Lab 3a: Measures of Spread and Normal Curve

Stat 131A, Fall 2018

Learning Objectives:

- Find the mean and median from different representations of data.
- Develop number sense with mean and median by creating different data sets with a given mean or median.
- Normal curve.
- Conversion to z-scores (i.e. standard units).

General Instructions

- Write your solutions in an `Rmd` (R markdown) file.
- Name this file as `lab03a-first-last.Rmd`, where `first` and `last` are your first and last names (e.g. `lab03a-gaston-sanchez.Rmd`).
- Knit your `Rmd` file as an html document (default option).
- Submit your `Rmd` and `html` files to bCourses, in the corresponding lab assignment.

Problem 1

The Governor of California proposes to give all state employees a raise:

a. One option is to give a flat raise of $250 a month. What would this do to the average monthly salary of state employees? to the Stadandrd Deviation?

b. Another option is a raise of a 5% increase in the monthly salaries. What would this do to the average and SD?

Problem 2

The length of time a person spends waiting in a physician's office for an appointment can be a frustrating experience. The results below are the responses of internists from two multispecialty group practices to the following question: "When patients arrive for midmorning or midafternoon appointments, how long do they generally have to wait before they are seen by you?" The table below contains a sample of the responses in minutes.

Group A	Group B
21	13
8	19
17	35
13	4
12	0

Try using R to do calculations.

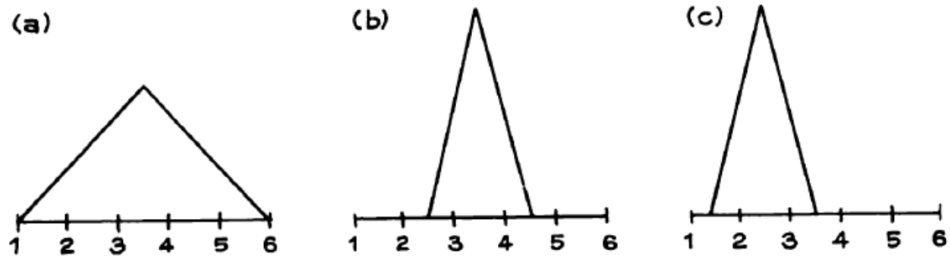
a. Compute the mean, median, and standard deviation of the internists' responses of group A.

b. Compute the mean, median, and standard deviation of the internists' responses of group B.

c. If you had to make an appointment and wanted some assurance that you would not spend more than 30 minutes waiting in the office, which would you choose? Why?

Problem 3

Below are sketches of histograms for three sets of numbers. Match the sketch with the description. Some descriptions will be left over. Give your reasoning in each case.



i. mean ≈ 3.5, SD ≈ 1

ii. mean ≈ 3.5, SD ≈ 0.5

iii. mean ≈ 3.5, SD ≈ 2

iv. mean ≈ 2.5, SD ≈ 1

v. mean ≈ 2.5, SD ≈ 0.5

vi. mean ≈ 4.5, SD ≈ 0.5

Problem 4

A statistics professor gives a test with 15 questions. This professor is a hard grader. The professor assigns points in the following manner: a correct answer = +1 point, an incorrect answer = -1 point, no answer = 0 points.

In theory, scores on the test could range from +15 to -15. The GSI calculates the standard deviation for scores on the test and reports SD = -2.3. What can we conclude given a standard deviation of -2.3?

a. There is an outlier that has a negative value close to -15.

b. Most students received negative scores.

c. Most students scored below the mean.

d. The standard deviation was calculated incorrectly.

Problem 5

A college statistics class conducted a survey of how students spend their money. They asked 25 students to estimate how much money they typically spend each week on fast food. They determined that the mean amount spent on fast food is $31.52 with a standard deviation of $21.60. Later they realized that a value entered as $3 should have been $30. They recalculate the mean and standard deviation. The mean is now $32.60.

Which of the following is true about the standard deviation?

a. The standard deviation will increase, because we have increased the value of a data point.

b. The standard deviation will stay the same, because the standard deviation is not affected by a change in a single measurement.

c. The standard deviation will decrease, because this change moved a data point closer to the mean.

Problem 6

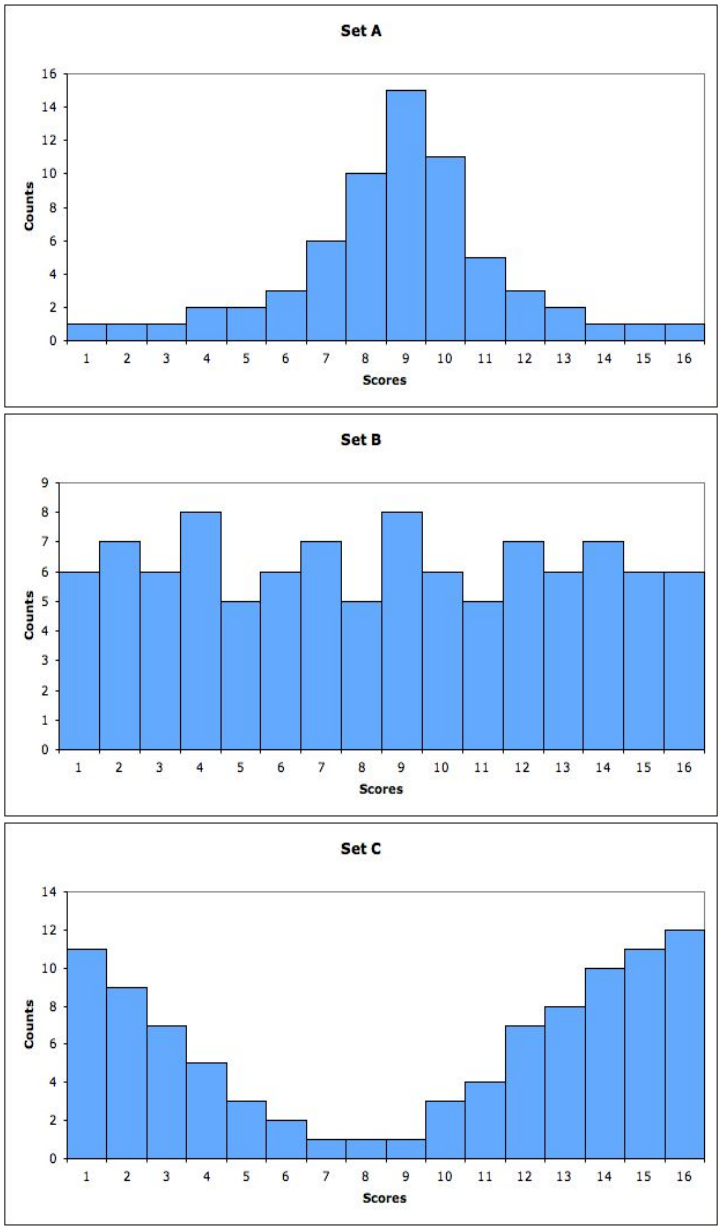
Consider the following two quantitative data sets:

- *Set A:* The times (in minutes) of all competitors in the 1,500-meter running track-and-field event at the most recent Olympic Games.
- *Set B:* The times (in minutes) of all competitors in the 1,500-meter running track-and-field event at all high school meets in the United States last year.

Which variable would you expect to have the larger SD?

Problem 7

Three math classes took a quiz, and their scores are displayed in the histograms below.



a. Which of the three sets has the largest standard deviation? Why?

i) Set A, because it has a larger central peak.
ii) Set B, because it is uniformly distributed.
iii) Set C, because it has a larger proportion of its values far away from the center.
iv) The three sets have the same standard deviation because they have equal ranges.

b. Which of the three sets has the smallest standard deviation? Why?

i) Set A, because it has most of its values near the center.
ii) Set B, because it is uniformly distributed.
iii) Set C, because of the lack of a large central peak.

Problem 8

One term, about 700 Statistics students at UC Berkeley, were asked how many college mathematics courses they had taken, other than Statistics. The mean number of courses was about 1.1; the SD was about 1.5. Would the histogram for the data look like (i), (ii), or (iii)? Why?

Problem 9

On a certain exam, the mean of the scores was 50 and the SD was 10.

a. Convert each of the following scores to standard units: 60, 45, and 75.

b. Find the scores which in standard units are: 0, +1.5, -2.8.

Problem 10

A group of people have an average temperature of 98.6 degrees Fahrenheit, with an SD of 0.3 degrees.

a) Translate these results into degrees Celsius.

b) Someone's temperature is 1.5 SDs above

Problem 11

See the documentation for the family of Normal distribution functions in R: `?Normal`. Use the function that allows you to find the area under the normal curve, and compute the following areas:

a. to the left of 0.80
b. to the right of 1.25
c. to the left of -0.40
d. between 0.40 and 1.30
e. between -0.30 and 0.90
f. outside -1.5 to 1.5

Problem 12

For Berkeley freshmen, the average GPA (grade point average) is around 3.0; the SD is about 0.5. The histogram follows the normal curve. Estimate the 30th percentile of the GPA distribution.

Problem 13

Among applicants to one law school, the average LSAT score was about 169, the SD was about 9, and the highest score was 178. Did the LSAT scores follow the normal curve?

Problem 14

Among freshmen at a certain university, scores on the Math SAT followed the normal curve, with an average of 550 and an SD of 100. Fill in the blanks; explain briefly.

a. A student who scored 400 on the Math SAT was at the _____th percentile of the score distribution.

b. To be at the 75th percentile of the distribution, a student needed a score of about _____ points on the Math SAT.

Problem 15

The Public Health Service found that for boys age 11 in a national study, the mean height was 146 cm and the SD was 8 cm. Fill in the blanks.

a. One boy was 170 cm tall. He was above the mean, by _____ SDs.

b. Another boy was 148 cm tall. He was above the mean, by _____ SDs.

c. A third boy was 1.5 SDs below the mean. He was _____ cm tall.

d. If a boy was within 2.25 SDs of average height, the shortest he could have been is _____ cm and the tallest is _____ cm.

e. Here are the heights of four boys: 150 cm, 130 cm, 165 cm, 140 cm. Match the heights with the descriptions. A description may be used twice: i) unusually short, ii) about average, iii) unusually tall

f. About what percentage of boys age 11 in the study had heights between 138 cm and 154 cm?

g. About what percentage of boys age 11 in the study had heights between 130 cm and 162 cm?

Problem 16

The heights of the NBA players averaged 78.96 inches; the SD was 3.48 inches. Assume that the height follows a symmetric bell-shaped distribution so that it is safe to use the normal approximation for data. Use the function `pnorm()` to find the percentage of players with heights:

a. below 75 inches.

b. above 80 inches.

c. between 77.5 and 82 inches.

d. below 73.5 inches or above 83.5 inches.

Problem 17

For women age 25-34 with full time jobs, the average income in 2004 was $32,000. The SD was $26,000, and 1/4 of 1% had incomes above $150,000. Was the percentage with incomes in the range from $32,000 to $150,000 about 40%, 50%, or 60%? Choose one option and explain briefly.