1. A recent poll conducted from Nov. 11-15, 2010 collected data from 1,000 adults randomly selected from throughout the U.S. The margin of error in this poll was ±3.1 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, 47.3% were in favor of the proposal and 52.7% 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 3.1% margin of error, we can see that somehwere between 44.1% and 50.4% of the population are in favor of the propocal, and somewhere between 49.6% and 55.8% 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 that are against it.

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

   Selection Bias
   Non-response Bias
   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 72, 74, and 82 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 70?

   Let x be the missing score. To get an 70 average, we need to have $\dfrac{72 + 74 + 82 + x}{4} = 70$, or $72 + 74 + 82 + x = 280$. Therefore, we must have $228 + x = 280$, or $x = 52$.
   Hence the student would need to get 52 on the exam in order to end up with an average of 70 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 80?

Let x be the missing score. To get an 80 average, we need to have $\dfrac{72 + 74 + 82 + x}{4} = 80$, or $72+74+82+x = 320$. Therefore, we must have $228 + x = 320$, or $x = 92$.

Hence the student would need to get 92 on the exam in order to end up with an average of 80 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 $\dfrac{72 + 74 + 82 + 90}{4} = 79.5$, while the median is $\dfrac{74 + 82}{2} = 78$, so the student would prefer that the mean was used.

4. The following data set gives the daily high temperature in Fargo, ND from November 1st to November 20th, 2010 (source: www.wunderground.com) $\{52, 55, 54, 45, 49, 57, 62, 61, 66, 56, 47, 48, 39, 33, 34, 37, 34, 32, 33, 18\}$

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

```
1 | 8
2 |
3 | 2 3 3 4 4 7 9
4 | 5 7 8 9
5 | 2 4 5 6 7
6 | 1 2 6
```
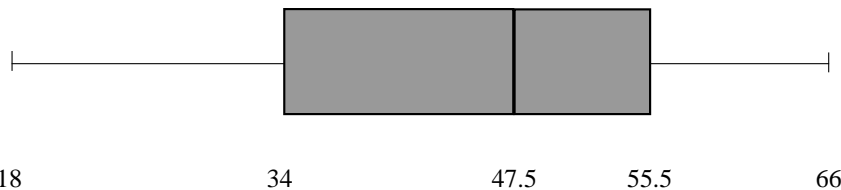
(b) (6 points) Find the mean and midrange of this data set.

mean: $\bar{x} = \dfrac{912}{20} \approx 45.6°F$

midrange: $\dfrac{18 + 66}{2} = 42°F$

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

min: 18, Q_1: 34, Q_2: 47.5, Q_3: 55.5, Max: 66

18   34   47.5   55.5   66

(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 (18). Because of this, we would expect that the median would be the best measure of center. In practice, the mean is $45.6°F$, the median is $47.5°F$, the midrange is $42°F$, and the mode is either 33 or 34 (two values are tied). The positive outlier seems to be effecting the midrange, so it is not the best measure. There are two modes, but both involve only 10% of the data, and do not seem to give a good representation of the "middle" in this case anyway. The median is slightly less than the mean, so the outlier is effecting the mean to some degree, so the median appears to be the best measure of center for this 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.

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

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

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

standard deviation: $s = \sqrt{\dfrac{218}{19}} \approx 3.39$

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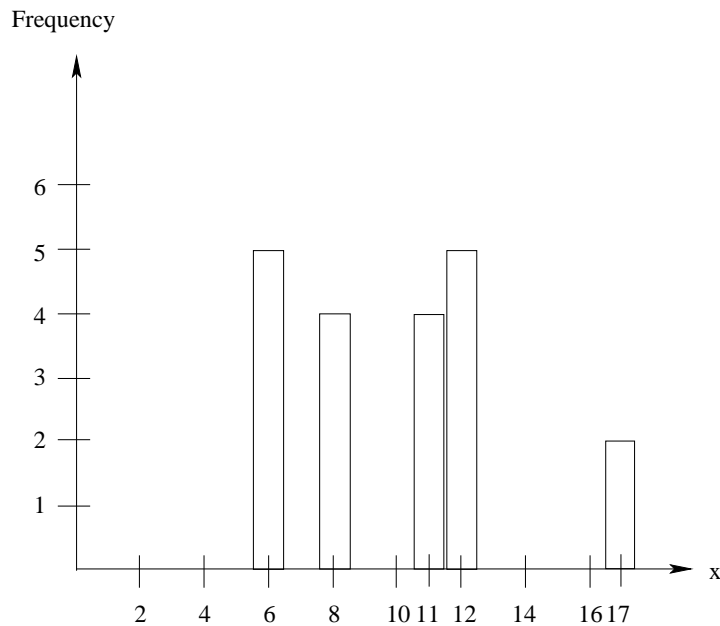
6. (5 points each) Suppose that the height of adults in the Fargo metropolitan area is normally distributed with a mean of 68 inches and a standard deviation of 4 inches. Also suppose that the total adult population of the Fargo metropolitan area is 150,000 people.

(a) What *percentage* of the population is *between* 68 and 74 inches tall?
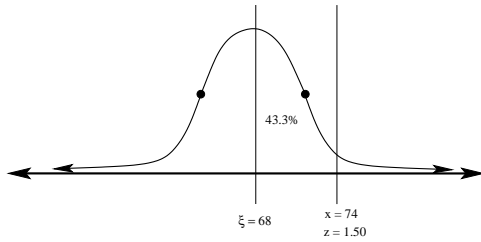
$$z = \frac{74 - 68}{4} = 1.50. \text{ Therefore, using the } z\text{-table and the diagram above, } A = 0.433, \text{ or } 43.3\%.$$
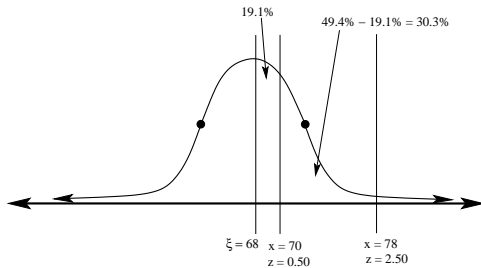
(b) What *percentage* of the population is *less than* 60 inches tall?
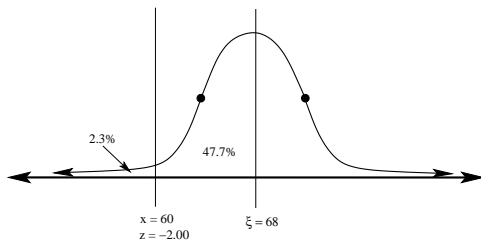
$$z = \frac{60 - 68}{4} \approx -2.00. \text{ From the } z\text{-table, we see } A = 0.477.$$

Since we want the percentage of the population whose height is **less than** 60 inches, we subtract $50 - 47.7 = 2.3\%$.
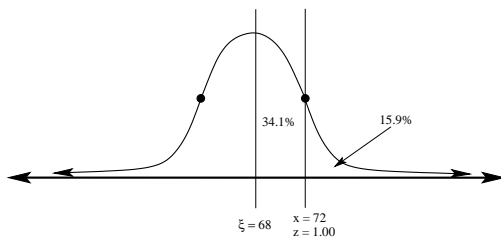
(c) What *percentage* of the population are *between* 70 and 78 inches tall?

$$z_1 = \frac{70 - 68}{4} = 0.50, \ z_2 = \frac{78 - 68}{4} = 2.50, \text{ so } A_1 = 0.191 \text{ and } A_2 = 0.494$$

Therefore, using the z-table and the diagram above, $A = 49.4\% - 19.1\%$, or 30.3%.

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

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

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* 99% of the people in the Fargo metro area?
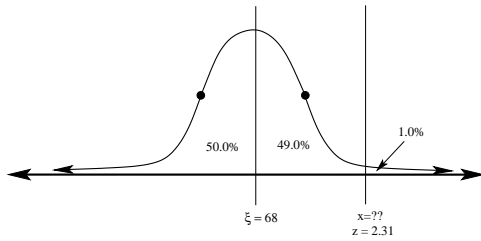
We want to find the x-score so that 99% of the population is below this score. We see from the z-table that when $A = 0.490$, $z \approx 2.31$.

Therefore, to find our raw score, $x = (z)(\sigma) + \mu = (2.31)(4) + 68 = 77.24$.

Hence a person must be at least 77.24 inches tall in order to be taller than 99% of the adult popoulation 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.

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

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.