

Problem Session 1

Stats 60/160

July 14, 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)?

Answer GPA is often left skewed, so the median is right to the mean.

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?

Answer It is probably 2. This is because the center is around 1, most data points are between distance 1 and 3 to the mean.

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

Answer We go through the four steps in the handout.

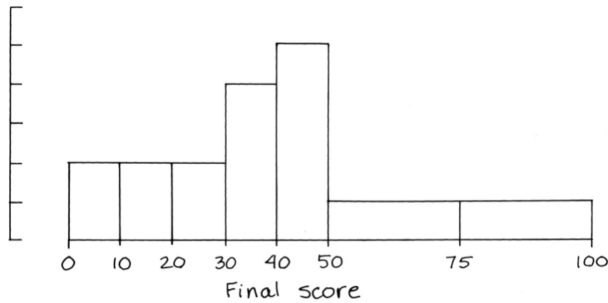
1. The mean is $(1 + 3 + 4 + 5 + 7)/5 = 4$.
2. The deviations from the mean are $[-3, -1, 0, 1, 3]$.
3. The mean of the squares of deviations is $(9 + 1 + 0 + 1 + 9)/5 = 4$.
4. The square root is 2.

The SD is 2. Also, this is the list from problem 2.1 shifted by 3!

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?



Answer If the total area is 100%, then the area on the left of 60th percentile is 60%.

Let's say the first rectangle represents an area a . The the area in each interval is $a, a, a, 2a, 2.5a, 1.25a$, and the total area is $10 \times a$, which means $a = 10\%$.

The area on the left of 40 is 50%, to get another 10% we need to add $10/25 \times 10 = 4$, thus the 60th percentile is 44.

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?

Answer This is equivalent to area between 1 - 2 SD above the mean, so the area is

$$(95\% - 68\%)/2 = 13.5\%.$$

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?

Answer This means 71 is 2SD above the mean radar gun reading at the speed limit, which is 65, thus one SD is

$$\text{SD} = (71 - 65)/2 = 3.$$

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.

Answer Note that if Linda is a bank teller and is active in the feminist movement, she is necessarily a bank teller. However, Linda may be a bank teller without being active in the feminist movement. Therefore, it is more likely that she is a bank teller. This can also be seen using the conditional probability rule: The probability that Linda is a bank teller and is active in the feminist movement equals the probability that Linda is active in the feminist movement given she is a bank teller times the probability that she is a bank teller, which is less than the probability that she is a bank teller.

Problem 5.2 Four draws are going to be made from the box $\boxed{1 \ 2 \ 2 \ 3 \ 3}$. Find the chance that $\boxed{2}$ is drawn at least once if ...

- (a) ... the draws are made with replacement.
- (b) ... the draws are made without replacement.

Answer (a) If the draws are with replacement, then results from each draw are independent (results from the first draw does not affect results from the second draw). Each draw, a $\boxed{2}$ is drawn with probability $2/5$. Using the compliment rule,

$$\Pr(\text{at least one } \boxed{2} \text{ is drawn}) = 1 - \Pr(\text{no } \boxed{2}\text{'s are drawn}) = 1 - \left(\frac{3}{5}\right)^4 = 0.87.$$

(b) There are only three tickets which are not a $\boxed{2}$, so if four tickets are drawn, at least one must be a $\boxed{2}$.

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?

Answer Write L for the event that the lie detector goes off, and S for the event that the employee is skimming money. The information given in the problem can be summarized as $\Pr(S) = 0.1$, $\Pr(L|S) = 0.8$, and $\Pr(L|\text{not } S) = 0.25$. To find the chance that an employee is lying about his innocence, we need to find $\Pr(S|L)$. For this calculation, we will use two more rules of probability

- *Bayes' Rule:* The order of conditioning can be reversed using the relation $\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)}$.
- *The Law of Total Probability:* $\Pr(B) = \Pr(B|A) \times \Pr(A) + \Pr(B|\text{not } A) \times \Pr(\text{not } A)$.

Using Bayes' Rule in combination with the law of total probability,

$$\Pr(S|L) = \frac{\Pr(L|S)\Pr(S)}{\Pr(L)} = \frac{\Pr(L|S)\Pr(S)}{\Pr(L|S)\Pr(S) + \Pr(L|\text{not } S)\Pr(\text{not } S)}.$$

Plugging in the information given in the problem,

$$\Pr(S|L) = \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.25 \times 0.9} = 0.26.$$

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

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

Answer (a) The chance that the first four cards drawn are of any particular denomination (e.g. the chance that the first four cards drawn are all 2's) is

$$\frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} \times \frac{48}{48} = 3 \times 10^{-6}.$$

Since there are 13 possible denominations, the chance that the first four cards will be of the same denomination is

$$13 \times \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} \times \frac{48}{48} = 5 \times 10^{-5}.$$

Alternatively, you can enumerate the number of ways the cards could be dealt. For any particular denomination of the first four cards, there are $4 \times 3 \times 2 \times 1$ ways to deal these cards (the first could be any one of the four suits, the second one has three options and so on). Since there are 13 denominations, altogether there are $13 \times 4 \times 3 \times 2 \times 1$ ways. There are 48 options for the last card. The total number of ways to deal 5 cards is $52 \times 51 \times 50 \times 49 \times 48$. Each way is equally likely, so we can use the counting principle.

(b) If a hand is a “four of a kind,” we must have 4 cards of the same denomination, and one card of any other denomination. The card of a different denomination can be placed in any of five positions (for instance we may have four cards of the same denomination followed by a different card, or two cards of one denomination followed by a card of a different denomination followed by two cards of the original denomination). The chance that four cards of the same denomination are drawn, with the card of a different denomination in a fixed position is given by the probability calculated in the previous question. So, the chance of getting “four of a kind” is

$$5 \times 13 \times \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} \times \frac{48}{48} = 0.0002.$$

You can also enumerate the number of ways a “four of a kind” occurs, there are 5 ways to pick location for the other denomination, 48 ways to pick this card, and 13×24 ways to pick the other four cards, and the denominator is again $52 \times \dots \times 48$.

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.)

- (a) What's the chance you get all 4 questions wrong?
- (b) What's the probability you get exactly 1 correct?
- (c) What's the probability you get exactly 2 correct?
- (d) What's the probability you get any correct?

Answer (a) Because results of the guesses are independent, we use the multiplication rule to get

$$\begin{aligned} \Pr(\text{all 4 wrong}) &= \Pr(\text{question 1 wrong}) \times \dots \times \Pr(\text{question 4 wrong}) \\ &= \left(\frac{4}{5}\right)^4 = 0.41. \end{aligned}$$

(b) The chance you get the first question correct and the others wrong is $1/5 \times (4/5)^3 = 0.0432$. There are four locations this correct question could be (first, second, third or fourth), each has equal probability, so the chance to get one question correct is $4 \times 0.0432 = 0.1728$.

(c) The chance you get the questions 1 and 2 correct and questions 3 and 4 wrong is $(1/5)^2(4/5)^2 = 0.0256$. There are 6 ways to choose 2 correct questions out of 4 questions (1&2, 1&3, 1&4, 2&3, 2&4, 3&4). Thus the chance to get 2 questions correct is $6 \times 0.0256 = 0.1536$.

(d) Use the complement rule to get

$$\Pr(\text{any correct}) = 1 - \Pr(\text{all 4 wrong}) = 1 - 0.41 = 0.59.$$

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).

(a) What is the probability you can successfully buy a ticket for the concert if two of you order? What about three of you order?

(c) 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?

Answer This question is very similar to problem 5.5.

(a) Whether each of you get a ticket or not is independent of another, so the probability the two of you fail to get a ticket is

$$\left(\frac{2}{3}\right)^2 = 0.44.$$

By the complement rule, the success probability is $1 - 0.44 = 0.56$.

(b) If three of you put in an order, the chance all of you fail to get a ticket is $(2/3)^3 = 0.296$. So the success probability is $1 - 0.296 = 0.704$.

(c) Similarly, we can calculate the success probability if 4 or 5 of you put in orders, and the success probabilities are 80% and 87% respectively. Thus if 5 of you put in an order, the success probability is at least 85%.