- **Descriptive statistics**
- **Inferential statistics**
- **Data graphics (1)**

---

*Background preparation:*

- *Kaplan (2016), Appendix, "Statistics brief reference"*

# 0. Overview

- Statistics
  - Descriptive
  - Inferential
    - + Role of probability

- Data graphics
  - How to make
  - How to read
  - As a pattern-finding step in an analysis

# 1. Descriptive statistics

- Purpose: To **summarize** data sets
  - Mean
  - Standard deviation
  - Correlation

# 1. Descriptive statistics

## Mean

- What is the mean of these values?

  4    4   2   4   16

  - How did you calculate this?

# 1. Descriptive statistics

## Mean

- What is the **concept** behind the mean?

# 1. Descriptive statistics

## Mean

- What is the **concept** behind the mean?
  - The amount each item would contribute to the total if all contributions were equal
  - Center of gravity
    - Picture a balancing beam with an equal weight at each value
    - The mean is the balancing point

# 1. Descriptive statistics

## Mean

- How is the mean (potentially) useful for:
  - **Describing** a data set?
  - Making **predictions**?

# 1. Descriptive statistics

## Mean

- How is the mean (potentially) useful for:
  - **Describing** a data set?
    - Describes the **central tendency** of the data set
  - Making **predictions**?
    - Gives us an “**expected value**” for future cases

# 1. Descriptive statistics

## Mean

- What are some potential **pitfalls** with using the mean in these ways?

# 1. Descriptive statistics

## Mean

- What are some potential pitfalls with using the mean in these ways?
- *Descriptive and predictive:* The mean might not resemble any actual value in the data set
  - Extreme **outliers** can skew the mean
    - Geography majors at UNC—the highest average salary after graduation?
  - The data set might be **bimodal** (age: parents and toddlers)

# 1. Descriptive statistics

## Mean

- What are some potential pitfalls with using the mean in these ways?
- *Predictive:* Were the items measured actually **representative** of their category?
  → *inferential statistics*

# 1. Descriptive statistics

## Standard deviation

- What does the standard deviation of a set of numbers indicate?

# 1. Descriptive statistics

- **Standard deviation** | Kaplan (2016: 266; my emphasis):

> The standard deviation reflects the **amount of ‘spread’** in the data: a small SD means that the numbers in the set are clustered tightly around the mean, while a larger SD means that the numbers are more spread out.
>
> As a rule of thumb, more than half of the numbers in the set will fall within one standard deviation of the mean, and the vast majority will fall within two standard deviations—
>
> but all this depends on the specific properties of the set, and there are no guarantees.

# 1. Descriptive statistics

**Standard deviation**

- How to calculate standard deviation (FYI only!)
  - *Deviance* (for each data point): Find the difference from the mean
  - *Sum of squared deviances*: Square each deviance value and add them up
  - Variance: Divide the sum of squared deviances by (the number of data points minus 1)
  - ***Standard deviation:*** Square root of the variance

A good source for basic statistics: *Concepts & Applications of Inferential Statistics*, by Richard Lowry—free online textbook at [*http://vassarstats.net/textbook/*](http://vassarstats.net/textbook/)

# 1. Descriptive statistics

## Standard deviation

- Why is it important? (Kaplan 2016: 266)

Table A.1 *Average pitch (fundamental frequency), duration, and loudness (intensity) of consonant-vowel-consonant words in sober and intoxicated speech.*

| | Sober | | Intoxicated | |
|---|---|---|---|---|
| Measure | *M* | (*SD*) | *M* | (*SD*) |
| Pitch (Hz) | 154.0 | (51.7) | 156.4 | (52.3) |
| Duration (ms) | 442 | (120) | 467 | (127) |
| Loudness (dB) | 78.2 | (5.8) | 75.1 | (5.3) |

→ Words *meaningfully* longer in intoxicated speech?

# 1. Descriptive statistics

## Correlation

- What is correlation?  What kinds of correlation are there?

- What does correlation look like in a **scatterplot**?

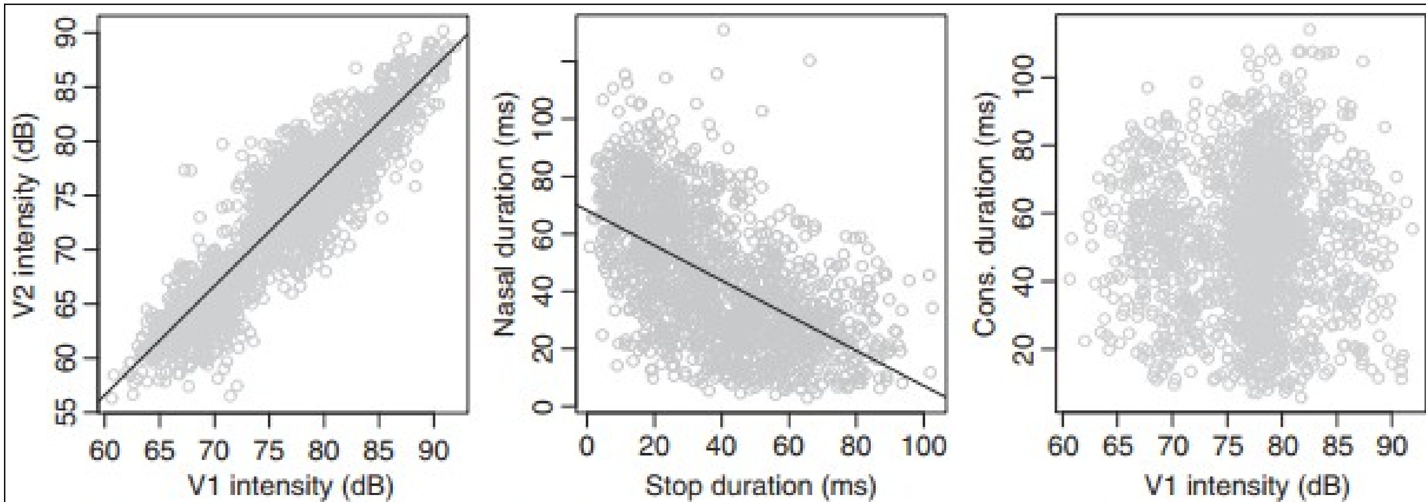
# 1. Descriptive statistics

## Correlation

- Correlation (*r*): Measures to what extent the value of variable *y* is predicted by the value of variable *x*
  - If we know how long each student studied for an exam, can we predict how well they did?
- Correlation can be positive, zero, or negative
  - Positive ($0 < r \leq 1$): when *x* increases, *y* increases
  - Zero correlation: no relationship between *x*, *y*
  - Negative ($-1 \leq r < 0$): when *x* increases, *y* decreases
- $r^2$ shows what **% of variation** in *y* is explained by *x*

# 1. Descriptive statistics

**Correlation** | In a **scatterplot** (Kaplan 2016:268)




(a) Relationship between intensity of the first and second vowels in words like *rapid*, $r = .894$.

(b) Relationship between duration of the stop (*p*) and nasal (*m*) in words like *trumpet*, $r = -.565$.

(c) Relationship between intensity of the first vowel and duration of the middle consonant in words like *rapid*, $r = -.001$.

Figure A.1 Examples of strong (a), moderate (b), and no (c) correlation.

- Can you get a sense of what it means to say that knowing *x* does (or does not) help predict *y*?

# 2. Correlation and causation

- We’ve all heard this: “Correlation is not causation!”
  - What does this actually mean?
  - How should we use it in interpreting experiment results?

# 2. Correlation and causation

- We've all heard this: "Correlation is not causation!"
  - What does this actually mean?

- What this does NOT mean:
  - Correlation is not "real"
  - Correlation tells us nothing
  - Correlation means sample size wasn't large enough

# 2. Correlation and causation

- We've all heard this: "Correlation is not causation!"
  - What does this actually mean?
- What this DOES mean:
  - Finding that $x$ and $y$ are correlated is **not enough** for us to immediately conclude that $x$ causes $y$ (or that $y$ causes $x$)

# 2. Correlation and causation

The following example, with tables and data graphics, comes from Richard Lowry's *Concepts & Applications of Inferential Statistics*, http://vassarstats.net/textbook/ch3pt1.html#top

- SAT scores by state in 1993:

"Among the states near the top of the list in 1993 (verbal and math SAT averages combined) were Iowa, weighing in at 1103; North Dakota, at 1101; South Dakota, at 1060, and Kansas, at 1042. And down near the bottom were the oft-maligned "rust belt" states of the northeast: Connecticut, at 904; Massachusetts, at 903; New Jersey, at 892; and New York, more that 200 points below Iowa, at 887. You can easily imagine the joy in Des Moines and Topeka that day, and the despair in Trenton and Albany. For surely the implication is clear: The state educational systems in Iowa, North Dakota, South Dakota, and Kansas must have been doing something right, while those in Connecticut, Massachusetts, New Jersey, and New York must have been doing something not so right."

# 2. Correlation and causation

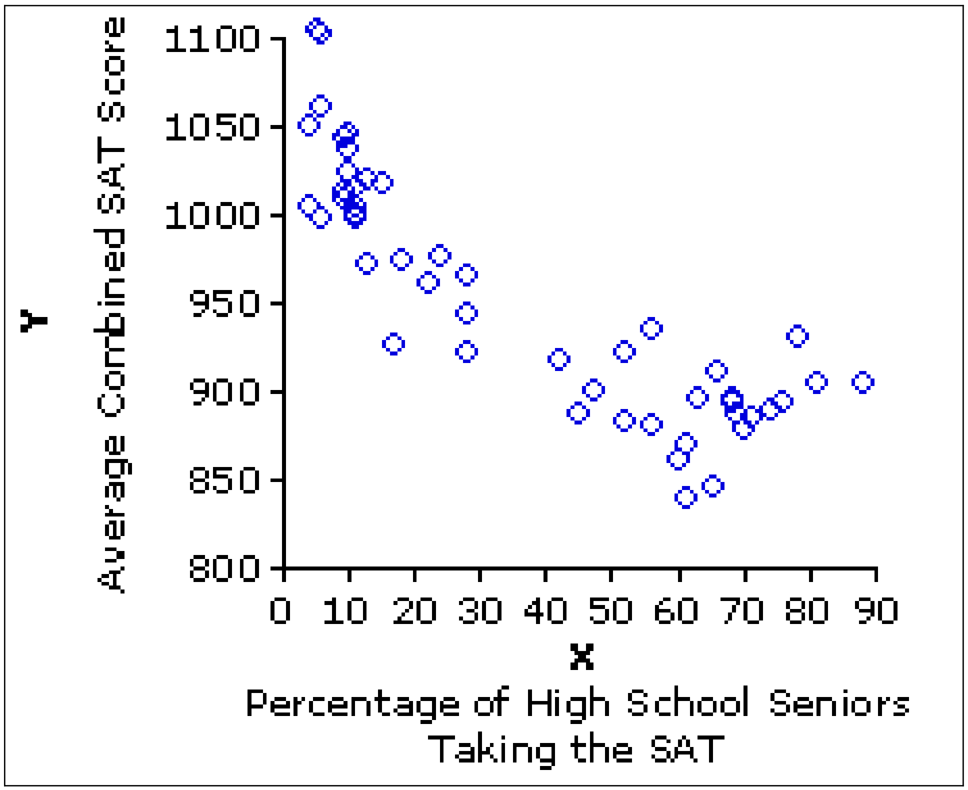
- But then, we learn the following extra information...

| State | Percentage taking SAT | Average SAT score |
|---|---|---|
| Iowa | 5 | 1103 |
| North Dakota | 6 | 1101 |
| South Dakota | 6 | 1060 |
| Kansas | 9 | 1042 |
| Connecticut | 88 | 904 |
| Massachusetts | 81 | 903 |
| New Jersey | 76 | 892 |
| New York | 74 | 887 |

- How can we pursue this?
  - What kind of **data graphic** would help us visually look for evidence of a **correlation**?

# 2. Correlation and causation

- Discuss: What are the **dependent** and **independent** variables? What preliminary conclusions can we draw?



# 2. Correlation and causation

- Suppose we find X and Y are **correlated**. Can we conclude that X **causes** Y?
  - Maybe X causes Y
  - Maybe Y causes X
  - Maybe Z causes both X and Y: in this case, Z is a **confounding factor**

- See a rather timely example: https://www.nytimes.com/2021/08/30/briefing/vaccine-immunity-booster-shots.html

# 2. Correlation and causation

- Take-home points:
  - Correlation does NOT imply causation, but it *can* still be informative
  - Try to minimize confounding factors in experiments

# 3. Inferential statistics

- **Descriptive statistics** — a summary of the information in a particular data set
  - We can describe phenomena we have observed
- But usually, we do experiments to understand general questions, not specific cases
  - Do the phenomena we have observed allow us to make broader predictions about the world?
- **Inferential statistics** — Did the patterns in our data arise by coincidence, or because they represent real facts about the world?

# 3. Inferential statistics

- What is the structure of a typical experiment?

# 3. Inferential statistics

- In its most basic form, an experiment **compares** two **conditions** to see if they are **different**
  - Women and men | *how much they talk*
  - Nouns and verbs | *how early they are learned*
  - French and English vowels | *how nasal they are*
- The **null hypothesis** ('move along, nothing to see here!') is that there is **no difference** between the two conditions

# 3. Inferential statistics

- Suppose we get a numerical difference between two conditions — is it ‘real’ (meaningful)?
  - Ask: What is the **probability (*p*)** that a difference of this size would arise by chance if the null hypothesis is actually correct?

# 3. Inferential statistics

- What is the **probability (*p*)** that a difference of this size would arise by chance *if* the null hypothesis is correct?
  - Low probability → unlikely to have arisen by chance → **statistically significant**

- "Low" probability — how low is low enough?

| | |
|---|---|
| $p<0.001$ | very highly significant |
| $p<0.01$ | highly significant |
| $p<0.05$ | significant |
| $p<0.1$ | 'marginally significant' (sometimes noted) |

# 3. Inferential statistics

- Trade-off: There is no magically "right" *p*-value
  - Threshold ($\alpha$) *too* low? Might reject results too often
  - But $p$<0.05 is sometimes too high (xkcd #882)

- Recent trend in research: Focus on measures such as effect size and confidence intervals rather than *p*-values

# 3. Inferential statistics

- What is the **probability (*p*)** that a difference of this size would <u>arise by chance</u> if null hyp. is correct?

  [See alsoVassarStats Binomial Distributions, Binomial Probabilities]

(Kaplan 2016: 272)

Table A.2 p-*values for various outcomes of a coin-tossing experiment, testing the null hypothesis that heads and tails are equally likely.*

| Tosses | Heads | *p* |
|---|---|---|
| 10 | 6 | .754 |
| 20 | 12 | .503 |
| 50 | 30 | .203 |
| 100 | 60 | .0569 |
| 200 | 120 | .00569 |
| 500 | 300 | .00000894 |

- Which of these coins do you think are unfair?

# 3. Inferential statistics

- Reading about experiment results: What to look for
  - What was the **null hypothesis**?
    (might be assumed rather than stated explicitly!)
  - What **statistical test** was performed?
  - Were any comparisons **statistically significant**?
  - Do the results show
    - a **main effect** (factor matters in the same way across all experiment conditions)?
    - an **interaction** (factor matters differently in different conditions)?

# 3. Inferential statistics

- Some things to watch out for…

  Kaplan (2016: 274)

  - It's tempting to use the *p*-value of a statistical test as a binary decision-making tool: if $p < 0.05$, the result is real; otherwise, it's not.

  - Null results
  - Correlation does not prove causation
  - Statistically significant — but does it matter?

## Null results and experimental power

- If an experiment is too small, it may not produce a low enough *p*-value, even if the effect is real
  - A 'null result' doesn't prove there is no effect
  - But we can trust a null result more confidently if the experiment was large, or many experiments have found a null result
  - Compare the coin-toss example above...
    If we get **60% heads**, is the coin unfair?

# 3. Inferential statistics

## Correlation and causation

- See discussion above!
  (Suppose we find X and Y are **correlated**.  Can we conclude that X **causes** Y?)

# 3. Inferential statistics

**Statistically significant — but does it matter?**

- Given a powerful enough experiment (lots of data), factors may be statistically significant, but the differences involved may not be relevant for real-world decisions
  - Suppose a particular teaching method made a statistically significant difference in test scores
  - But the difference was 1 percentage point
  - Worth the effort to implement?

# 4. Next time

- More about data graphics
- Experimental design