Math 247: Test 1 (Take Home Key) **Name:_____**

(20 points)

- *This exam is due <u>at the beginning of class</u> on Thursday, 2/8/2018. You may work with other people in the class but not with tutors, other instructors, etc. <u>Be sure that all answers are written in your own words</u>; i.e., do not write verbatim the same answer as another student.*
- *Your work can be typed or NEATLY handwritten.*
- *Your work should incorporate all of the technology work mentioned below; i.e., **<u>copy and paste the Minitab work into your write-up.</u>***

Scoring will be based on organization of your work, accuracy, and thoughtful, well-written answers.

Answer all of the following questions on another piece of paper and <u>attach to this cover sheet</u>. Use complete sentences in your answers!

Pandora took a random sample of 25 listeners to find out how many minutes they spent listening to internet radio during the last week. The (sorted) results were as follows:

0	170	180	200	240	230
230	250	280	280	290	310
310	310	330	350	350	360
370	390	400	400	430	450
480					

1.) Use Minitab to graph a histogram from the data. Adjust the classes (bins) so that there are 9 classes (intervals) and they begin at 0. You'll need to double click on the bars of the histogram, select the "binning" tab, then select "cutpoint" and "number of intervals".

- Describe the distribution, including shape, spread, potential outliers, and "typical value".

2.) Use Minitab to find the Descriptive Statistics" for the data.

- Explain whether or not it be appropriate to apply the Empirical Rule to this data set. Write a sentence or two (complete sentences, please!) to explain and reference the histogram in your explanation.

  Based on the roughly symmetric and unimodal shape of the histogram and the observation that the mean and the median are close in value (they differ by only 10.4 minutes), we can reasonably apply the Empirical Rule. We would not expect the Empirical Rule to be a perfect fit, though, since the data is not perfectly symmetrical in its distribution.

- By hand, showing work, find the range of listening times that are 1 standard deviation from the mean, 2 standard deviations from the mean, 3 standard deviations from the mean:

  Fill these values by hand on your histogram.

- Determine how many actual data values (using the given Pandora data) fall in each of the ranges above, then compute what percent of the total (total of 25 data values) is in each range: Report these numbers and percentages as shown below:

  Actual data within 1 standard deviation of the mean: _____/25 = _____ %
  Actual data within 2 standard deviations of the mean: _____/25 = _____ %
  Actual data within 3 standard deviations of the mean: _____/25 = _____ %

- Relate the actual data percentages to those predicted by the Empirical Rule? Include the predicted values from the Empirical Rule as well as the percentages you found above.

3.) By hand, find the outlier limits (show work!) and explain what this tells you about any potential outliers.

- Use Minitab to graph a boxplot and explain whether or not it confirms your outlier analysis.

- Find the z-score for the outlier and interpret it in the context of the problem. In your explanation, connect the z-score to what it means to be an outlier.

Take Home Test Answers:

Pandora took a random sample of 25 listeners to find out how many minutes they spent listening to internet radio during the last week. The results were as follows:

0	170	180	200	240	230
230	250	280	280	290	310
310	310	330	350	350	360
370	390	400	400	430	450
480					

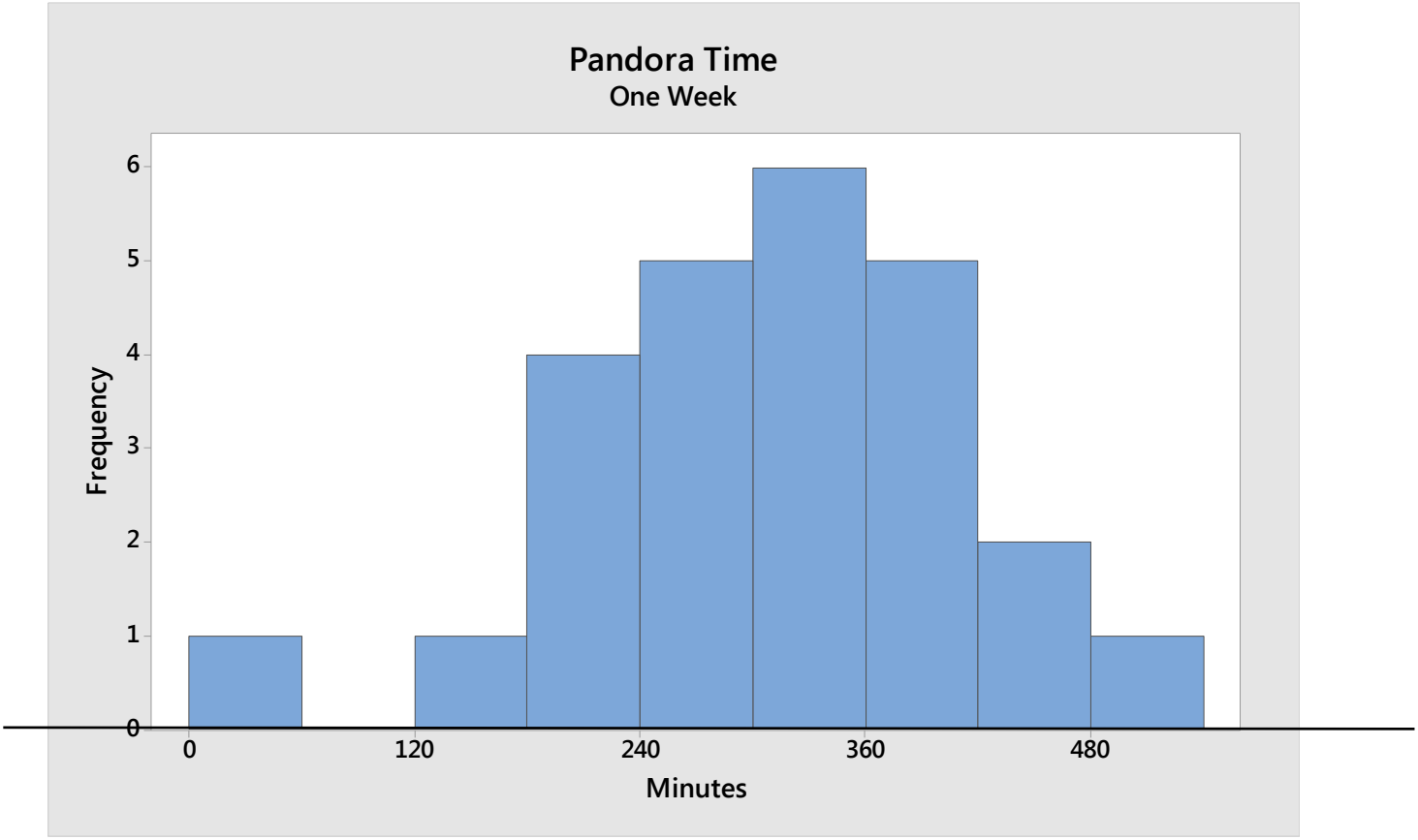
1.) **Histogram and analysis:**

Shape: The distribution is unimodal and roughly symmetric with some slight skewing to the left.

Spread: The data ranges between 0 and 450 minutes. The data is clustered around the Modal Class (tallest bar).

Outliers: There is a potential outlier of 0 minutes.

Typical value: The modal class (tallest bar) is between 300 and 360 minutes, so a "typical value" for listening time would be in this range, approximately 330 minutes (5 ½ hours), if we had to estimate.



2.) **Descriptive Statistics**

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Minutes	25	0	303.6	20.9	104.6	0.0	235.0	310.0	380.0	480.0

Applying the Empirical Rule:
Based on the roughly symmetric and unimodal shape of the histogram and the observation that the mean and the median are close in value (they differ by only 10.4 minutes), we can reasonably apply the Empirical Rule. We would not expect the Empirical Rule to be a perfect fit, though, since the data is not perfectly symmetrical in its distribution.

Ranges according to standard deviations:
According to the Descriptive Statistics values found using Minitab (see printout above), the mean listening time is 303.6 minutes and the standard deviation is 104.6 minutes.

Parking Lot
$\bar{x} = 303.6 \text{ min}$
$s = 104.6 \text{ min}$

Calculations:
1 S.D. from the mean:
$\bar{x} - 1s = 303.6 - 104.6 = 199 \text{ min}$ $\bar{x} + 1s = 303.6 + 104.6 = 408.2 \text{ min}$

2 S.D.'s from the mean:
$\bar{x} - 2s = 303.6 - 2(104.6) = 94.4 \text{ min}$ $\bar{x} + 2s = 303.6 + 2(104.6) = 512.8 \text{ min}$

3 S.D.'s from the mean:
$\bar{x} - 3s = 303.6 - 3(104.6) = -10.2 \text{ min}$ $\bar{x} + 3s = 303.6 + 3(104.6) = 617.4 \text{ min}$

Percentage of actual data captured in each range
Using the ORIGINAL DATA (see instructions!), we get the following results:

Actual data within 1 standard deviation of the mean: 18 /25 = 72%

Actual data within 2 standard deviations of the mean: 24 /25 = 96%

Actual data within 3 standard deviations of the mean: 25 /25 = 100%

Empirical Rule vs. Actual Percentages:

The Empirical Rule states that about 68% of the data should be within 1 standard deviation of the mean. We found that 72% of the actual data was located within one standard deviation. This is not perfect agreement between the theory (Empirical Rule) and the actual data values, but we expected this due to the lack of perfect symmetry of the distribution.

The Empirical Rule states that about 95% of the data should be within 2 standard deviations of the mean. We found 96% of the data located there so this is much better agreement! Note that this tells us that even with non-symmetrical distributions, finding data values beyond 2 standard deviations from the mean is unusual!

The Empirical Rule states that all or almost all (99.7%) of the data should be within 3 standard deviations of the mean. We did, in fact, find that 100% of the data fell within 3 S.D.s of the mean, so this is perfect agreement!

3.) **Finding outlier limits and outliers**

Variable	Minimum	Q1	Median	Q3	Maximum
Minutes	0.0	235.0	310.0	380.0	480.0

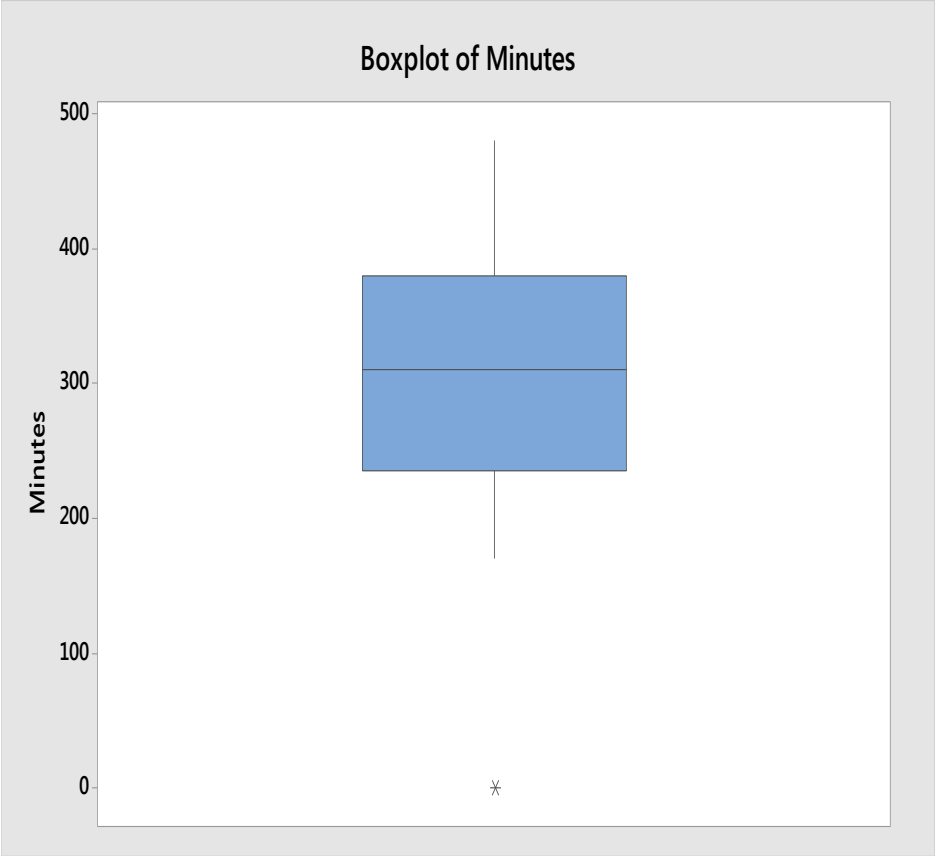
IQR = $Q_3 - Q_1$ = 380.0 – 235.0 = 145 minutes

Lower Outlier Limit = $Q_1 - 1.5IQR$ = 233 – 1.5(145) = 17.5 min

Upper Outlier Limit = $Q_3 + 1.5IQR$ = 380 +1.5(145) = 597.5 min

The listening time of 0 minutes is a lower outlier since it is below the lower fence.
There are no upper outliers.

The Minitab boxplot graph confirms our analysis since the value of 0 min is shown as an asterisk, indicating it is an outlier by Minitab's calculations.



z-score for 0 minutes:

$$z = \frac{x - \bar{x}}{s} = \frac{0 - 303.6}{104.6} = -2.902$$

The outlier of 0 minutes of listening time is 2.902 standard deviations below the mean. Since this value is well over 2 standard deviations from the mean (almost 3 standard deviations!), we see that it is a very unusual value in this data set.