Math 107 Sample Exam number 1, Fall 2009

Dr. Richard Mikula, Instructor

October 4, 2009

Contents

1 Tables and Graphs

Suppose that the 6 p.m. temperatures (in degrees Fahrenheit) for a 36-day period are as follows:

$$72, 78, 86, 93, 101, 96, 98, 82, 81, 77, 67, 72,$$

$$81, 85, 82, 83, 76, 78, 73, 81, 86, 80, 76, 60,$$

$$56, 45, 70, 81, 82, 77, 73, 76, 80, 81, 77, 70$$

1. Draw a frequency histogram for this data set, by first filling in the table

Temperature Range	Frequency
< 40	
40-49	
50-59	
60-69	
70-79	
80-89	
90-99	
100-109	
110 +	

2. Does the histogram appear to be skewed to the right, or skewed to the left (if not skewed symmetric or symmetric, please say so)?

3. Does there appear to be any outliers?

To examine the answers to this problem, we must first reorder the elements:

$$45, 56, 60, 67, 70, 70, 72, 72, 73, 73, 76, 76, 76,$$

$$77, 77, 77, 78, 78, 80, 80, 81, 81, 81, 81, 81$$

$$82, 82, 82, 83, 85, 86, 86, 93, 96, 98, 101$$

Filling in the table, we may then see what the histogram looks like:

Temperature Range	Frequency
< 40	0
40-49	1
50-99	1
60-69	2
70-79	14
80-89	14
90-99	3
100-109	1
110 +	0

The histogram appears to be slightly skewed to the left. We can better see this after we calculate the mean $\bar{x}$ and the median.

To see if there are any outliers, I will use a simple test, which requires that we know the first and third quartiles Q_1, Q_3.

By inspection, it seems that 45 is an outlier. However, the rest of the extreme values in our data set are some what questionable. Thus, we need a quantitative test for outliers.

One such test is to first compute the difference $Q_3 - Q_1$, which is the third quartile minus the first quartile.

Then take this number and multiply by 1.5. That is compute

$$1.5 \cdot (Q_3 - Q_1).$$

For simplicity, call this d. Then numbers less that $Q_1 - d$ and numbers greater than $Q_3 + d$ are considered outliers.

In our case, let us compute Q_1, Q_3:

For Q_1, we take $p = 0.25$. In this case $np = 36 \cdot 0.25 = 9$. Since this is an integer, Q_1 is the average of the 9th and 10th term in our ordered data set. That is

$$Q_1 = \frac{73 + 73}{2} = 73.$$

For the third quartile Q_3 we take $p = 0.75$. Thus $np = 36 \cdot 0.75 = 27$. Since this is an integer, we take the third quartile to be the average of the 27th and 28th terms in our ordered data set. That is,

$$Q_3 = \frac{82 + 82}{2} = 82.$$

Next $d = 1.5(82 - 73) = 13.5$. Thus, any value in our data set less than

$$Q_1 - d = 73 - 13.5 = 59.5$$

will be considered an outlier. Also, any number in our data set greater than

$$Q_3 + d = 82 + 13.5 = 95.5$$

will be considered an outlier. Thus, our outliers are

$$45, 56, 96, 98, 101.$$

2 Measures of Center

For the data set given above (the temperatures at 6 p.m.):

1. Find the mean and median.
2. If there are any outliers, what is the mean and median for the data set after you dispose of these values.
3. Also, based on what you observe here. Is the histogram skewed to the left or right? Please explain.

The mean $\bar{x}$ is the average of the numbers in our data set. In this case, it is

$$\bar{x} = \frac{2812}{36} \approx 78.1.$$

The median is the 50th percentile. Here we compute $np = 36 \cdot 0.5 = 18$. Thus the median is the average of the 18th and 19th terms in our ordered data set

$$\text{median} = \frac{78 + 80}{2} = 79.$$

Since the median is larger than the mean $\bar{x}$, we say that the histogram is skewed to the left (although only slightly).

If we throw away the outliers, there are 31 terms remaining. To compute the median for this new data set we look at $np = 31 \cdot 0.5 = 15.5$. Thus the median is the 16th term in the ordered data set without the outliers. That is

$$\text{median} = 78.$$

The mean of the data set without the outliers is

$$\bar{x} = \frac{2416}{31} \approx 77.935.$$

3 Measures of Dispersion

For the temperatures at 6 p.m. data set:

1. Compute the standard deviation for the data set with and without any outliers.
2. Find the range for the data set with and without outliers.
3. Find the first and third quartiles.
4. Find the 67th percentile.
5. For $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9$ find the $100p$th percentiles. Using these numbers, does there appear to be and clumping in the data set? Please explain.

We have already found $Q_1 = 73$ and $Q_3 = 82$. Let us find the other percentiles first:

The 67th percentile: Here $p = 0.67$. Thus $np = 36 \cdot 0.67 = 24.12$. thus the 67th percentile is the 25th term in our ordered data set, which is 81.

For the other percentiles q look at the table below:

p	np	q
0.1	$36 \cdot 0.1 = 3.6$	67
0.2	$36 \cdot 0.2 = 7.2$	72
0.3	$36 \cdot 0.3 = 10.8$	76
0.4	$36 \cdot 0.4 = 14.4$	77
0.5	$36 \cdot 0.5 = 18$	median = 79
0.6	$36 \cdot 0.6 = 21.6$	81
0.7	$36 \cdot 0.7 = 25.2$	82
0.8	$36 \cdot 0.8 = 28.8$	83
0.9	$36 \cdot 0.9 = 32.4$	93

There appears to be clumping around 76 or 77 and again around 81, 82, 83. We see this by how the q values only increase slightly around $p = 0.3$ and $p = 0.6$

The range of the data set with the outliers is $101 - 45 = 56$. The range of the data set without the outliers is $93 - 60 = 33$.

Now, we compute the standard deviation (with outliers), which we do with the aide of a table:

i	data value a_i	data value squared a_i^2
1	45	$45^2 = 2025$
2	56	$56^2 = 3136$
3	60	$60^2 = 3600$
4	67	$67^2 = 4489$
5	70	$70^2 = 4900$
6	70	$70^2 = 4900$
7	72	$72^2 = 5184$
8	72	$72^2 = 5184$
9	73	$73^2 = 5329$
10	73	$73^2 = 5329$
11	76	$76^2 = 5776$
12	76	$76^2 = 5776$
13	76	$76^2 = 5776$
14	77	$77^2 = 5929$
15	77	$77^2 = 5929$
16	77	$77^2 = 5929$
17	78	$78^2 = 6084$
18	78	$78^2 = 6084$
19	80	$80^2 = 6400$
20	80	$80^2 = 6400$
21	81	$81^2 = 6561$
22	81	$81^2 = 6561$
23	81	$81^2 = 6561$
24	81	$81^2 = 6561$
25	81	$81^2 = 6561$
26	82	$82^2 = 6724$
27	82	$82^2 = 6724$
28	82	$82^2 = 6724$
29	83	$83^2 = 6889$
30	85	$85^2 = 7225$
31	86	$86^2 = 7396$
32	86	$86^2 = 7396$
33	93	$93^2 = 8649$
34	96	$96^2 = 9216$
35	98	$98^2 = 9604$
36	101	$101^2 = 10201$
$n = 36$	$\bar{x} = \frac{1}{n} \sum a_i = \frac{2812}{36} \approx 78.11$	$\sum a_i^2 = 223712$

Thus we see that the standard deviation is :

$$s = \sqrt{\frac{1}{n-1}(\sum_{i=1}^{n} a_i^2 - n \cdot \bar{x}^2)}$$

$$\approx \sqrt{\frac{1}{35}(223712 - 36 \cdot (78.11)^2)}$$

$$\approx \sqrt{\frac{1}{35}(223712 - 36 \cdot 6101.17)}$$

$$\approx \sqrt{\frac{1}{35}(223712 - 219642.12)}$$

$$= \sqrt{\frac{1}{35}(4069.88)}$$

$$\approx \sqrt{116.28}$$

$$\approx 10.78.$$

To calculate the standard deviation without the outliers, we simply take the above table and cut out the first two rows and the last three rows, getting:

i	data value a_i	data value squared a_i^2
3	60	$60^2 = 3600$
4	67	$67^2 = 4489$
5	70	$70^2 = 4900$
6	70	$70^2 = 4900$
7	72	$72^2 = 5184$
8	72	$72^2 = 5184$
9	73	$73^2 = 5329$
10	73	$73^2 = 5329$
11	76	$76^2 = 5776$
12	76	$76^2 = 5776$
13	76	$76^2 = 5776$
14	77	$77^2 = 5929$
15	77	$77^2 = 5929$
16	77	$77^2 = 5929$
17	78	$78^2 = 6084$
18	78	$78^2 = 6084$
19	80	$80^2 = 6400$
20	80	$80^2 = 6400$
21	81	$81^2 = 6561$
22	81	$81^2 = 6561$
23	81	$81^2 = 6561$
24	81	$81^2 = 6561$
25	81	$81^2 = 6561$
26	82	$82^2 = 6724$
27	82	$82^2 = 6724$
28	82	$82^2 = 6724$
29	83	$83^2 = 6889$
30	85	$85^2 = 7225$
31	86	$86^2 = 7396$
32	86	$86^2 = 7396$
33	93	$93^2 = 8649$
$n = 31$	$\bar{x} = \frac{1}{n}\sum a_i = \frac{2416}{31} \approx 77.94$	$\sum a_i^2 = 189530$

Thus we see that the standard deviation for the data set without the outliers is :

$$\begin{aligned} s &= \sqrt{\frac{1}{n-1}(\sum a_i^2 - n \cdot \bar{x}^2)} \\ &\approx \sqrt{\frac{1}{30}(189530 - 31 \cdot (77.94)^2)} \\ &\approx \sqrt{\frac{1}{30}(189530 - 31 \cdot 6074.64)} \\ &= \sqrt{\frac{1}{30}(189530 - 188313.84)} \\ &= \sqrt{\frac{1}{30}(1216.16)} \\ &\approx \sqrt{40.54} \\ &\approx 6.37. \end{aligned}$$