

Q. (4 points) Let $X = \{-6, -3, -2, 0, 1, 4\}$.

a) Calculate the mean, \bar{x} .

$$\begin{aligned}\bar{x} &= \frac{1}{6}(-6 + (-3) + (-2) + 0 + 1 + 4) \\ &= \frac{1}{6}(-6) \\ &= -1\end{aligned}$$

b) Find the standard deviation, s_X .

1. Subtract -1 from each value to find the deviations from the mean: $\{-5, -2, -1, 1, 2, 5\}$.

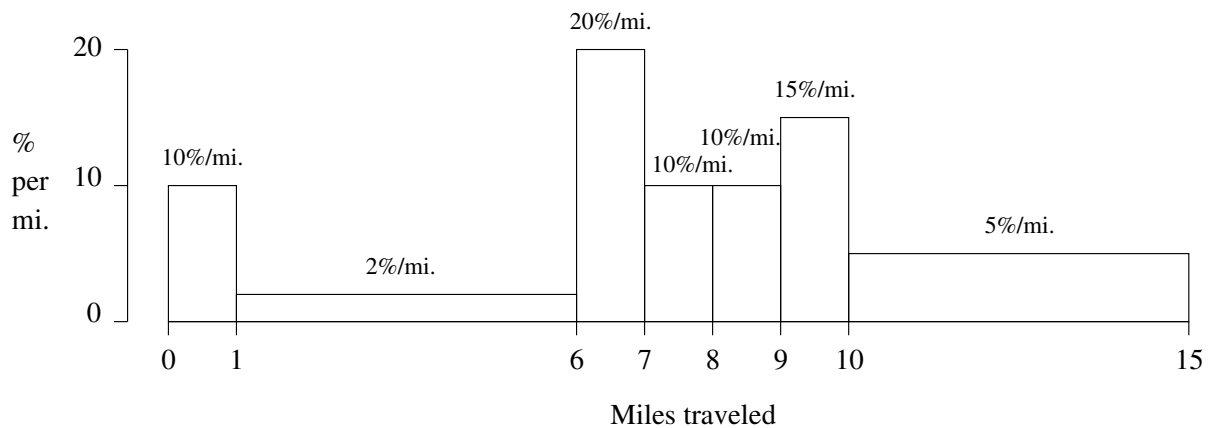
2. Square each value: $\{25, 4, 1, 1, 4, 25\}$.

3. Find the mean of the new list:

$$\begin{aligned}\frac{1}{6}(25 + 4 + 1 + 1 + 4 + 25) &= \frac{1}{6}(60) \\ &= 10.\end{aligned}$$

4. Take the square root: $s_X = \sqrt{10}$.

Q. (3 points) (Hypothetical). The ACME corporation asked its employees to measure the distance of their commutes. The results are summarized in the relative frequency histogram below.



a) What percent of employees travel more than 10 miles?

$$5\text{mi.} \times 5 \frac{\%}{\text{mi.}} = 25\%$$

b) About what percent travel more than 7.5 miles?

Add up the areas to the right of 7.5.

$$0.5 \times 10 + 1 \times 10 + 1 \times 15 + 5 \times 5 = 55\%$$

c) What is the median distance traveled?

Because 55% of employees travel more than 7.5 miles the median must be greater than 7.5. If we look at (b) we see that the area to the right of 8 is

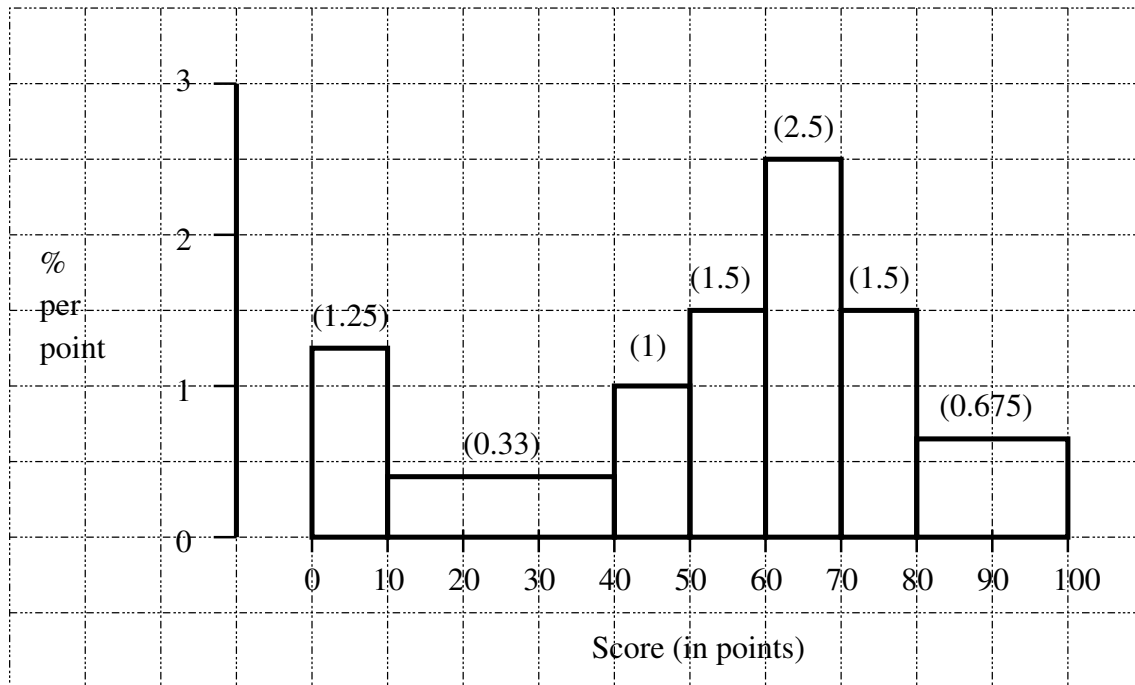
$$1 \times 10 + 1 \times 15 + 5 \times 5 = 50\%.$$

Therefore the median is 8 miles.

Q. (5 points) The results of a test in a college math class are summarized by the table below. Class intervals include left endpoints but not right endpoints.

Score range	Percent	Width	Height ($= \frac{\text{percent}}{\text{width}}$)
0 - 10	12.5	10	$\frac{12.5}{10} = 1.25$
10 - 40	10	30	$\frac{10}{30} = 0.33$
40 - 50	10	10	$\frac{10}{10} = 1$
50 - 60	15	10	$\frac{15}{10} = 1.5$
60 - 70	25	10	$\frac{25}{10} = 2.5$
70 - 80	15	10	$\frac{15}{10} = 1.5$
80 - 100	12.5	20	$\frac{12.5}{20} = 0.675$

Draw a relative frequency histogram for the data on the grid below (you may use as much or as little of the grid as you wish). Indicate the height of each part of the histogram and label the axes.



Q. (2 points) Ads for ADT Security Systems claim

When you go on vacation, burglars go to work According to FBI statistics, over 25% of home burglaries occur between Memorial Day and Labor Day.

Memorial Day is May 25 and Labor day is the first Monday of September. Do these statistics support the argument that burglars go to work when other people go on vacation? Answer yes or no, and explain your answer.

Memorial Day to Labor Day is the traditional summer vacation time, so that part of ADT's claim makes sense. The other part is the implicit claim that burglaries are more common in the months between Memorial Day and Labor Day than in other months. The statistic that 25% of burglaries occur between Memorial Day and Labor Day **does not support this claim**. This is because Memorial Day to Labor Day is about 25% of the year. We would expect 25% of burglaries to occur between Memorial Day and Labor Day if the rate of burglary were the same through the whole year. Thus the statistic is evidence that burglars don't care whether or not people are on vacation.

Q. (2 points) Both of the following lists have an average of 50.

(i) 50, 40, 60, 34, 66, 30, 70, 25, 75

(ii) 50, 40, 60, 34, 66, 30, 70, 25, 75, 99, 1

Which has a larger standard deviation, and why? No computations are necessary.

List (ii) has a larger standard deviation. This is not just because there are more numbers in list (ii) than in list (i). The list 50, 40, 60, 34, 66, 30, 70, 25, 75, 51, 49 has a smaller SD than list (i), for example. List (ii) has a larger standard deviation because 99 and 1 are farther from the mean than numbers in list (i) tend to be.

Q. (5 points) From the mid-1960s to the early 1990s, there was a slow but steady decrease in SAT scores. For example, in 1967 the average Verbal SAT score was about 540; in 1994 the average Verbal SAT score had dropped to about 500. In both years the standard deviation was about 100 and the scores were approximately normally distributed.

- a) Estimate the percentage of students scoring over 700 on the Verbal SAT in 1967. Standardize the score (700) and use the z -table.

$$z = \frac{700 - 540}{100} = \frac{160}{100} = 1.6$$

From the table the percentage of students scoring **below** 700 is 94.52%. The percentage scoring above 700 is then $100 - 94.52 = 5.48\%$.

- b) Estimate the percentage of students scoring over 700 on the Verbal SAT in 1994. Standardize the score (700) and use the z -table.

$$z = \frac{700 - 500}{100} = \frac{200}{100} = 2$$

From the table the percentage of students scoring **below** 700 is 97.72%. The percentage scoring above 700 is then $100 - 97.72 = 2.28\%$.

- c) Estimate the 95th percentile for the Verbal SAT in 1967.

Look on the z -table for the first number greater than or equal to 95. This happens for $z = 1.65$. Therefore the 95th percentile is 1.65 SDs above the mean. We unstandardize to find the score.

$$x = 1.65 \times 100 + 540 = 165 + 540 = 705$$

The 95th percentile is 705.

Q. (4 points) (Hypothetical). Scientists are conducting experiments to determine if lack of sleep causes colds. A group of volunteers was exposed to the cold virus. The volunteers then recorded the amount of time they slept and whether or not they got a cold. After 10 days the rate of illness among the volunteers who reported less than 60 hours of sleep in total was found to be higher than the rate of illness among volunteers who reported more than 60 hours of sleep.

- a) Is this a controlled experiment or an observational study?

This is an observational study. The volunteers were divided into groups based on how much they slept, something the researchers had no control over.

- b) Identify the control and treatment groups.

The control group consists of those people who got more than 60 hours of sleep in total. The treatment group consists of those people who got less than 60 hours of sleep in total.

Note: it would also have been reasonable in this study to group people according to whether or not they got sick. In this case, however, the results would have been stated differently: those who got sick were more likely to have slept less than 60 hours in total than those who didn't get sick.

c) What are two possible confounding factors? Explain how these confounding factors might affect the conclusions you can draw from the study.

There are many possible answers.

Age is one confounding factor. Older people tend to sleep less than younger people and could also be more likely to get sick.

Another confounding factor is stress. Stress is thought to impair the functioning of the immune system and so might lead a subject to get sick. Stress can also affect sleep.

Q. (3 points) In Physics Lab you are conducting an experiment to measure acceleration due to gravity. In the experiment you measure the amount of time it takes a ball to roll down an inclined plane. Each student repeats the experiment 20 times and then takes the average of his/her measurements. (This average, along with the length and angle of the inclined plane, will allow you to calculate the acceleration of gravity).

a) Why is the lab instructor making you repeat the experiment 20 times instead of just doing the experiment once?

Each measurement will have some chance error and thus may be relatively far from the actual value. Fortunately chance error on average is 0. Therefore repeating the measurement many times and taking the average should give us an accurate estimate of the actual value.

b) The mean and standard deviation of Katie's measurements are 4.16 seconds and 0.18 seconds, respectively. One of Bill's measurements was 3.74 seconds. Is this difference between Bill's measurement and Katie's average likely the result of chance error or bias? Explain your answer.

If we standardize Bill's measurement we see that it is 2.33 SDs below Katie's average. The probability of a measurement this low being the result of chance error is 0.99%. This is quite small and so it would be reasonable to suspect that there is some bias at work. Possibly Bill is doing something different from Katie in performing the experiment.

Q. (2 points) Which would you expect to be larger: the average (mean) age of people in this class or the median age? Explain your answer.

Most people in the class are probably between 18 and 25. Therefore the median age for the class is probably in this range.

There are a couple of students who are considerably older (in their 30s) but none who are considerably younger (no one below 17). Therefore the mean is likely to be higher than the median.