Math 102

Exam 4 Practice Problem Solutions

1. A group of researchers wishes to find out which professional sport is currently the most popular in the United States. To shed some light on this, they decide to go to the Mall of America on a Sunday afternoon and ask people the following question: "Is football your favorite professional sport, or do you prefer a different sport like basketball, baseball, hockey, or soccer?" Although some people refuse to answer their question, they eventually get 100 responses. Of those that responded, 53 prefer football, 20 prefer basketball, 18 prefer baseball, 7 prefer hockey, and 2 prefer soccer.

   (a) What is the population in this survey? What is the sample?

   The population is all people in the United states (since this is the group the researchers want to find out about).

   For the sample, I would accept either the 100 people who actually answered the survey, or, better yet, all the people who were asked the survey question, including those who refused to answer.

   (b) What forms of bias, if any, may have effected the data collected in this survey? Explain your reasoning.

   Selection bias: Only people in a specific location at a specific time had an opportunity to answer the survey. This was not a random sample of all people in the U.S.

   Leading Question bias: The way the question is phrased both seems to limit the choices and makes football the predisposed response.

   Non-response bias: It is implied in the description above that several people refused to participate in the survey.

   (c) Based on this study, what conclusions, if any, can be reached about which sport is most popular in the United States today? Explain your reasoning.

   Although it is likely that football is the most popular sport among people in the U.S., due to the numerous design flaws in this study, no clear conclusions can be drawn from the data collected in this survey.

2. (a) Give an example of a real life situation where the mean is the most appropriate measure of central tendency.

   I would accept any clearly described example that does not have a lot of outliers.

   For example, the mean is a good measure of center to describe the average number of M&Ms in a snack size bag, or an individual student's average on 4 exams.

   (b) Give an example of a real life situation where the median is the most appropriate measure of central tendency.

   The median is a more appropriate measure of center in situations where the data has outliers. For example, the median would be best to use to measure the average household income in the U.S., since outliers like Bill Gates and other billionaires and multi-millionaires would skew the mean quite a bit.

   (c) Give an example of a real life situation where the mode is the most appropriate measure of central tendency.

   The mode is the best measure of center is situations where we only want to know the most popular response, and in situations where the data collected is not numerical. One example would in the data collected in an election or a pre-election poll.

3. A company with 19 employees has a mean salary of \$38,000. Suppose a new employee is hired at a salary of \$28,000

   (a) What is the mean salary for the employees in this company after the new hire?

   Since there are already 19 current employees whose mean salary is \$38,000, to find the new mean, we must compute $\dfrac{(19)\cdot(\$38,000)+\$28,000}{19+1} = \dfrac{750,000}{20} = \$37,500$

   (b) What, of anything can we determine about the impact of this new hire on the median salary in this company? Explain your reasoning.

   Since we don't know the actual salaries of the original 19 employees, it is not possible to be precise about how the addition of the new employee impacts the median salary of this company. We suspect that this new salary is in the lower half of all salaries, in which case that median would either go down, or it could possibly stay the same (if there are repeated salaries in the middle of the data). If there are any salaries that are a lot above the mean, for example if the president of the company makes \$250,000 a year, then the new hire may actually be in the top half of the data, in which case, the new hire may raise the median.

4. Given the data set $\{4, 11, 17, 25, 23, 19, 17, 7, 38, 20\}$

5. (a) Find the mean, mode, and midrange of this data set.

   Mean: $\frac{4+11+17+25+23+19+17+7+38+20}{10} = \frac{181}{10} = 18.1$

   Mode: 17, the only data point that occurs twice.

   Midrange: $\frac{\text{max + min}}{2} = \frac{38+4}{2} = 21$

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

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

   (c) Find the 5 number summary of this data set.

   First, we put the data in order: $\{4, 7, 11, 17, 17|19, 20, 23, 25, 38\}$

   min: 4

   Q1: 11

   Median: $\frac{17+19}{2} = 18$

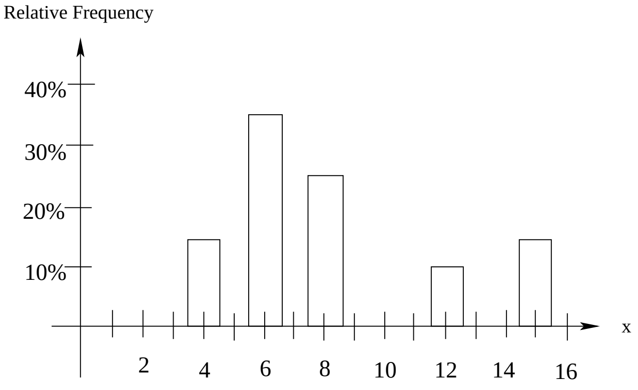
   Q3: 23

   Max: 38

6. Given the following frequency table:

x	frequency	rel. freq
4	3	.15
6	7	.35
8	5	.25
12	2	.10
15	3	.15

Relative Frequency Histogram:



(a) Complete the relative frequency column in the table given above.

Note: These are computed by dividing the frequency in each row by 20, the total number of data points, giving the proportion on times each type of data point occurs (See table above)

(b) In the space provided above, make a relative frequency histogram for the data in the table above.

Note: In a relative frequency histogram, the horizontal scale represents each value represented among our data points, and the vertical scale gives the relative frequencies of each value. (See figure above)

(c) Compute the mean and median of the data in this table.

mean: $\overline{x} = \dfrac{(4)(3) + (6)(7) + (8)(5) + (12)(2) + (15)(3)}{20} = \dfrac{163}{20} = 8.15$

median: since there are 20 data points, we average the 10th and 11th entries $\frac{6+8}{2} = 7$

7. Find the mean and standard deviation of the data set: $\{2, 5, 7, 12, 14\}$

x	$x - \overline{x}$	$(x - \overline{x})^2$
2	$2 - 8 = -6$	36
5	$5 - 8 = -3$	9
7	$7 - 8 = -1$	1
12	$12 - 8 = 4$	16
14	$14 - 8 = 6$	36

First notice that $\overline{x} = \frac{2+5+7+12+14}{5} = \frac{40}{5} = 8$.

Also, $n = 5$, so $n - 1 = 4$

By adding the entries in the last column of table on the left,

we see that $\Sigma(x - \overline{x})^2 = 98$.

Then $s = \sqrt{\dfrac{\Sigma(x - \overline{x})^2}{n-1}} = \sqrt{\dfrac{98}{4}} \approx 4.95$

8. Suppose Stock A had a average price of \$25.50 last year, with a standard deviation of \$3.50, which Stock B had an average price of \$76.00 last year, with a standard deviation of \$10.

(a) Find the coefficient of variance for each of these stocks last year.

Recall that the coefficient of variance can be found by dividing the standard deviation of a data set by its mean, and then multiplying by 100%. This gives a measure of how much the data varies in comparison with the average size of a data point.

Therefore, for Stock A, $CV = \frac{3.50}{25.50} \cdot 100\% \approx 13.725\%$.

Similarly, for Stock B, $CV = \frac{10.00}{76.00} \cdot 100\% \approx 13.158\%$.

(b) Which stock was more volatile? Explain your reasoning.

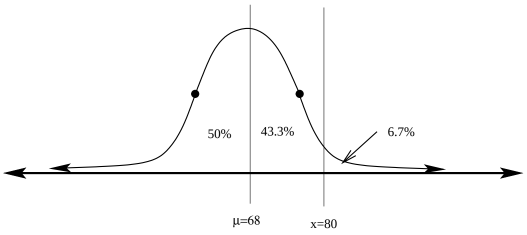
Based on the coefficients of variance computed above, Stock A is slightly more volatile that Stock B.

9. Suppose that 500 test scores (on a 100 point test) are approximately normally distributed with a mean of 68, and a standard deviation of 8.

(a) What percentage of scores are above 80 points?

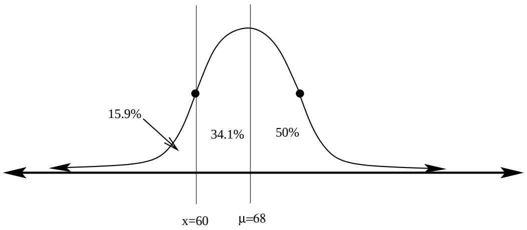
Since $\mu = 68$ and $\sigma = 8$ for this population, we can compute the z-score for 80.

$z = \frac{80-68}{8} = 1.5$. From the z-table, we look up this entry and find the corresponding area to be $A = .433$, or 43.3%. This represents the area under the normal curve between $x = 68$ and $x = 80$. To find the percentage of scores *above* 80, we must subtract: $50 - 43.3\% = 6.7\%$.



(b) How **many** scores are below 60 points?

First notice that the z-score for $x = 60$ is $z = \frac{60-68}{8} = -1$. From the z-table, we look up this entry and find the corresponding area to be $A = .341$, or 34.1% (or, just use the 68% rule). This represents the area under the normal curve between $x = 60$ and $x = 68$. To find the percentage of scores below 60, we must subtract: $50 - 34.1\% = 15.9\%$. Finally, we multiply this percentage by the total number of scores in order to find the *number* of scores below 60: $(500) \cdot (.159) \approx 80$ scores.

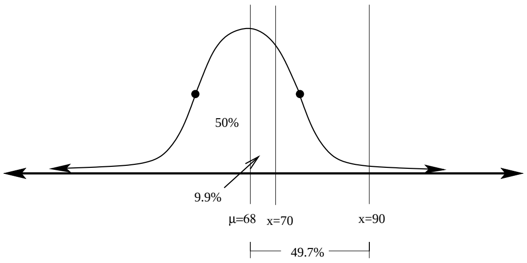


(c) What percentage of scores are between 70 and 90 points?

Here, we compute the z-score for each of these x-values:

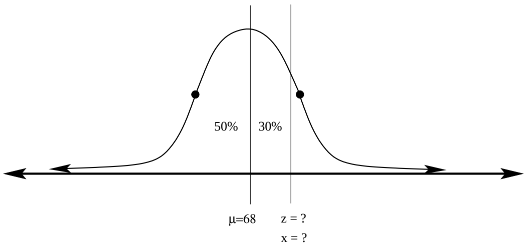
$z_1 = \frac{70-68}{8} = .25$, and $z_2 = \frac{90-68}{8} = 2.75$.

Next, we look up the area values for each of these in the z-table: $A_1 = .099$ and $A_2 = .497$



From the figure above, we see that to find the percentage of scores between the two, we subtract the areas we found above: $.497 - .099 = .398$, or 39.8%.

(d) What score would a person need to get on the test in order to have scored higher than 80% of the people who took this test?



We need to find the x-value which is above 80% of all scores. To do this, we first need to find the appropriate z-score. We look in the table for the z-score whose area correxponds to 30%, and find that $A = .300$ when $z = .84$. Working backwards from this, $x = z \cdot \sigma + \mu = (.84)(8) + 68 = 74.72$.