Statistics Assignment: Differences Between Mean and Median

1. You're given $n = 5$ measurements: 0,7,1,1,4. Find the mean, median, and mode. Show your work.

You can calculate the mean by adding all the values and dividing:

$$(0 + 7 + 1 + 1 + 4)/5 = \boxed{2.6}$$

You can calculate the median by ordering the values and finding the middle:

$$0, 1, \boxed{1}, 4, 7$$

You can calculate the mode by ordering the values and finding the middle:

$$0, \boxed{1, 1}, 4, 7$$

2. The VCR is a common fixture in most American homes. In fact, most American households have VCR, and many have more than one. You take a simple random sample of households and get these numbers – each one represents a count of VCR from one house. Based on these data and the fact that they represent a simple random sample, how many VCR would you say most American households have? Which measure of center did you use, and why?

I used the mean because the distribution is mostly symmetrical, with no outliers. There are 59 families with 1 VCRs, 20 families with 0 VCRs, and 21 families with 3 VCRs. Therefore, the mean is

$$((59 \cdot 1) + (20 \cdot 0) + (21 \cdot 3))/100 = \boxed{1.22}$$

Based on this data, I would say that *most* American households have 1 VCR.

3. The top ten movies and their profits (in millions) from the weekend of May 23-25,1997 are presented here. Find the mean, median and which one represents the center of data better.

MEAN

$$\frac{(90.2 + 11.4 + 8.0 + 5.6 + 5.4 + 4.7 + 3.0 + 2.3 + 2.1 + 1.73)}{10} = \boxed{13.42 \text{ million}}$$

MEDIAN

$$1.73, 2.1, 2.3, 3.0, 4.7, \boxed{(5.05)}, 5.4, 5.6, 8.0, 11.4, 90.2$$

The median better describes the center of the data, as the outlier present in the profits of *Jurassic Park* increase the mean by a lot.

4. The school system has a salary dispute: schoolteachers and administrators claim they're being underpaid, while school district officials disagree.

   a. It would be best for the district budgeting officials to use the **mean** salary because some of the administrators earn much more than some of the teachers, so the district officials can pretend that the teachers make a high salary already.

   b. The teachers should use the median in order to rule out the outliers caused by the salaries earned by top administrators.

   c. The median is probably a better way to measure the pay for most teachers and administrators because it rules out the high outliers, which skews the data to the right.

5. In a psychological experiment, the time on task was recorded for ten subjects.

   MEAN

   $$\frac{175 + 200 + 190 + 185 + 250 + 190 + 230 + 225 + 240 + 265}{10} = \boxed{215}$$

   MEDIAN

   $$175, 185, 190, 190, 200, \boxed{(212.5)}, 225, 230, 240, 250, 265$$

   I would use the mean because it seems like this data isn't very skewed, and the mean would be the best measure of central tendency for symmetrical data.

   If there was an extra subject with a time of 784 seconds, the mean would be $\boxed{266}$ and the median would be $\boxed{225}$. The scientists might decide to throw out this outlier because their hypothesis might depend on a shorter amount of time to do a task based on changing another variable. Therefore, a time of 784 seconds wouldn't help their cause.

6. Consider the following histogram of class GPAs. Which measure of central tendency would be greater, the mean or the median, and why?

The mean would be greater as the data is skewed towards the right. However, the median wouldn't change much even if the data was symmetrical.

7. Suppose the yearly salaries in a small law firm are as follows:

MEAN

$$\frac{45000 + 50000 + 50000 + 65000 + 50000 + 280000}{6} = \boxed{90{,}000}$$

MEDIAN

$$45000, 50000, 50000, \boxed{(50000)}, 50000, 65000, 280000$$

You can increase the outlier salary (280,000) to 1 million to increase the mean, and also decrease the other 6 salaries to somewhere below 30,000.

You can decrease the outlier salary to 10,000 to decrease the mean and increase the other 6 salaries to above 50,000 to increase the median.