The following questions will help you with the midterm questions. Keep all your work. Do not turn it in.

Question 1 is worth 24 points, questions 2 worth 4 points, question 3 worth 2 points. Total = 30 points.

1. At an outpatient-testing center, a sample of 20 days showed the following number of cardiograms done each day. 

a. Construct a grouped frequency distribution with 6 classes. (Let the first lower class limit be 1. It must include class limits, class frequencies, class boundaries, midpoints, relative frequencies, and percent cumulative frequencies in that order.)

b. Construct a one row stem-and-leaf display.

c. Draw a histogram.

d. Draw an Ogive (with the percent cum. freq. on the y-axis.)

e. Find the mean and the standard deviation.

f. Use part a to find the mean and standard deviation of your grouped frequency distribution.

g. The answers in parts e (using the raw data) and f (using the grouped frequency distribution) are usually different. Explain why.

h. Draw a box-and-whisker plot.

i. From the work you have done so far, do you think the distribution is symmetric? If not, what kind is it? Explain.

j. Use part d to find the 30th percentile.

k. Use part d to determine 20 cardiograms corresponds to what percentile.

l. Find the Pearson's Index of Skewness. What does it tell you about the shape of the distribution? Interpret the result.

2. The mean of a distribution is 50 with a standard deviation of 5. Use Chebyshev’s theorem to answer the following questions.

a. At least what percent of the data lie between 35 and 65?

b. Find the range in which at least 75% of the data values will be.
3. Given the following two set of data; determine which set of data is more variable.

Explain.

\[ \bar{X}_1 = 55 \quad \bar{X}_2 = 133 \]
\[ s_1^2 = 10 \quad s_2^2 = 12 \]

One way to organize a set of date is creating a **frequency distribution**. It is a table, which shows the number of times each number, or a group of numbers occurs. If the numbers are into groups, then it is called a **grouped frequency distribution**. Each group is called a class. In general number of classes is between 5 - 20. Remember that a gfd is a tool from which we gather a lot of information at a glance.

Can you think of disadvantages of a gfd?

The width of each class is called class interval.

Read section 2.1 to learn how to create a gfd plus other things (like midpoints, class boundaries, relative freq, and cumulative freq). You will need most of these later on.

There are two rules in section 2.4: Empirical Rule (which we study in length in chapter five) and Chebychev’s Theorem. Even though CT is very simple to understand and apply, students at first have difficulty to comprehend it.

**Chebychev’s Theorem:** At least \( 1 - \frac{1}{k^2} \) \( \cdot \) 100\% of any set of data lie within \( k \) standard

deviation about the mean.

The beauty of this theorem is that it applies any distribution with any shape.

Example: Suppose the mean of a population is 20 feet and the standard deviation is 3 feet. Then, at least 75\% of the data lie within 2 standard deviation about the mean (replace k with 2 in the above formula). The range for this population when \( k = 2 \) will be 14–26 (since \( \bar{X} - ks = 20 - 2(3) = 14 \) and \( \bar{X} + ks = 20 + 2(3) = 26 \)). The interpretation here is: At least 75\% of the data lie between 14 and 26.

**Coefficient of Variation:** It is a very good tool to compare the variability of two sets of

data. The formula is \( CV = \frac{s}{\bar{X}} \) or \( CV = \frac{\sigma}{\mu} \).

Example: Suppose you randomly select a sample form population A and a random

sample from population B. The statistics you got from the first sample are:

\( \bar{x}_1 = 10 \), \( s_1 = 2 \), and for the second sample they are: \( \bar{x}_2 = 20 \), \( s_2 = 3 \). You want to know

which set is more variable than the other one. Compare \( CV_1 \) with \( CV_2 \). \( CV_1 = \frac{2}{10} = .2 \),

and \( CV_2 = \frac{3}{20} = .15 \).

Since \( CV_2 < CV_1 \), the first set of data is more variable than the second one.

Remark:

a. The lecture for this chapter is very short. If there are concepts that you have a difficult
time to understand, please let me know.

b. Since there are some graphs, you might have some difficulty to email your work. If you face any difficulty, let me know.
c. Excel does an excellent job for graphs. Here is an example:

**Using Excel to Construct a Histogram**

Suppose we take a random sample of size 25 from Richland students. The variable of interest is the age of students. We want to have a grouped frequency distribution with 6 classes. Then, we want to use Excel to construct a histogram for this frequency table. Let the numbers in the sample be:

56, 32, 70, 36, 22, 18, 46, 15, 38,
27, 27, 25, 45, 52, 48, 33, 22, 19,
20, 40, 28, 37, 23, 33, 35

1. First find the class width: \( CW = \frac{\text{Largest} - \text{smallest}}{\text{# of classes}} = \frac{70 - 15}{6} = \frac{55}{6} = 9.17 \). Your class width is: \( CW=10 \). Note that we have rounded this number up.

2. Create your table as shown below. Note: Your first class limit could be 15, 14, 13, 12, or 11. But, you have to make sure that all of your data will fit in your table. I will use 11 for the lowest class limit (the numbers are nicer when we use this number for the lowest class limit).

<table>
<thead>
<tr>
<th>Class limits</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 - 20</td>
<td>4</td>
</tr>
<tr>
<td>21 - 30</td>
<td>7</td>
</tr>
<tr>
<td>31 - 40</td>
<td>8</td>
</tr>
<tr>
<td>41 - 50</td>
<td>3</td>
</tr>
<tr>
<td>51 - 60</td>
<td>2</td>
</tr>
<tr>
<td>61 - 70</td>
<td>1</td>
</tr>
</tbody>
</table>

Create this table in Excel and follow the following steps to construct a histogram for this table.

3. Highlight these two columns.
4. On the menu list select Insert, and click on Chart.
5. Now you see a few histograms. Select one by clicking on the histogram and then click on NEXT. At this point you will see a bar graph. Select NEXT.
6. Now you will see your bar graph with some options. The first one is the title. For our example the title could be Distribution of ages of 25 randomly selected students at RCC. The x-axis will be labeled as age and the y-axis is freq. On the right side of your graph you should see a small box with the freq in it, click on it and then press the delete button to get rid of it. At this point you have a bar graph like this:
The following steps will change the bar graph to a histogram:

1. Right click on a bar.
2. Select the FORMAT DATA SERIES.
3. Select OPTIONS.
4. Reduce the Gap width to 0.
5. Click OK.
You should see this graph now.

If you like to get rid of the gray back ground do this:
Right click on the gray area and then click on CLEAR.
This is what you will see:
Distribution of ages of 25 students at RCC

If you don’t like to see the horizontal lines do the following:
Right click on one of the lines and select clear. You will see a nice histogram.
Note that all these graphs are done in Excel, and then transferred to Microsoft word. There are other ways that you can use Excel to construct a histogram. You could have different colors for different rectangles by double clicking on the bar you want to have a different color (the last graph).

Good luck. If you still have questions, let me know.