Chapter 9
Notes for Instructors

Content

Statistics is the focus of this chapter. Students should be able to:

1. represent data using various types of graphs including line plots, stem and leaf plots, histograms, line graphs, bar graphs, pie charts, and pictographs.
2. find the total area of the columns in a histogram
3. find the area under a line graph using the corresponding histogram
4. understand when graphical representations of data are inaccurate or misleading
5. recognize the need for measuring both the central tendency and the spread of a data set
6. recognize the need for all three measures of central tendency
7. construct a box and whisker plot for a set of data
8. use box and whisker plots to compare sets of data
9. understand the need for using samples to make inferences about the population
10. understand the definition of a random sample
11. understand that it s difficult to obtain a random sample
12. determine whether a given sample is random or not and explain their answer
13. construct a relative frequency histogram and a relative frequency polygon for a set of data
14. understand that the area under a relative frequency polygon is one square unit
15. display an intuitive understanding of the normal distribution
16. apply the 69-95-99.7 rule for normal distributions

You will probably need about two to two and a half weeks for this chapter.

Manipulatives and Technology

There is very little reference to manipulatives in Chapter 9, but I do think the exercise on the top of page 559 which shows the mean as the balance point is a very good exercise. This exercise will not take up much of your class time and can be done with the wooden cubes in the Mathskellar.

There are certainly many ways that you can incorporate technology into your lessons in this chapter. I hope that someday UK will require students in this class to purchase a TI-73. I am very impressed with some of the things it can do (see the "Window on Technology" on page 547). We do have a few TI-73's in the Mathskellar, but not enough for a whole class. It would be great if we had a classroom set. There is also a free program from Peanut Software that does some nice things with Statistics. Peanut Software has several free programs that you might find useful in this class and others. Winstats is their statistics program. It was on the computers in CB 313. If it has been removed from those computers, you could send a Help request to have it reinstalled. In reality, you probably do not need to have the program installed in CB 313, since you do not need to have administrator privileges to install it. Simply request that your students be given accounts in CB313 and show them how to install it. This way they will be able to install the program on their home computers. I have included a Winstats instruction sheet with this documentation.

Notes and Suggestions:

Notes on Section 9.1: *The Graphical Representation of Data*

- Students have a lot of difficulty remembering the differences between a line plot and a line graph. (Frankly, I can't keep track of them myself sometimes.)

- My students needed to *see* that the area under the histogram is the same as the area under the line graph. I am including a large histogram with this documentation. You could have the students form small groups. Give each group a histogram. (In my experience, it is wise to have a few extras on hand.) Have the students draw the line graph on the histogram. They may need a straight edge to do this. You will need to make sure that they are connecting the midpoints of the tops of the rectangles. Also, be sure that they do not skip "rectangles" of height zero. Have one students cut out the area above the line graph. They need to do this carefully, being sure to cut only to the right of the graph, on the line graph or to the left of the graph. From this cut-out the student will need to cut out the small triangles from the histogram that fell above the line graph. Place these triangles under the line graph to see that the area of the histogram and the area under the line graph are equal.

- I think it is good to go over the *Choosing Good Visualizations* example on pages 544 and 546. It is good preparation for Chapter 12 which deals with Measurement. A lot of students in this course have difficulty with perimeter, area, and volume.

- I really like the questions in the *Thinking Critically* portion of the problem set. Often these questions are very difficult for students, but I believe that the question in this problem set are certainly well within our students reach. Moreover, I think that they lead to excellent discussion topic cut only to the right of the graph, on the line graph or to the left of the graph. From this cut-out the student will need to cut out the small triangles from the histogram that fell above the line graph. Place these triangles under the line graph to see that the area of the histogram and the area under the line graph are equal. (**NOTE:** We do have scissors in the Mathskellar.)

- I think it is good to go over the *Choosing Good Visualizations* example on pages 544 and 546. It is good preparation for Chapter 12 which deals with Measurement. A lot of students in this course have difficulty with perimeter, area, and volume.

- I really like the questions in the *Thinking Critically* portion of the problem set. Often these questions are very difficult for students, but I believe that the question in this problem set are certainly well within our students reach. Moreover, I think that they lead to excellent discussion topics.

Notes on Section 9.2: *Measures of Central Tendency and Variability*

- The examples at the very beginning of this section are very good and should help students understand why we need measures for both the central tendency and the variability of data.

- I like the block activity at the top of page 559. Students need to understand that the mean is a balance point. We do have blocks in the Mathskellar.

- Students should recognize that the most commonly reported measure of central tendency is the mean. Students should come to understand that the mean can be very misleading since it is easy to construct a data set in which no data item is close to the mean. This provides some motivation for the median and the mode. Similarly, neither the median nor the mode provides a good indication of the central tendency of the data without the other measures.

- Students should understand the information provided by each of the measures of variability.

- It should be emphasized that the most accurate portrayal of the data is given when all measures of central tendency and variability are reported. Students may argue that it is better to simply report the data. If they make this argument, I like to ask them how they would analyze the data in several lists each containing a thousand or more test scores. This provides some motivation for the measures given in this section since it is much easier to compare the means, medians, modes, ranges, and standard deviations than it is to compare the scores themselves.

- Students should be able to explain why the differences in the standard deviation formula are squared. (See problem number 7 in Section 9.2.)

- I think it is good to keep your eyes out for statistics in Newspapers and magazines while you are teaching this section. I found that my students had a lot of difficulty understanding the statistics that they are reading. For example, I included the following question on my third exam, I found that many students could not identify the mean nor recognize that the mode had to be less than $365 per year.

In the article "The World's 10 Worst Dictators" (David Wallechinsky, *Parade* magazine, February 22, 2004), the author states that in Equatorial Guinea "the per capita income is $4500 a year," but "60% of the people live on less than $1 a day."

- What is the average yearly income in Equatorial Guinea?
- What can you say about the median yearly income in Equatorial Guinea?
- Why are the mean and the median so different for this problem?
- Which do you think is more representative of the yearly salary in Equatorial Guinea, the median or the mean?
- Speakers and writers often use statistics to persuade their audiences. Moreover, it is not uncommon for speakers and writers to report only one of the measures of central tendency. For example, a journalist might report only the mean while ignoring the mode and the median. How would you respond to a speaker or a writer who uses only the mean to support his or her view? Use this example to justify your response.

Notes on Section 9.3:*Statistical Inference*

- Students should understand the difference between a population and a sample.
- Students should understand the need for using samples to make inferences about a population.
- Students should understand the definition of a *random sample*. (See problem numbers 2–8 in Section 9.3.)
- This section can lead to the topic of random number generators. I really like the questions in the *Teaching Concepts* portion of the Problem Set 9.3. (See problem numbers 13–15 in Section 9.3.)
- You could perform a binomial experiment in class — actually, you could have the students perform the experiment outside of class. In fact, I suggest that you have the students perform the experiment outside of class and make them bring their data to class. Have each student flip a coin 100 times and record their results. Suppose that we flip a coin n times and perform t trials. (Note that each student has 20 trials is $n = 5$.) Let the outcome of each experiment be the number of head obtained in a trial. As a class draw several line graphs for various values of n and t. As t increases, students should see that the line graph more nearly approximates the normal curve. If you want, you could do this for small n such as $n = 10$ with the class so that they understand the binomial distribution. Then you could take the class to the computer lab and use the `Probability` $\rightarrow$ `Binomial` program in Winstats. This should work nicely if they understand the histogram, plus it will save you the headache of having coins rolling all over the floor.

Worksheets

I have included two worksheets with this documentation. I had six different data sets for the first question on the second worksheet. I split the class into six groups and assigned each group one data set. I have included all six versions of the second worksheet.