

Problem Session 1

Stats 60/160

July 9, 2020

1 Measure of Center, Skew

- **average** (or **mean**): avg of list = $\frac{\text{sum of list}}{n}$.

- **median**: the middle number when you order the list.

- **skew**

skewed to the left: mean < median (e.g., GPA)

skewed to the right: mean > median (e.g., income)

symmetric: mean = median (e.g., normal curve)

- Measures of center can be misleading because they do not take variability into account.

Problem 1.1 The mean undergraduate GPA at Stanford is 3.4. Do you expect (more than / less than / about) half of all undergraduates to have a GPA above 3.4 (or is it impossible to tell)?

2 Measures of Spread

- **standard deviation (SD)**: the deviation of the “typical” observation from the mean. To calculate it:

1. Calculate the deviations from the mean.

2. Square the deviations.

3. Calculate the mean of the squares.

4. Take the square root.

- Steps 2-4 calculate the root-mean-square (r.m.s.) of the deviations. The r.m.s. also measures center.

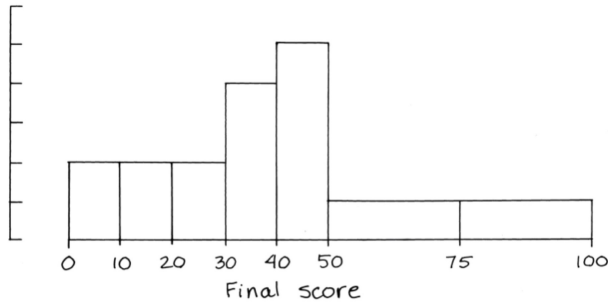
Problem 2.1 Without calculating it, guess the SD of the list $[4, 0, -2, 2, 1]$. Is it 1, 2, or 4?

Problem 2.2 What is the SD of the list $[1, 3, 4, 5, 7]$?

3 Histograms

- The mean and SD still do not paint a complete picture of the data.
- A **histogram** gives a more complete view.
 - Areas correspond to percentages.
 - Heights represent % per unit.
 - The areas must add up to 100%!

Problem 3.1 Shown below is a histogram of final exam scores. Can you estimate the 60th percentile?



4 Normal Curves and the Empirical Rule

- Many histograms based on data follow a normal curve.
- The **empirical rule** is a useful rule of thumb for normal curves.
 - 68% of data fall within 1 SD of the mean.
 - 95% of data fall within 2 SDs of the mean.
 - 99.7% of data fall within 3 SDs of the mean.
- For other SDs (e.g., 1.5), you will need to use a normal table.

Problem 4.1 IQ scores follow the normal curve with mean 100 and SD 15. People with an IQ between 115 and 130 are classified as “bright”. What percentage falls into this category?

Problem 4.2 The speed limit on the freeway is 65mph. Because of error in the radar gun readings, officers will not stop cars unless they are driving over 71mph. The police chief says that this ensures that no more than 2.5% of cars driving at the speed limit will be pulled over for speeding. Assuming radar gun readings follow a normal curve, what does this say about the SD of the readings?

5 Probability Rules

- *Counting Principle*: If all outcomes are equally likely, the probability of any event is
$$\Pr(A) = \frac{\# \text{ outcomes in } A}{\# \text{ possible outcomes}}.$$
- *Addition Rule*: If A and B are **mutually exclusive**, $\Pr(A \text{ OR } B) = \Pr(A) + \Pr(B)$.

- *Multiplication Rule*: If A and B are **independent**, $\Pr(A \text{ AND } B) = \Pr(A) \cdot \Pr(B)$.
- *Conditional Probability*: The probability of B given A is $\Pr(B|A) = \frac{\Pr(A \text{ AND } B)}{\Pr(A)}$. This is the same as $\Pr(A \text{ AND } B) = \Pr(A) \cdot \Pr(B|A)$, which allows us to calculate $\Pr(A \text{ AND } B)$ when events are not independent.
- *Complement Rule*: The probability of the **complement** (the *opposite*) of an event is $\Pr(\text{not } A) = 1 - \Pr(A)$.

Problem 5.1 Tversky and Kahneman (1982) asked subjects the following question.

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

Problem 5.2 Four draws are going to be made from the box

1	2	2	3	3
---	---	---	---	---

. Find the chance that

2

 is drawn at least once if ...

- ... the draws are made with replacement.
- ... the draws are made without replacement.

Problem 5.3 10% of employees at a department store have been skimming money from the cash register. The manager decides to subject all employees to a lie detector test. The lie detector goes off 80% of the time when a person is lying, but it also goes off 25% of the time when a person is telling the truth. The lie detector beeps for a worker who claims he didn't do it. What's the chance he's lying?

Problem 5.4 A poker hand of 5 cards is dealt from a single deck of 52 cards.

- What's the probability the first four cards are the same rank?
- What's the probability you get "four of a kind" (four cards of the same rank)?

Problem 5.5 You are in the middle of an SAT verbal section when the proctor calls out, "One minute remaining!" Oh no! You haven't even read the last passages, and there's only time to guess the answer to the 4 remaining questions. (Remember that each question has five answer choices.)

- What's the chance you get all 4 questions wrong?
- What's the probability you get exactly 1 correct?
- What's the probability you get exactly 2 correct?
- What's the probability you get any correct?

Problem 5.6 You and your friends want to go to a concert. Because it's very popular, each of you only have 1/3 chance of getting the ticket if you put in an order (one person can purchase for the group).

- What is the probability you can successfully buy a ticket for the concert if two of you order? What about three of you order?
- How many of your friends plus you need to put in orders to guarantee at least 85% chance to obtain a ticket for the group?