PSY 201: Statistics in Psychology

Lecture 02
Measurement scales
Descriptive statistics
What is our national security threat?

Greg Francis

Purdue University

Fall 2023

VARIABLES

factors that affect data e.g.
study performance of college students taking a statistics course
variables include

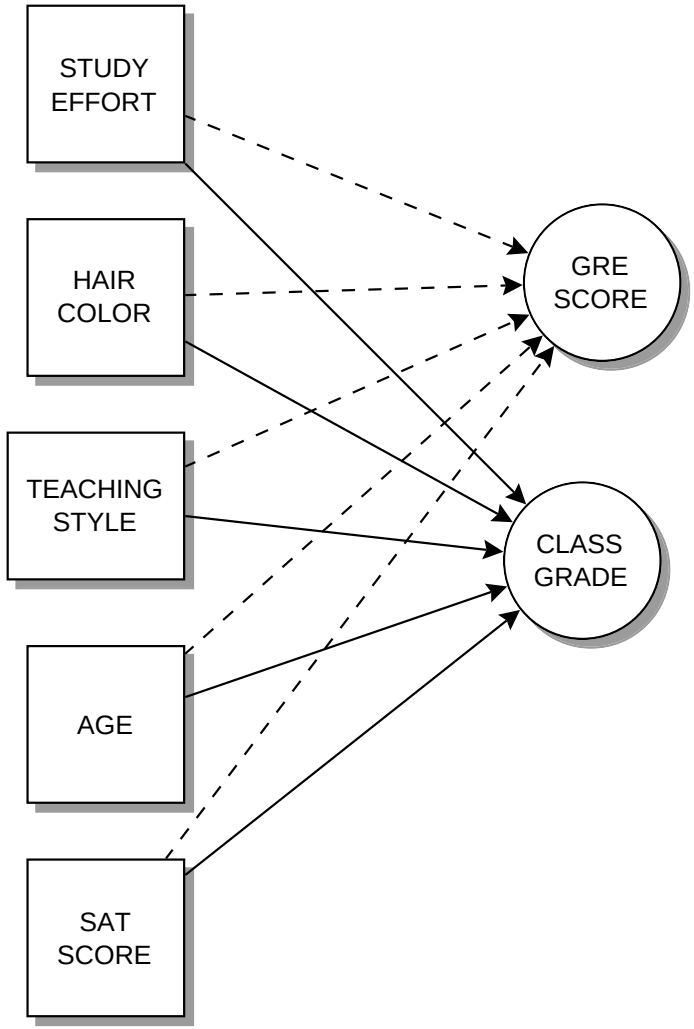
- teaching style
- age
- SAT scores
- class grade
- study effort
- hair color...

DEPENDENT VARIABLES

e.g.

class grade

GRE scores



DEPENDENT VARIABLES

- researchers are interested in how dependent variables change as other variables change
  - see how the dependent variables *depend* on other variables)
- other variables are called **independent** variables
  - researcher either keeps track of or controls the values of independent variables

INDEPENDENT VARIABLES

two types

1. researcher manipulates variable
   e.g. drug dosage, teaching style,...
2. variable classifies
   e.g. hair color, eye color, SAT scores,...

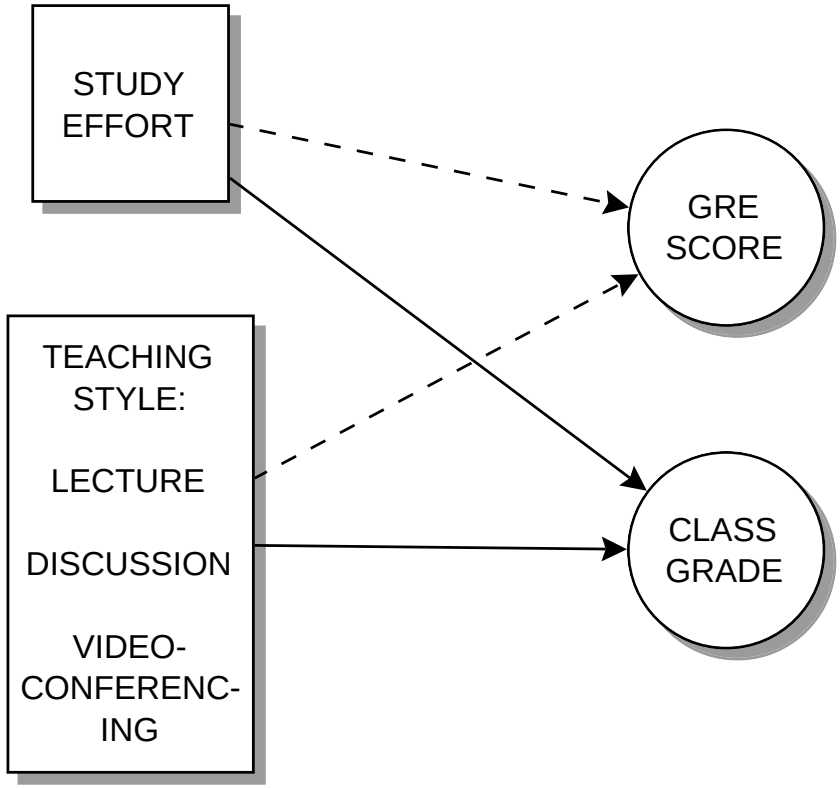
study wants to know how the dependent variable *changes* with changes in the independent variables

LEVELS OF VARIABLES

independent variables can have different **levels**

e.g.
three methods of teaching style

1. Lecture.
2. Discussion.
3. Videoconferencing



EXAMPLE

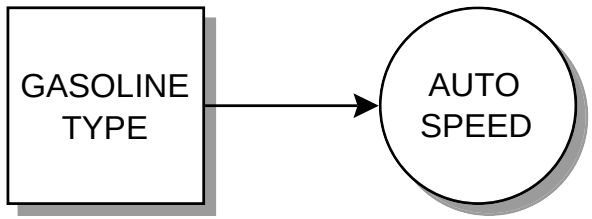
Does gasoline type affect car speed?

- take a car
  - fill it with different types of gasoline
  - measure top speed
- keep many things **constant**
  - same driver
  - same car
  - same course
  - same weather
  - ....

if you started changing these, they would become independent variables

VARIABLES

- independent variable: gasoline type
- levels of independent variable
  - Amoco
  - Sunoco
  - Crystal Flash
  - Marathon
  - ...
- dependent variable: auto speed



MEASUREMENT

- studies need to identify variables and measure them
- different variables have different scales of measurement
- four scales of measurement:
  least precise to most precise
  - nominal
  - ordinal
  - interval
  - ratio

NOMINAL SCALE

- classification of objects into categories
- e.g.
  - nationality
  - color of eyes
  - gender
  - names of objects
- no **order** to the categories!

NOMINAL SCALE

- two key properties
  - data categories are mutually exclusive.
  - data categories have no logical order.
- numbers can designate categories
  1. blue eyes
  2. brown eyes
  3. green eyes
- but the order of numbers does not imply order of categories, because there really is no order

ORDINAL SCALE

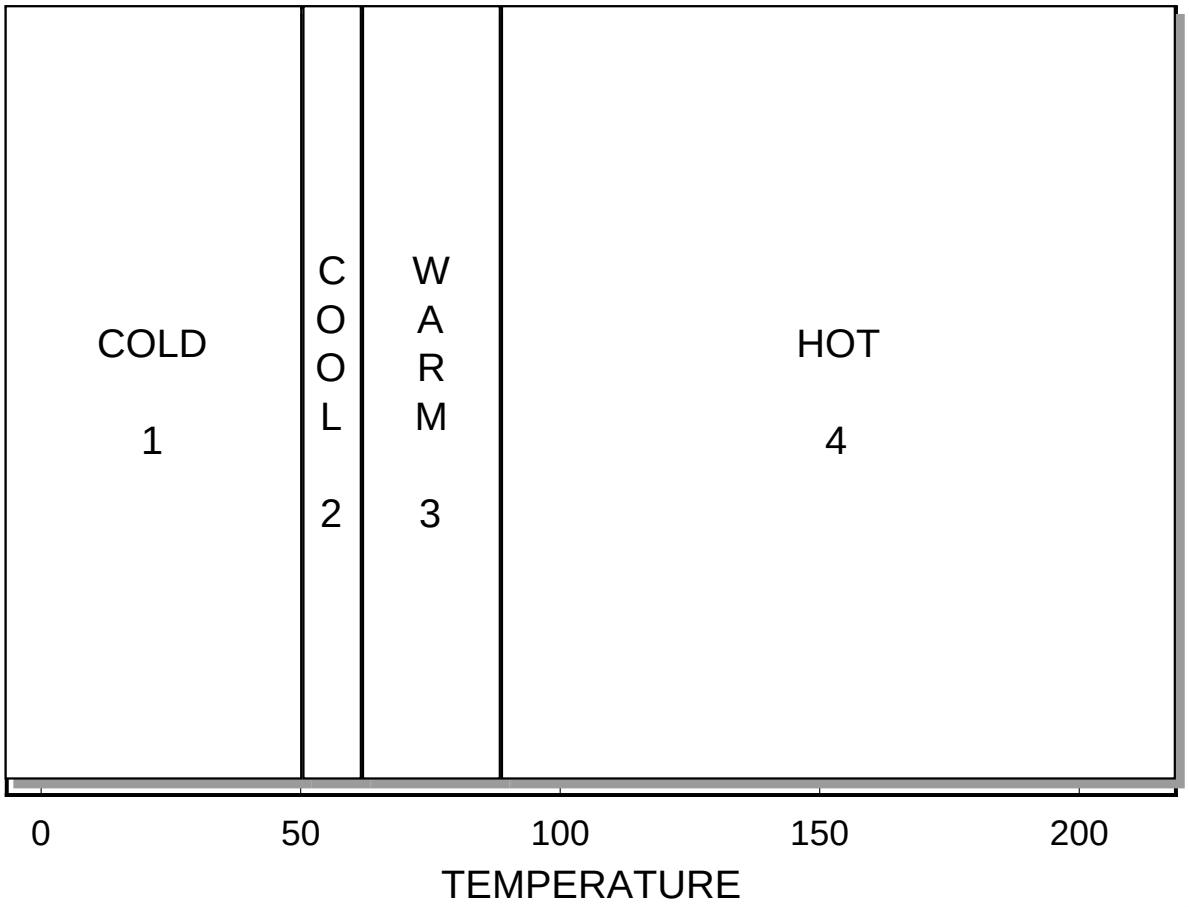
- ordered classification
- e.g.
  - grading system A,B,C,D,F
  - warmth: cold, cool, warm, hot
  - aggressive, timid
- **order** is important and means something

ORDINAL SCALE

- numbers can be used to designate categories
- e.g. warmth
  1. cold
  2. cool
  3. warm
  4. hot
- order of numbers agrees with order of categories

ORDINAL SCALE

- but size of number does not correspond to **amount** of relevant characteristic
- e.g., warm (3) does not necessarily have 2 more units of warmth than cold (1)



ORDINAL SCALE

characteristics

- data categories are mutually exclusive.
- data categories have some logical order.
- data categories are scaled according to the amount of the particular characteristic they posses.

USING SCALES

- One needs to pick items that have a "natural" scale to convey certain types of information
- Thus, for example, colors are typically at the nominal scale of measurement
- this makes them a poor choice for labeling of ordinal data because people do not automatically know what the different colors mean
- this was a problem for the National security warning system, which used colors to indicate different threat levels
- Which is more severe: green threat or blue threat?

MATCHING SCALES

- The problem was that the scales of threat (ordinal scale) and color (nominal scale) do not match. Thus, news reports of the threat level invariably do not list only the color but also the associated phrase with each report. The color scale was of no use at all (they were dropped in 2011).

INTERVAL SCALE

- equal unit scale
- e.g.
  - temperature (Fahrenheit or Celsius)
  - IQ scores (try to be)
  - most tests
- no beginning to scale
- zero point is just another category

INTERVAL SCALE

- numbers can be used to designate categories
  e.g.
  - 22° F $\rightarrow$ level of heat
  - 25° F $\rightarrow$ level of heat
  - 28° F $\rightarrow$ level of heat
- order of numbers agrees with order of categories
- number differences agree with characteristic differences (e.g., 3° F)

INTERVAL SCALE

- Intelligence quotient scores
  - 50 IQ
  - 100 IQ
  - 150 IQ
- an adult with a 50 IQ should have 50 fewer units of intelligence than a person with a 100 IQ
- a person with a 100 IQ should have 50 fewer units of intelligence than a person with a 150 IQ
- however, you cannot say that a genius (150 IQ) is 1.5 times as intelligent as an average (100 IQ)

INTERVAL SCALE

- zero point
- 0 temperature does not mean no heat (in F and C)
- 0 IQ does not mean no intelligence
- 50° F is *not* twice as hot as 25° F.
- an IQ of 100 is *not* twice as smart as an IQ of 50

WHY ZERO MATTERS

- I can create an equivalent interval scale that preserves all the differences

$$\text{NEW}_{\text{IQ}} = \text{OLD}_{\text{IQ}} + 20$$

- differences are still the same
  - $150 \rightarrow 170$
  - $100 \rightarrow 120$
  - $50 \rightarrow 70$
- but the ratios are all different 170 is not 1.5 times 120! Multiplication makes no sense!
- if zero meant absence of trait, I could not create an equivalent interval scale, zero would have to correspond to zero, and nothing else.

INTERVAL SCALE

characteristics

- data categories are mutually exclusive.
- data categories have some logical order.
- data categories are scaled according to the amount of the particular characteristic they posses.
- equal differences in the characteristic are represented by equal differences in the numbers.
- the value 0 is just another value on the scale.

RATIO SCALE

- what we normally think of as measurement
- e.g.
  - height
  - weight
  - energy
- zero point corresponds to the lack of a characteristic

RATIO SCALE

- numbers can be used to designate categories
  e.g.
  - 25 meters $\rightarrow$ distance
  - 5 meters $\rightarrow$ distance
  - 0 meters $\rightarrow$ no distance
- order of numbers agrees with order of categories
- number differences agree with characteristic differences

RATIO SCALE

- Kelvin temperature scale measures heat energy
- e.g.
  - 0° K $\rightarrow$ no heat energy
  - 25° K $\rightarrow$ heat energy
  - 50° K $\rightarrow$ heat energy

RATIO SCALE

- zero point
- 0 distance means no distance
- 0° K temperature means no heat
- 50 meters *is* twice as far as 25 meters
- 50° K *is* two times as much heat energy as 25° K.

RATIO SCALE

- data categories are mutually exclusive.
- data categories have some logical order.
- data categories are scaled according to the amount of the particular characteristic they posses.
- equal differences in the characteristic are represented by equal differences in the numbers.
- the value 0 reflects the absence of the characteristic.

MEASUREMENT SCALE

- how do you identify what scale is appropriate?
- measures at a "higher" scale can also be used at a lower scale, but not vice-versa
- the correct scale often depends on how you intend to *use* the data, and not so much on the intrinsic properties of the things you measure
- e.g. I can use person *names* as
  - nominal scale (code different people)
  - ordinal scale (alphabetize by name)

SCALES

- qualitative variables: generally discrete categories
  - nominal data
  - ordinal data
- quantitative variables: generally continuous
  - interval data
  - ratio data
- sometimes data looks like it is qualitative when it is actually quantitative (e. g., temperature readings do not usually use decimals, but they could)

POPULATION

- all members of a specified group
- e.g.,
  - all students in this class
  - all Purdue students
  - all patients with Alzheimer's disease
- measure of a population characteristic is called a **parameter**
- e.g.,
  - mean grade in class
  - highest grade in class
  - lowest grade in class

SAMPLE

- a subset of all members of a specified group, e.g.
  - all students in this class, relative to all Purdue students
  - all Purdue students, relative to all college students nationwide
  - all Alzheimer's patients, relative to all ill patients
- measures of a sample characteristic are called **statistics**, e.g.,
  - mean grade in class
  - highest grade in class
  - lowest grade in class
- we will use a statistic to **infer** properties of the corresponding population

CONCLUSIONS

- variables
  - dependent
  - independent
- measurement scales
- important issues for interpreting data
- important for applying statistical approaches

NEXT TIME

- working with data
- displaying data
- summarizing data

Why the space shuttle Challenger *blew up.*