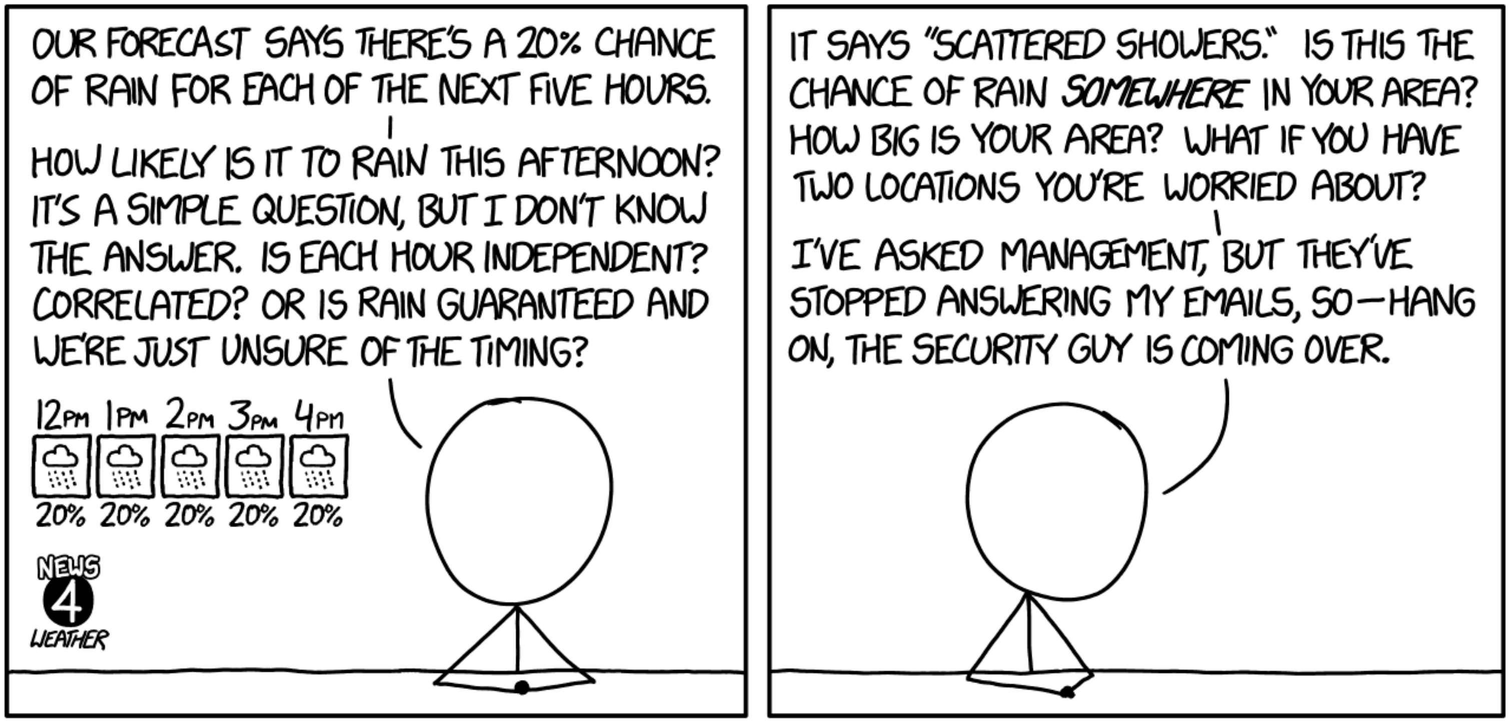
Stat 88: Probability and Mathematical Statistics in Data Science



https://imgs.xkcd.com/comics/meteorologist.png

Lecture 1: 1/20/2021

Course introduction and the basics

Shobhana M. Stoyanov

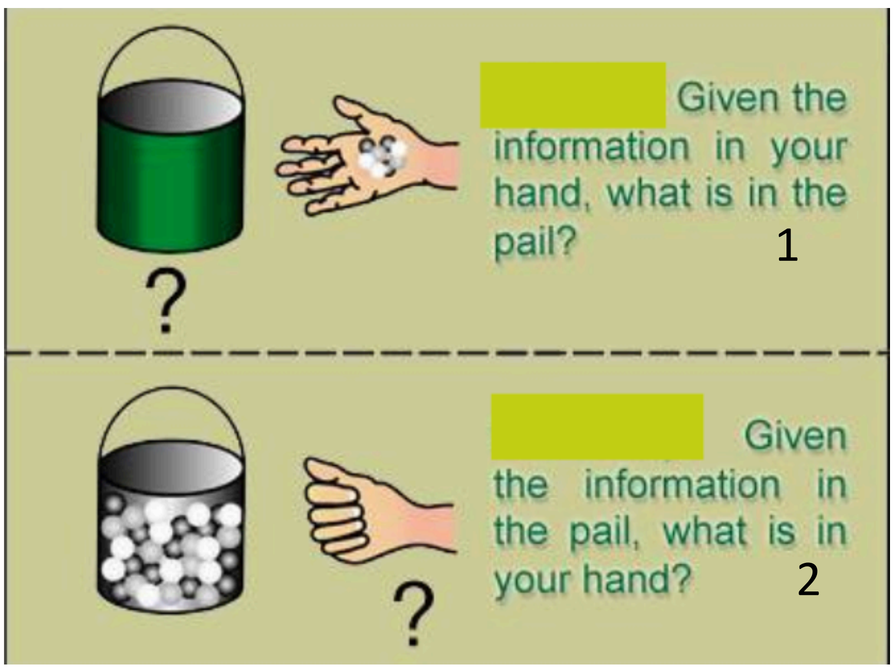
Agenda

- Course resources:
  - Course site: http://stat88.org
  - Announcements and discussions: Piazza
  - Assignments and grades: Gradescope
- Icebreaker padlet
- The Basics:
  - Section 1.1: Probabilities as Proportions
  - Section 1.2: Exact Calculation or Bound

Icebreaker

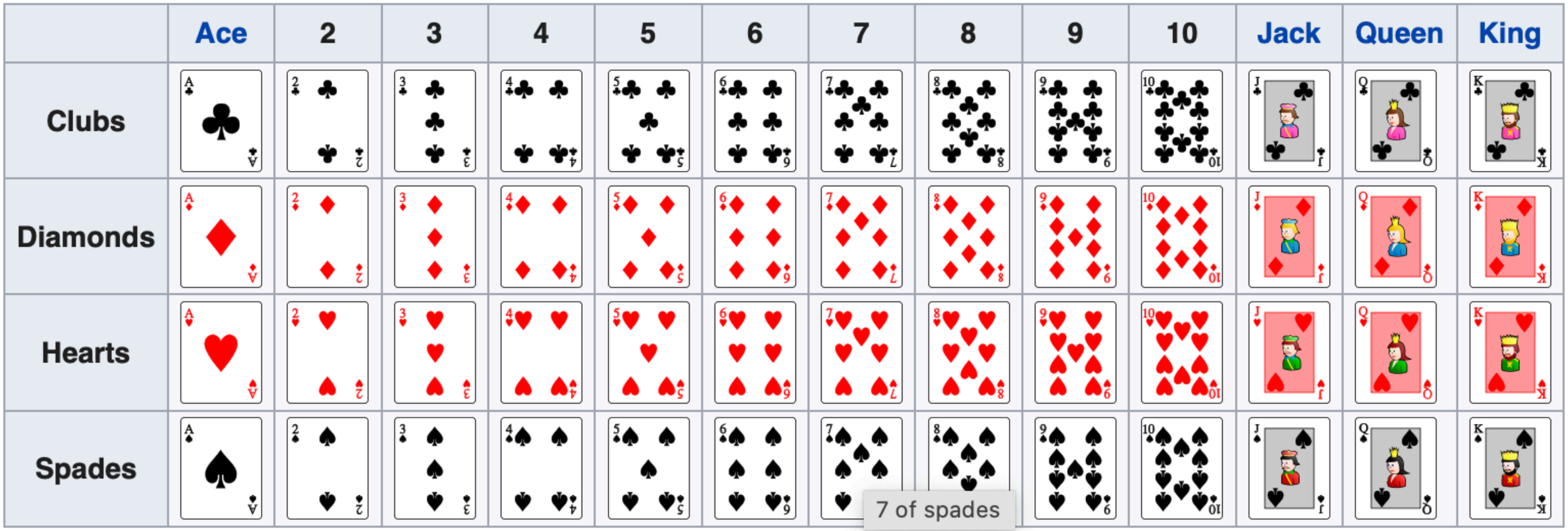
- In the breakout room, introduce yourself to your classmates. How are you feeling today? You can write your thoughts on the padlet.

- Discuss which is probability and which is statistics:



Cards

Example set of 52 playing cards; 13 of each suit clubs, diamonds, hearts, and spades



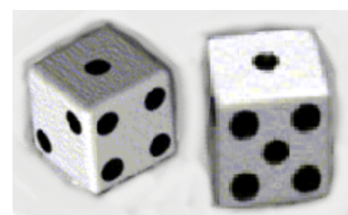
If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

If you deal 2 cards, what is the chance that at least *one* of them is a queen?

Section 1.1: Probabilities as proportions

- We can think about probability as a numerical measure of uncertainty, and we will define some basic principles for computing these numbers.
- These basic computational principles have been known for a long time, and in fact, gamblers thought about these ideas a lot. Then mathematicians investigated the principles.

- Famous problem: will the probability of **at least one six** in **four** throws of a die be equal to prob of **at least a double six** in 24 throws of a pair of dice.

- Note: single = die, plural = dice:  vs 

Origins of probability: de Méré's paradox

Questions that arose from gambling with dice.

Antoine Gombaud,
Chevalier de Méré

Blaise Pascal

Pierre de Fermat

The dice players
Georges de La Tour
(17th century)

Terminology

- Suppose we have an action that results in exactly one of several possible *outcomes* or results, and chance or randomness is involved - that is, each time we perform the action, the outcome will be different, and we don't know exactly which outcome will occur.
- Such an action is called an *experiment* or a *random experiment*.

- A collection of all possible outcomes of an action is called a *sample space* or an *outcome space* . Usually denoted by Ω (sometimes also by S).
- An *event* is a collection of outcomes, so a subset of Ω.

Computing probabilities

- If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

- How did you do this? What were your assumptions?

- Say we roll a die. What is Ω?

- What is the chance that the die shows a multiple of 3? What were your assumptions?

Chance of a particular outcome

- We usually think of the chance of a particular outcome (roll a 6, coin lands heads etc) as the number of ways to get that outcome divided by the total possible number of outcomes.

$$\frac{\#\ of\ particular\ outcomes\ of\ interest}{total\ \#\ of\ outcomes\ possible}$$

- So if A is an event (subset of Ω), then $P(A) = \frac{\#(A)}{\#(\Omega)}, A \subseteq \Omega$

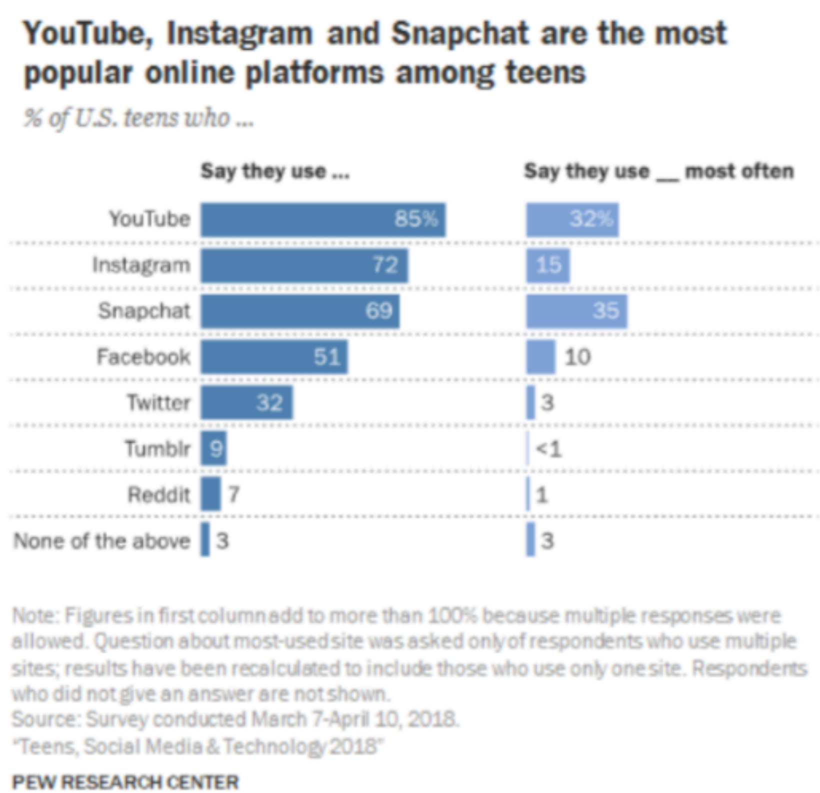
Cards

- If you have a well-shuffled deck of cards, and deal 1 card from the top, what is the chance of it being the queen of hearts? What is the chance that it is a queen (any suit)?

- If you deal 2 cards, what is the chance that at least *one* of them is a queen?

Not equally likely outcomes

- What if our assumptions of equally likely outcomes don't hold (as is often true in life, data are messier than nice examples).
- Here is a graphic from Pew Research displaying the results of a 2018 survey of social media use by US teens.

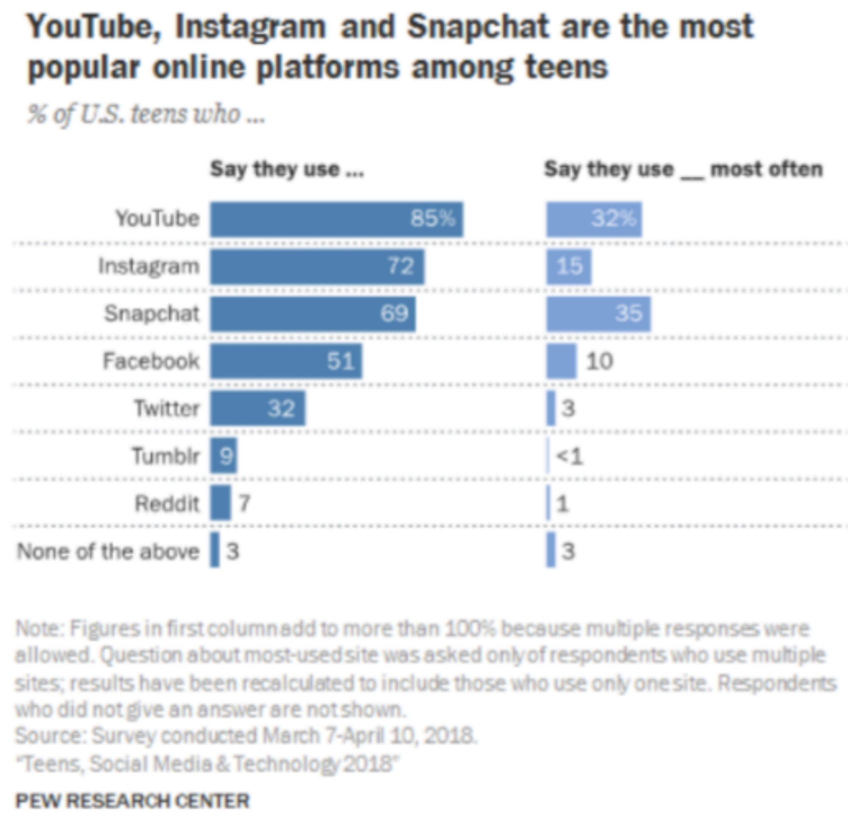


- What is the difference b/w 2 charts?

- Why do the % add up to more than 100 in the first graph?

- Second graph gives us a *distribution* of teens over the different categories

Not equally likely outcomes



1. What is the chance that a randomly picked teen uses FB most often?

2. What is the chance that a randomly picked teen did *not* use FB most often?

3. What is the chance that FB *or* Twitter was their favorite?

4. What is the chance that the teen used FB, just not most often?

5. Given that the teen used FB, what is the chance that they used it most often?

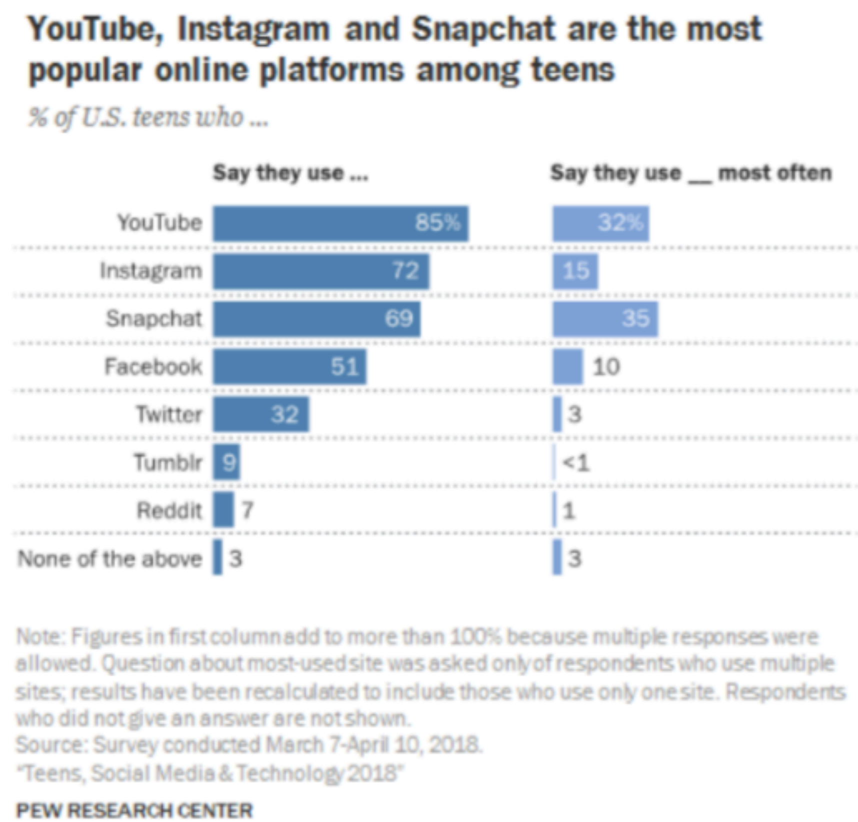
Conditional probability

- In the last question, we used the information that the teen used FB. We were told the teen used FB, and *then* asked to compute the chance that FB was their favorite.
- This is called the *conditional probability that the teen used Facebook most often, given that they used Facebook* and denoted by:

Conditional probability

- This probability we computed is called a ***conditional probability***. It puts a condition on the teen, and *changes* (restricts) the universe (the sample space) of the next outcome, a teen who likes FB best.
- To compute a conditional probability:
  - First restrict the set of all outcomes as well as the event to *only* the outcomes that *satisfy* the given **condition**
  - Then calculate proportions accordingly
- How do the probabilities in #1 and #5 compare?

Section 1.2: Exact Calculations, or Bound?



Recall #3 about FB or Twitter. What was the answer? What can you say about the chance that a randomly selected teen used FB or Twitter (not necessarily most often)?

Rules that we used:

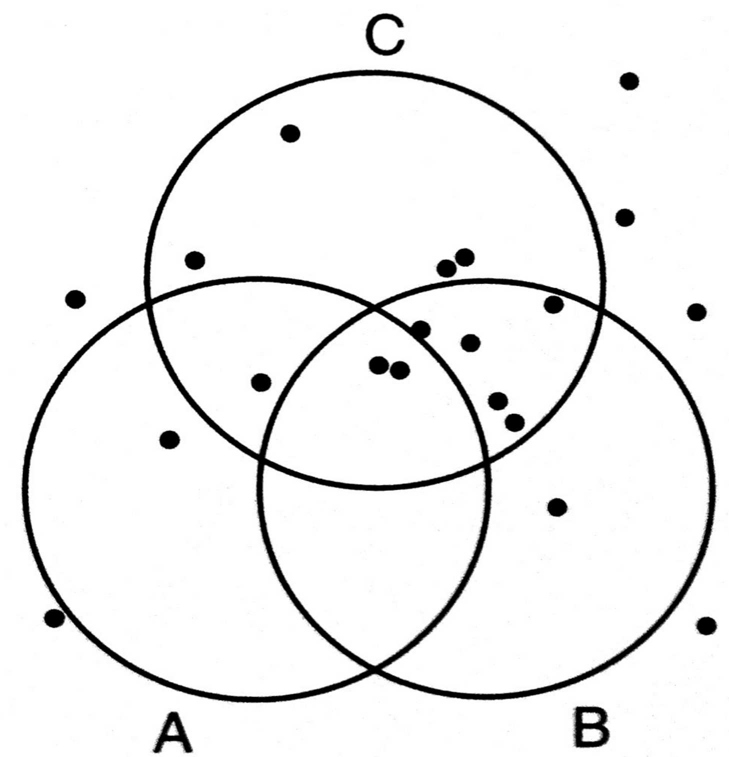
- If all the possible outcomes are *equally likely*, then each outcome has probability ***1/n***, where ***n*** = number of possible outcomes.
- If an event A contains k possible outcomes,

  then *P(A) = k/n*.

- Probabilities are between 0 and 1

- If two events A and B don't overlap, then the probability of A or B = P(A) + P(B) (since we can just add the number of outcomes in one and the other, and divide by the number of outcomes in Ω)

Extra problem 1



Consider the Venn diagram above. (The sample space consists of all the dots.) What is the probability of A? What about A or B? A or B or C?

Extra problem 2

- A ten-sided fair die is rolled twice:

  - If the first roll lands on 1, what is the chance that the second roll lands on a number bigger than 1?

  - Find the probability that the second number is greater than the twice the first number.