

1. A recent poll, conducted from April 11-15, 2012 collected data from 1,000 adults randomly selected from throughout the U.S. The margin of error in this poll was ± 2.7 percent. Poll participants were asked:

“As you may know, there is a proposal in Congress to allow foreigners who have jobs but are staying illegally in the United States to apply for legal, temporary-worker status. Are you in favor of this proposal or are you against this proposal?”

Of those polled, 48.3% were in favor of the proposal and 51.7% were against the proposal.

- (a) (3 points) What is the population in this survey?

Based on the description above, the population is adults currently living in the U.S.

- (b) (3 points) What is the sample in this survey?

The sample is the 1,000 adults who were randomly selected to take the poll.

- (c) (3 points) Based on this survey, what conclusions, if any, can you reach about the opinion the overall population holds about this proposal.

First, it is important to note that this question is asking about the *population*, **not** the *sample*, so simply observing the percentage of responses for the sample is not enough here. Most of you observed that the sample seems to be fairly evenly divided on this issue. This big clue that points us toward information about the population is the margin of error in this poll. Given a 2.7% margin of error, we can see that somewhere between 45.6% and 51.0% of the population are in favor of the proposal, and somewhere between 49.0% and 54.4% of the population are against the proposal. Based on this, we cannot be sure whether or not more people are in favor of this proposal than are against it.

2. (a) (3 points) List the three forms of bias that were discussed in class.

Selection Bias, Non-response Bias, and Leading Question Bias.

- (b) (3 points) Pick one of these forms of bias and define it precisely.

Selection Bias: the method used to select the sample for a study systematically excludes part of the population of interest.

Non-response Bias: a significant portion of the sample does not give a response.

Leading Question Bias: the phrasing of the question (either accidentally or by design) encourages one or more responses and/or to discourage ones or more responses.

3. A student has scored 70, 76, and 84 on the first three exams in a Physics class. Suppose that a fourth exam will be given next week and that the student's final grade will be computed based solely on their scores on these four exams (all four are out of a maximum of 100 points)

- (a) (4 points) What score would the student need to get on this exam in order for their final average to be 80?

Let x be the missing score. To get an 80 average, we need to have $\frac{70 + 76 + 84 + x}{4} = 80$, or $70 + 76 + 84 + x = 320$.
Therefore, we must have $230 + x = 320$, or $x = 90$.

Hence the student would need to get 90 on the exam in order to end up with an average of 80 in the course.

- (b) (4 points) What score would the student need to get on this exam in order for their final average to be 90?

Let x be the missing score. To get an 90 average, we need to have $\frac{70 + 76 + 84 + x}{4} = 90$, or $70 + 76 + 84 + x = 360$.
Therefore, we must have $230 + x = 360$, or $x = 130$.

From this, since we are told that the maximum possible score is 100, it is not possible for the student to end up with an average of 90 in the course.

- (c) (4 points) Supposing that they scored 90 on the fourth exam, would the student rather have their final grade computed using the median or the mean? Be sure to justify your answer.

The mean of these 4 scores is $\frac{70 + 76 + 84 + 90}{4} = 80$, while the median is $\frac{76 + 84}{2} = 80$, so the student has no preference since the resulting grade would be the same either way.

4. The following data set gives the daily high temperature in Fargo, ND from April 1st to April 15th, 2012 (source: www.wunderground.com) {73, 59, 63, 64, 66, 67, 58, 60, 45, 41, 50, 68, 55, 71, 59}

(a) (6 points) Make a stem and leaf display for this data set.

We begin by putting this data set in increasing order. This will help us both here and in part (c). This gives us:

{41, 45, 50, 55, 58, 59, 59, 60, 63, 64, 66, 67, 68, 71, 73}

```

1 |
2 |
3 |
4 | 1 5
5 | 0 5 8 9 9
6 | 0 3 4 6 7 8
7 | 1 3

```

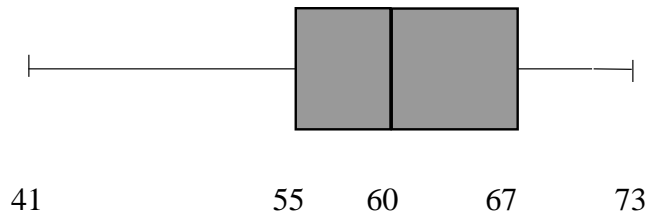
(b) (6 points) Find the *mean* and *midrange* of this data set.

$$\text{mean: } \bar{x} = \frac{899}{15} \approx 59.9^\circ F$$

$$\text{midrange: } \frac{41 + 73}{2} = 57.0^\circ F$$

(c) (8 points) Find the 5 number summary for this data set and draw a “Box-and-Whisker” plot.

min: 41, Q_1 : 55, Q_2 : 60, Q_3 : 67, Max: 73



(d) (5 points) Which measure of center do you think best describes the “middle” of this data set? Explain your reasoning.

Notice that this data set has a negative outlier (41). Because of this, we would expect that the median would be the best measure of center. In practice, the mean is $59.9^\circ F$, the median is $60.0^\circ F$, the midrange is $57^\circ F$, and the mode is 59. The outlier seems to be effecting the midrange, so it is not the best measure. Since the data is numerical we would not typically use the mode. Both the mean and the median seem to give a good representation of the “middle” in this particular data set.

5. Given the following frequency table:

x	Freq.	$x \cdot f$	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 \cdot f$
6	5	30	-4	16	80
8	4	32	-2	4	16
11	4	44	1	1	4
12	5	60	2	4	20
17	2	34	7	49	98
Total:	20	200			218

(a) (8 points) Compute the mean and median of the data in this table.

$$\text{mean: } \bar{x} = \frac{200}{20} = 10$$

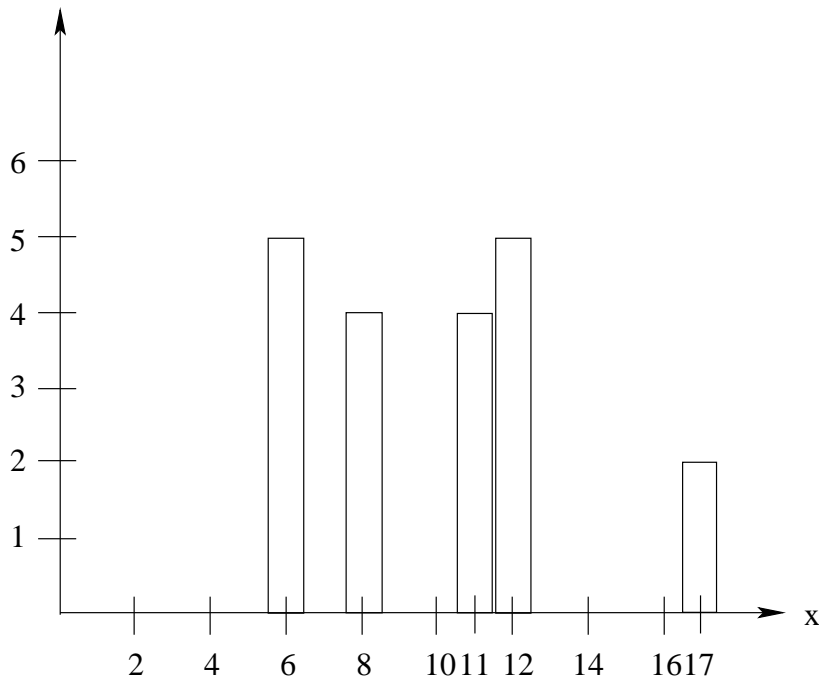
$$\text{median: (the average of the 10th and 11th data values)} = \frac{11 + 11}{2} = 11$$

(b) (10 points) Compute the standard deviation of the data by completing the table above.

$$\text{standard deviation: } s = \sqrt{\frac{218}{20 - 1}} = \sqrt{\frac{218}{19}} \approx 3.39$$

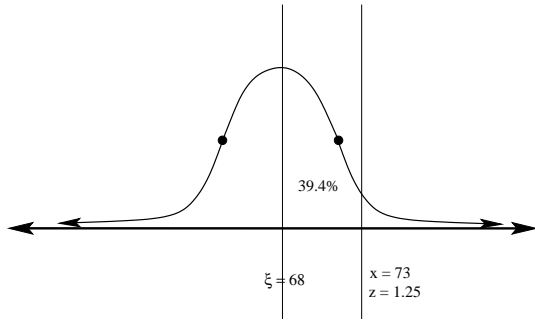
(c) (6 points) In the space provided, make a frequency bar graph for the data in the table above. Be sure to label your axes.

Frequency



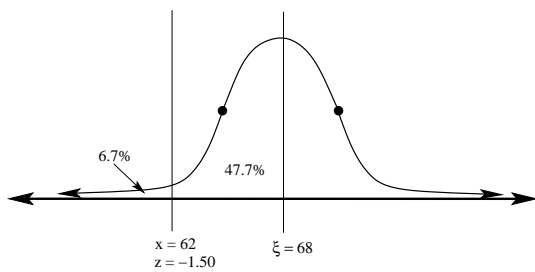
6. (5 points each) Suppose that the height of adults in the Fargo metropolitan area is normally distributed with a mean of 68 inches and a standard deviation of 4 inches. Also suppose that the total adult population of the Fargo metropolitan area is 150,000 people.

(a) What *percentage* of the population is *between* 68 and 73 inches tall?



$z = \frac{73 - 68}{4} = 1.25$. Therefore, using the z -table and the diagram above, $A = 0.394$, or 39.4%.

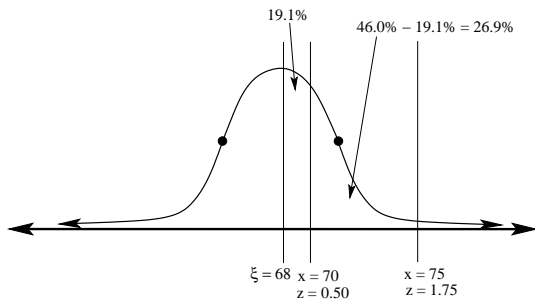
(b) What *percentage* of the population is *less than* 62 inches tall?



$z = \frac{62 - 68}{4} \approx -1.50$. From the z -table, we see $A = 0.433$.

Since we want the percentage of the population whose height is **less than** 62 inches, we subtract $50 - 43.3 = 6.7\%$.

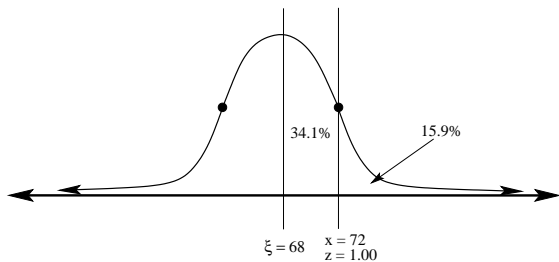
(c) What *percentage* of the population are *between* 70 and 75 inches tall?



$z_1 = \frac{70 - 68}{4} = 0.50$, $z_2 = \frac{75 - 68}{4} = 1.75$, so $A_1 = 0.191$ and $A_2 = 0.460$

Therefore, using the z -table and the diagram above, $A = 46.0\% - 19.1\%$, or 26.9%.

(d) **How many adults** in Fargo are *over* 72 inches tall?

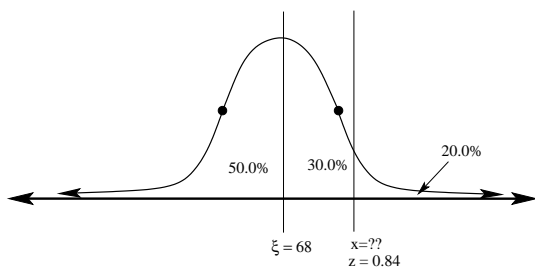


$z = \frac{72 - 68}{4} \approx 1.00$. Once again, using the z -table, we see $A = 0.341$.

To find the percentage of the population with height *over* 72 inches, we compute $50 - 34.1 = 15.9\%$.

However, we want the *number of people* that satisfy this description, so we compute $(150,000)(0.159) = 23,850$. We conclude that around 23,850 people in the Fargo metro area are over 72 inches tall.

(e) How tall would a person need to be in order to be *taller than* 80% of the people in the Fargo metro area?



We want to find the x -score so that 80% of the population is below this score. We see from the z -table that when $A = 0.300$, $z \approx 0.84$.

Therefore, to find our raw score, $x = (z)(\sigma) + \mu = (0.84)(4) + 68 = 71.36$.

Hence a person must be at least 71.36 inches tall in order to be taller than 80% of the adult population in the Fargo metro area.

Extra Credit: (5 points) A company with 100 employees currently has a mean salary of 50,000 per year. How many employees each making exactly 100,000 per year would need to be laid off in order for the new mean salary of the company to be less than 45,000 after these layoffs? [Find the minimum number of layoffs needed]

Using algebra, suppose we lay off x employees. To find the new mean salary, we subtract the salary of the laid off employees from the total of the salaries at the company and then divide by the new lower total number of employees.

$$\text{Then } \bar{x} = \frac{100(50,000) - x(100,000)}{100 - x} = 45,000.$$

Then $45,000(100 - x) = 5,000,000 - 100,000x$, or, $4,500,000 - 45,000x = 5,000,000 - 100,000x$. Then, rearranging terms, $55,000x = 500,000$, so $x \approx 9.09$.

Therefore, 10 employees each making 100,000 per year would need to be laid off in order to get the average salary below 45,000 per year.

Note: This problem could also be solved using guided guessing, provided you showed enough detail about how you know the answer you found is correct.