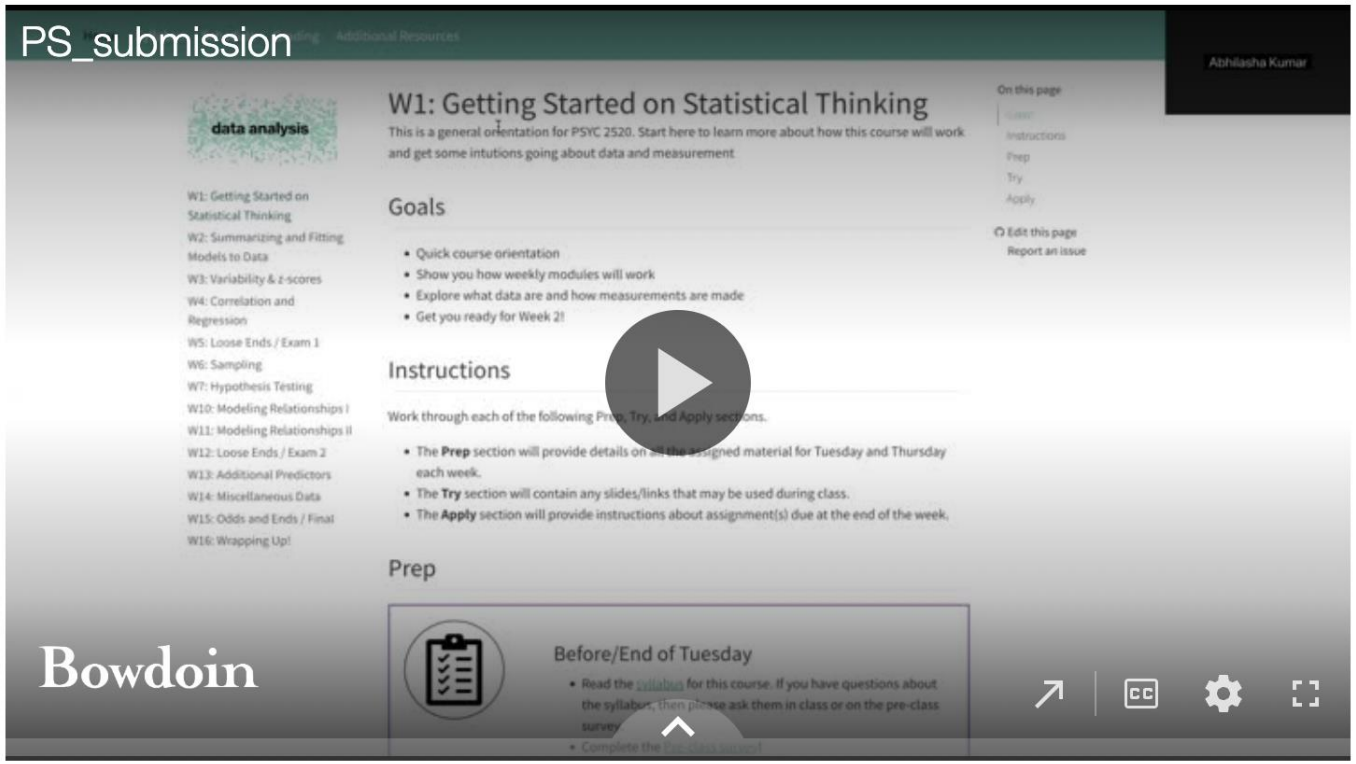
Problem Set 1: First Attempt

Start Assignment

Due Feb 3 by 11:59pm **Points** 2.5 **Submitting** a file upload **File Types** pdf

Please complete the problem set available here. **Please submit a PDF of your solution sheet.**

- Make sure you are looking at the correct problem set.
- Please follow the template provided in the link above for the specific problem set.
- **Please watch the video below** to make sure you are following the submission guidelines for problem sets.



Problem Set 1 (summarizing & means)

Attempt 1 due date: **Feb 3, 2025**

PS1: Solution Template **[Use this template to create your own solution sheet]**
PS1 worksheet template **[Use this template to create your own worksheet]**

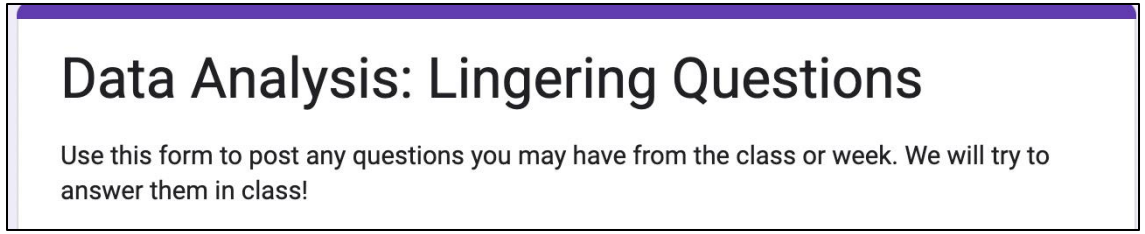
Please watch this video that describes how to **submit** problem sets

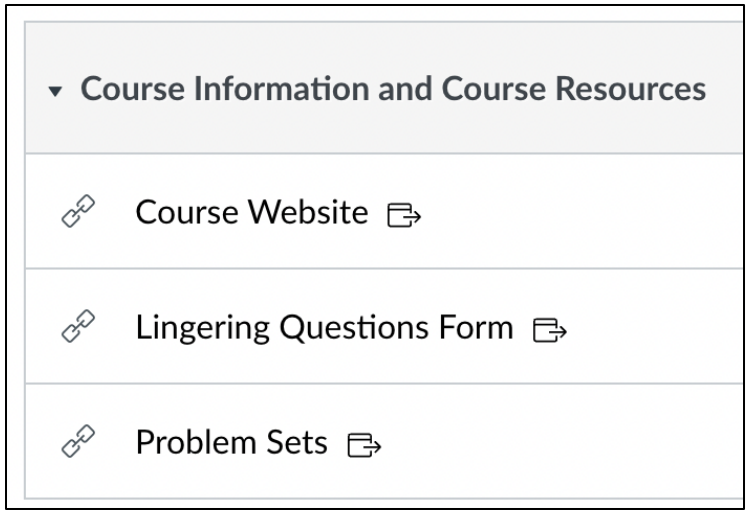
Total number of problems (including sub-parts): 32
75% cutoff for a reasonable first attempt: 24

- Chapter 1 Problems: 8, 10, 18, 20, 22
  - 22a has a typo: it should be “add the scores and then square the sum”
- Chapter 2 Problems: 4, 6, 12, 14, 18,
- Chapter 3 Problems: 10, 12, 14, 20, 22

more Qs

- Is there flexibility for students with accommodations?
- Do we need to buy the textbook? Can we use it online.
- What is your favorite part about teaching data analysis?





Here are the to-do's for this week:

- Submit Week 1 Quiz (due Sunday midnight)
- Start working on Problem Set 1 (Chapter 1 problems)
- If you haven't already, take the Skills Assessment Exam from the textbook (Appendix A) and submit a reflection. This will help me support you better in the course.
- Submit any lingering questions here!
- Extra credit opportunities:
  - Submit Pre-class survey (1 point)
  - Submit Exra Credit Questions (1 point for 8 submissions)
  - Submit Optional Meme Submission (1 point for winners!)

today's agenda

introduce statistical thinking

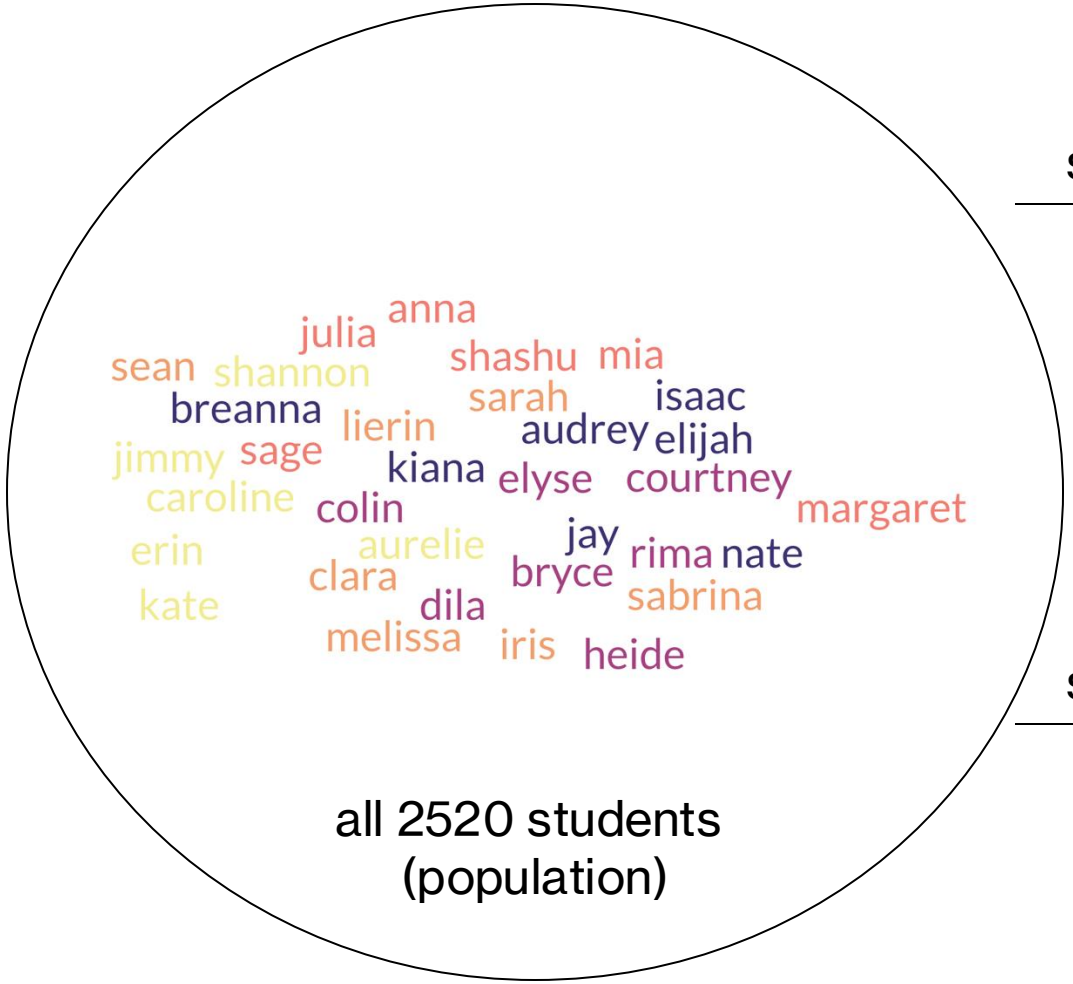
define population / sample / data

discuss scales of measurement / reliability

what is statistical thinking?

- understanding the complex world in simple terms
  - summarization + uncertainty
- different from other forms of thinking, e.g., human intuition, heuristics, etc.
- three key uses: describe (the world), decide (something), predict (something)
- key concepts:
  - learning from data: we let the data guide us
  - aggregation: we “summarize” raw data
  - uncertainty: we assess how well our raw data maps on to the summarization
  - sampling: we acknowledge that our data are samples from a population

populations and samples



sampling

sample 1

Melissa
Iris
Nate
Isaac
Julia
Sage

sampling

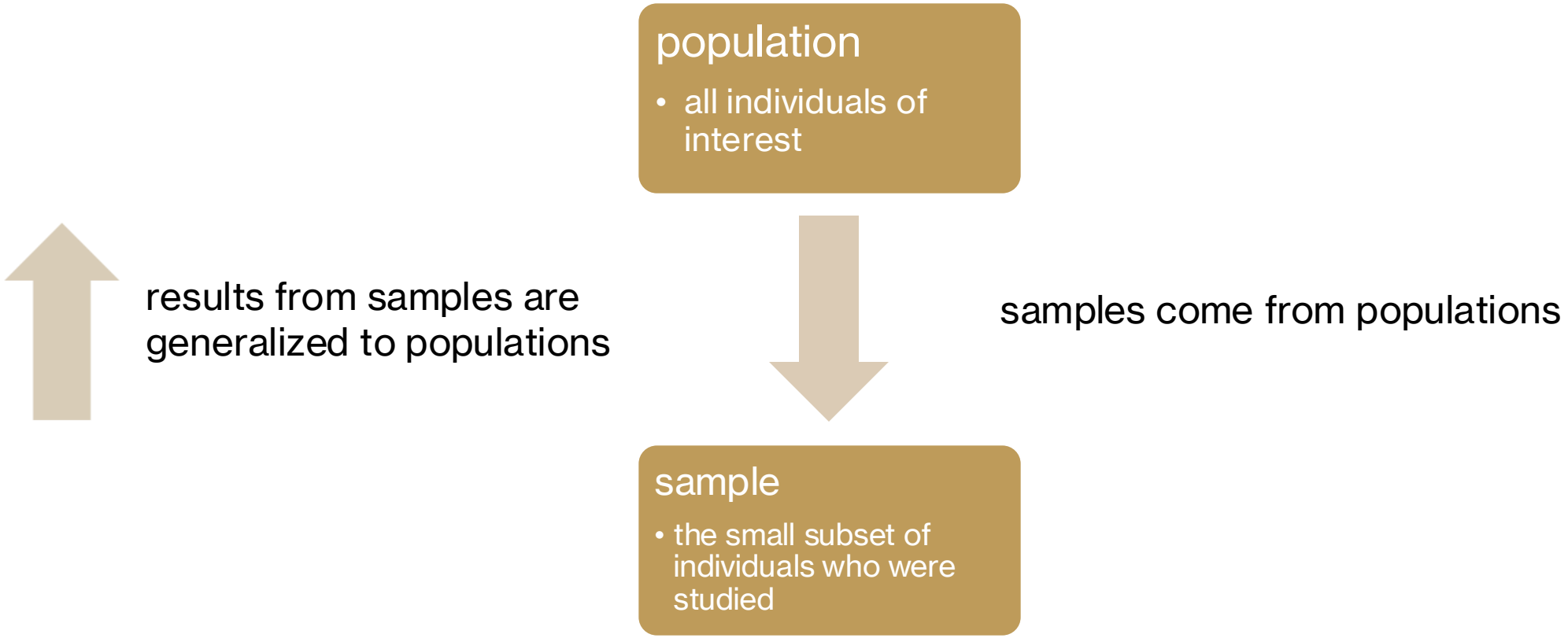
sample 2

Audrey
Kiana
Lierin
Sarah
Jimmy
Elyse

samples should be

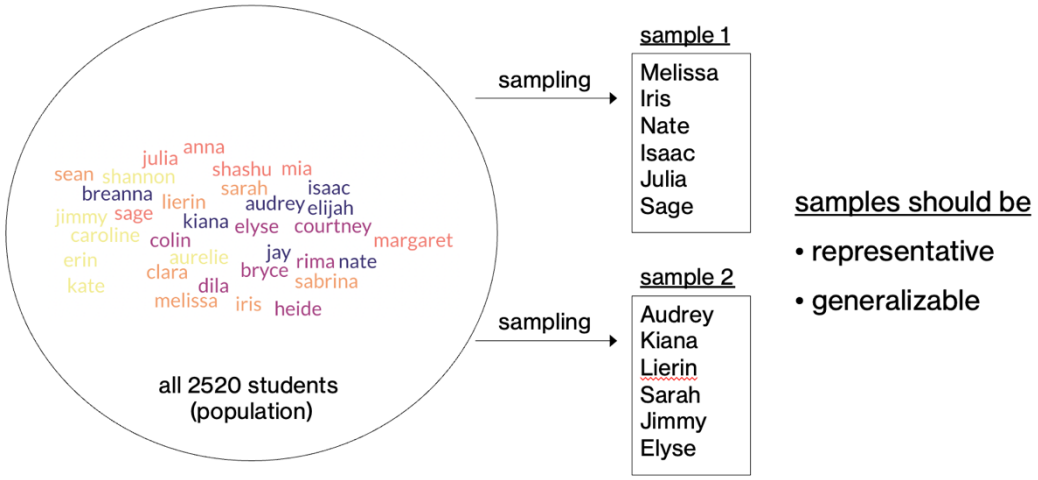
- representative
- generalizable

populations and samples



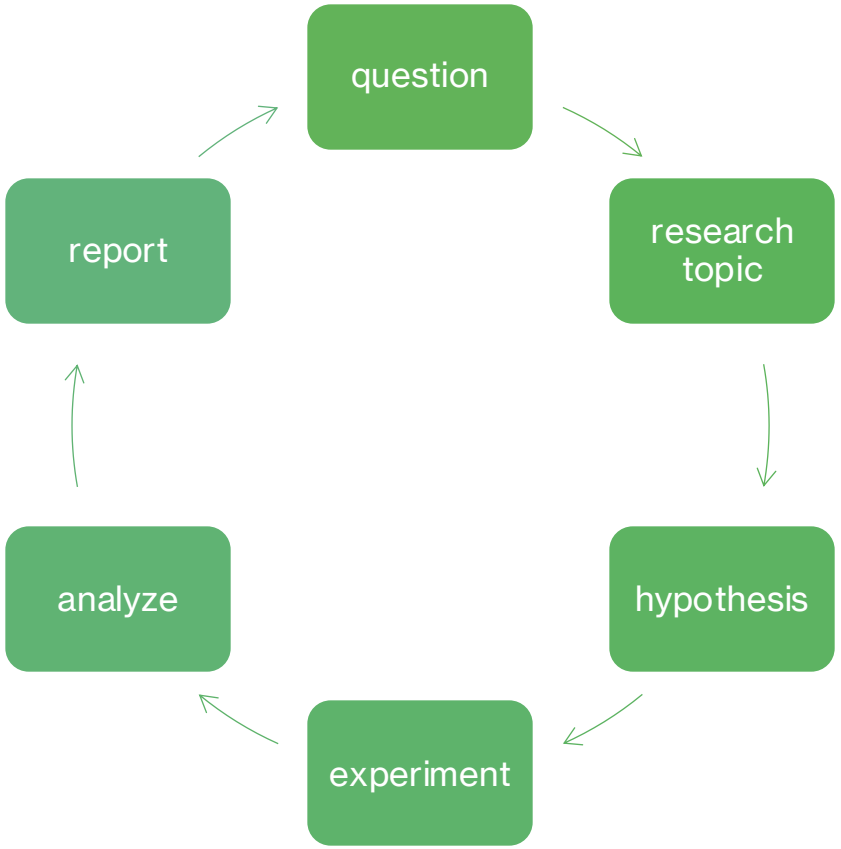
parameters, statistics, sampling error

- **p**arameter: something that describes a **p**opulation
- **s**tatistic: something that describes a **s**ample
- sampling error: the discrepancy between the sample statistic and the true population parameter it is estimating
- to reduce sampling error:
  - use a sufficiently large sample
  - use random selection: selecting individuals from the population at random for your sample to create an unbiased sample



the scientific method

- the scientific method is a method for acquiring knowledge by making predictions, carrying out experiments to test those predictions, and making inferences based on the observed outcomes
- constructs, variables, and constants
  - construct: something intangible that we operationalize
  - variable: a characteristic that changes across conditions
  - constant: a characteristic that is fixed across conditions
- to make inferences, we manipulate a variable of interest, and observe the effect on an outcome variable, holding all other variables constant



samples in research

experimental research

- test a manipulation to establish a cause-and-effect relationship between two variables

non-experimental research

- quasi-experimental research
  - no actual manipulation, groups/variables defined due to natural variations
- descriptive research
  - single or collection of variables are observed and summarized
- correlational research
  - at least two variables are observed to determine a relationship

research terminology: review

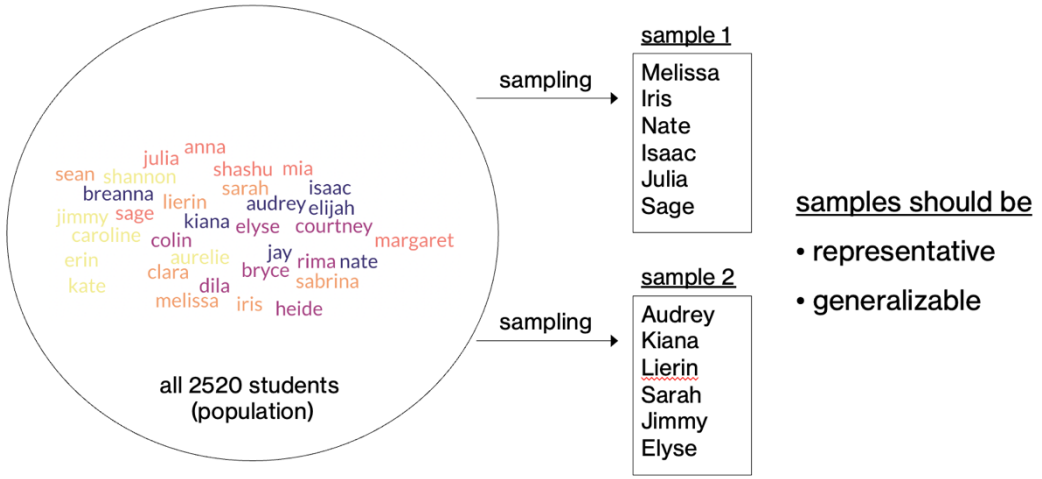
- independent variable (what is being manipulated?)
  - levels denote the types of “conditions” that a participant could be assigned to
- dependent variable (what is being measured?)
- design type (within- or between-subjects/participants)
  - were all participants exposed to all levels of the independent variable?
- key ideas for controlling other extraneous variables:
  - random assignment
  - matching/holding constant
  - control conditions

practice scenario

- A clinical psychologist is interested in the effectiveness of a new anti-depression drug. He collects depression scores from a group of individuals diagnosed with depression at time 1. All individuals then take the drug, and are measured again a month later at time 2.
  - what kind of study is it (experimental / non-experimental)?
  - independent and dependent variables?
  - design type (within- or between-participant)?
  - what would the data look like? what would a plot of results look like?

from samples to data

- samples provide us with information
- data ***are*** measurements or observations obtained from a sample
  - a dataset is a collection of measurements or observations
  - a datum is a single measurement or observation

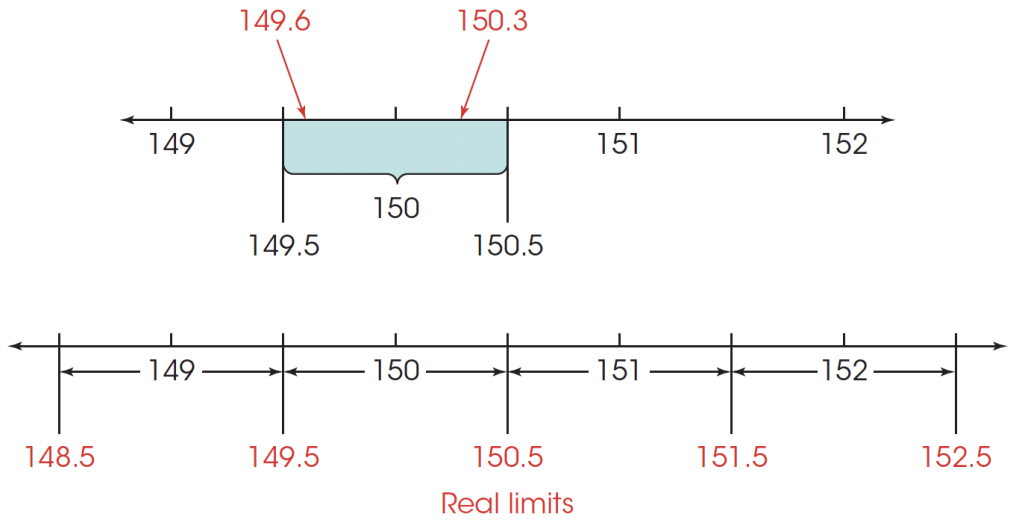


scales of measurement

- data can be measured in several ways:
  - qualitative (put things into categories) vs. quantitative (assign numbers) data
  - discrete: separate, indivisible values. no values can exist between two neighboring values; integer scales
  - continuous: an infinite number of possible values fall between any two observed values. hypothetically divisible into an infinite number of fractional parts.
- how data are measured determines:
  - what kinds of mathematical operations can be applied
  - what kind of statistical computations can be computed

real limits for continuous data

- only applies to continuous data
- the real limit separates two adjacent scores, and is located halfway between the scores
  - each score has an upper real limit (UL) and a lower real limit (LL)
  - lower limit for 150 is 149.5; upper limit is 150.5



scales of measurement

NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	**magnitude**	**equal intervals**	**absolute zero**
nominal				
ordinal				
interval				
ratio				

scales of measurement

NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	**magnitude**	**equal intervals**	**absolute zero**
nominal	✅			
ordinal	✅	✅		
interval	✅	✅	✅	
ratio	✅	✅	✅	✅

practice

- A researcher is testing the effect of alcohol on memory performance. He randomly gives one group of subjects a bottle of vodka, and another a nonalcoholic substance that tastes like vodka. Each group then learns a list of words, and attempts to recall them. Number of words correctly recalled for each group is recorded
  - what is the scale of the independent variable?
  - what is the scale of the dependent variable?

activity

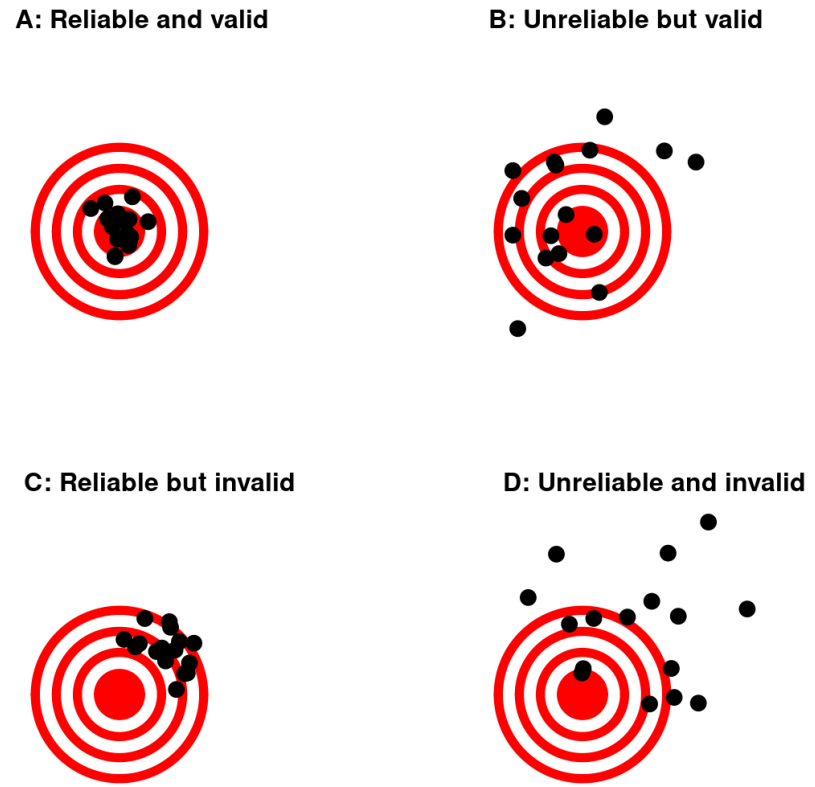
NOIR	each value has a unique meaning	a value has a sense of quantity, some values are larger, some are smaller	units along the scale of measurement are equal to one another	the scale has a true meaningful zero point
	identity	**magnitude**	**equal intervals**	**absolute zero**
nominal	✅			
ordinal	✅	✅		
interval	✅	✅	✅	
ratio	✅	✅	✅	✅

- assign a data type to each variable (NOIR) and whether it is discrete / continuous

variable	NOIR	discrete/continuous
numbers on basketball jerseys		
sizes of Starbucks orders		
weight		
calendar years		
IQ scores		

reliability and validity

- reliability: consistency of measurements
  - test-retest reliability
  - inter-rater reliability
- validity: are we measuring what we think we are measuring?
  - ***face*** validity: reality check, does it make sense?
  - ***construct*** validity: is it related to other measurements in a logical manner? convergent vs. divergent validity
  - ***predictive*** validity: can it predict future data?



next time

- why/how do we summarize data?
- how do we “explain” data?

Prep

Before Tuesday

- Watch: Summarizing Data.
- Read Chapter 2 from the Gravetter & Wallnau (2017) textbook.

Before Thursday

- Watch: Central Tendencies.
- Read Chapter 3 from the Gravetter & Wallnau (2017) textbook.

After Thursday

- See Apply section.

Here are the to-do’s for this week:

- Submit Week 1 Quiz (due Sunday midnight)
- Start working on Problem Set 1 (Chapter 1 problems)
- If you haven’t already, take the Skills Assessment Exam from the textbook (Appendix A) and submit a reflection. This will help me support you better in the course.
- Submit any lingering questions here!
- Extra credit opportunities:
  - Submit Pre-class survey (1 point)
  - Submit Exra Credit Questions (1 point for 8 submissions)
  - Submit Optional Meme Submission (1 point for winners!)

bonus practice: scenario

- A social psychologist is interested in gender differences in math performance. She randomly selects students from Bowdoin and has them solve a series of equations. Number of equations correctly solved for each participant is recorded.
  - what kind of study is it (experimental / non-experimental)?
  - independent and dependent variables?
  - design type (within- or between-participant)?
  - what would the data look like? what would a plot of results look like?

bonus practice: data in abstracts

- table groups
- go to the abstract document and read over the abstract
- make note of (you will need to make a copy to edit the document):
  - independent variable(s) and data type(s)
  - dependent variable(s) and data type(s)
- predicted graph of results?
- key takeaway?