MA 111 Review for Exam 6: Solutions

Exam 6 (given in class on Thursday, December 3) will cover Unit 6: Descriptive Statistics. Use Chapter 14 in the textbook as a study tool.

You should be familiar with the following key ideas:

- Understand how to make and/or interpret the following graphical representations of data:
  - frequency table
  - bar graph
  - histogram
  - pictogram
  - pie chart
- Remember the differences between continuous and discrete variables, and between numerical and categorical variables. Understand which of the above graph types are best suited for these different types of variables.
- Be able to critique a graph. What are some common ways that graphs can be made so they mislead the viewer?
- Know the definition of the mean (or average) and how to calculate it for a given data set.
- Know the definition of the pth percentile and how to find it for a data set.
- Understand the meanings of median, first quartile, and third quartile:
  - The first quartile Q_1 is the 25th percentile.
  - The median M is the 50th percentile.
  - The third quartile Q_3 is the 75th percentile.
- Be able to make and/or interpret a box-and-whisker plot.
- Know the definition of the range and the interquartile range.

Practice Problems

Use the odd exercises on pages 545-554 of the text to supplement these. You can check your answers in the back of the book.

1. Suppose you are given the following data table:

Favorite Colors of a Third Grade Class	
red	3
orange	1
yellow	4
green	3
blue	7
purple	2

   (a) Choose a graphical representation and use it to represent the data in the table.

   You could make a bar graph or a pie chart. A pictogram might also work, maybe with crayons making up the bars of the bar graph.

   Here's the bar graph:

   (b) Other than the graph you just chose, name another type of graph that would be appropriate for representing this data.

   As above, a bar graph, pie chart, or pictogram is appropriate

   (c) Name a type of graph that would be inappropriate for representing this type of data, and explain why.

   A histogram would be inappropriate, because the variable is not numerical and continuous. Remember that histograms are best for continuous variables like height, weight, etc.

2. Suppose you want to make a pie chart of the age of students in MA111. If 39% of the students are 19 years old, how big would the corresponding slice of the pie be? Calculate the size of the central angle of the slice.

   The central angle would be $39 \cdot 3.6° = 140.4°$.

3. Consider the following two data sets of exam scores for two different sections of a course:

$$\begin{aligned} \text{Class A} &= \{40, 52, 65, 66, 68, 71, 72, 73, 77, 77, 78, 78, 80, 81, 81, 83, 86, 87, 92, 99\} \\ \text{Class B} &= \{55, 65, 66, 68, 72, 72, 73, 74, 76, 78, 79, 79, 80, 81, 85, 85, 85, 87, 90, 95\} \end{aligned}$$

(a) Find the median, first quartile Q_1 and third quartile Q_3 for each class.

Both classes have $N = 20$ students. Notice the lists are already sorted.

Class A: The median is the average of the 10th and 11th values: $M = 77.5$.

The locator for Q_1 is $L = .25(20) = 5$. So Q_1 is the average of the 5th and 6th values: $Q_1 = 69.5$.

The locator for Q_3 is $L = .75(20) = 15$. So Q_3 is the average of the 15th and 16th values: $Q_3 = 82$.

For Class B, the locators are all the same because N is the same.

So $M = 78.5$, $Q_1 = 72$, and $Q_3 = 85$.

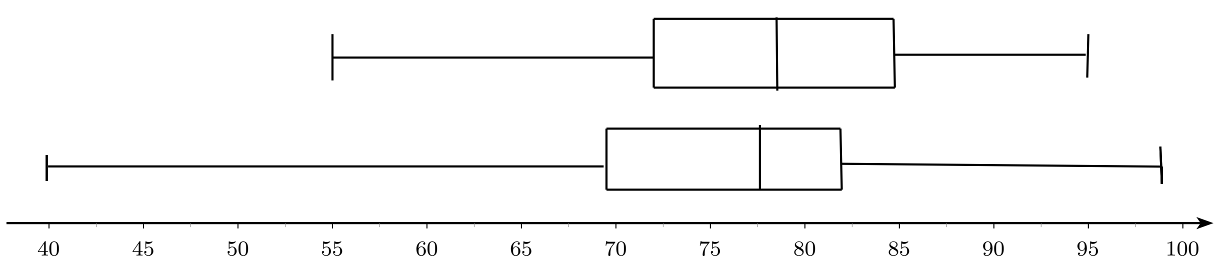
(b) The professor wants to recommend the top 10% of each class for a scholarship. How many students will she recommend? (Remember, she will recommend students from *each* class).

The top 10% of each class is found by computing the 90th percentile.

So $L = .9(20) = 18$. The 90th percentile is the average of the 18th and 19th values of each class.

So there are two students in Class A (they got a 92 and a 99), and two students in Class B (they got a 90 and a 95). Overall, the professor will recommend 4 students.

(c) Create box and whisker plots for both classes on a single axis.



4. Mike's average on the first five exams is 88. What must he earn on the next exam to raise his exam average to 90?

Since the first five exams averaged to be 88, the sum of all of those five exam scores was $88(5) = 440$.

Let the next exam grade be g. Then Mike wants:

$$\begin{aligned} \frac{440+g}{6} &= 90 \\ 440+g &= 90(6) \\ 440+g &= 540 \\ g &= 540-440 \\ g &= 100. \end{aligned}$$

5. A professor is teaching two sections of MA111. Section 1 has 35 students, and Section 2 has 28 students. On the last exam, the average score for Section 1 was 73, while the average score for Section 2 was 78.

   For all of the students combined, find the average score for the exam.

   The sum of scores in Section 1 was $73(35) = 2555$. The sum of scores in Section 2 was $78(28) = 2184$.

   So the average of all the scores is

$$\frac{2555+2184}{35+28} = \frac{4739}{63} = 75.2$$

6. Look at the hand-out of graphs I distributed at the beginning of this unit. Be able to identify the ways that each graph is misleading, and suggest a way to improve the graphs.

   Some of the ways graphs are misleading include:

   - altering the vertical axis (especially cutting it off before 0) in order to make a graph appear more dramatic than it actually is;
   - using items in a pictograph that actually change by volume or area, not height. The heights may be correct, but the resulting change in volume (or area) appears more dramatic.
   - not labeling or mislabeling axes.

   If you have any questions about a specific example, please ask.

7. Consider the following data set:

$$\{7, 59, 25, 27, 64, 70, 68, 11, 45, 5, 17, 45, 52, 21, 26\}$$

   (a) Find the 30th percentile of the data set.

   We first need to sort the data:

$$\{5, 7, 11, 17, 21, 25, 26, 27, 45, 45, 52, 59, 64, 68, 70\}$$

Notice $N = 15$ items in the list.
So $L = .3(15) = 4.5$. We round this up to 5.
The 30th percentile is 21.

(b) Find the 60th percentile of the data set.
$L = .6(15) = 9$. We take the average of the 9th and 10th values.
The 60th percentile is 48.5.

(c) Find the 95th percentile of the data set.
$L = .95(15) = 14.25$. We round up to 15.
The 95th percentile is 70.

(d) Find the 50th percentile of the data set.
This will be the median. Since N is odd, we take the middle value: 27.

8. Consider the data set $\{2, 18, 19, 22, 24, 24, 25, 78\}$.

   (a) Find the range.
   $78 - 2 = 76$.

   (b) Find the interquartile range.
   We first have to find Q_1 and Q_3. There are $N = 8$ data values.
   For Q_1, $L = .25(8) = 2$. So $Q_1 = 18.5$.
   For Q_3, $L = .75(8) = 6$. So $Q_3 = 24.5$.
   The interquartile range is $24.5 - 18.5 = 6$.

   (c) Why might we consider the interquartile range a better measure of the spread of this particular set?
   The minimum of 2 and the maximum of 78 seem to be outliers of this set. Most of the values are between 18 and 25. The interquartile range represents this fact better than the range.

9. Give an example of a data set with $N = 5$ with the median less than the mean.

   There are many correct answers:

$$\{3, 4, 5, 35, 40\}$$

   The median is 5 but the mean is 17.4.

10. Try questions 45 and 46 on pages 549-550 of the textbook. Be able to interpret box-and-whisker plots to answer similar questions.

    Question 46 has two typos. The question should read:

    (a) Fill in the blank: Of the 612 engineering graduates, at most ______ had a starting salary greater than **$45,000**.

(b) Fill in the blank: If there were 240 agriculture graduates with starting salaries of **\$35,000** or less, the total number of agriculture graduates is approximately ______.

The answers to 45 are in the back of the book.

For 46a, \$45,000 corresponds to Q_1. So 25% of the students had salary \$45,000 or less; thus 75% of the students had a salary greater than \$45,000. 75% of 612 is: $.25(612) = 459$ students.

For 46b, \$35,000 corresponds to Q_1. So 240 students make up 25% of the total agriculture grads. Therefore there are $240(4) = 960$ agriculture graduates in total.

11. In 2006, the median SAT score was $d_{756,155}$, where $\{d_1, d_2, \ldots, d_N\}$ denotes the data set of all SAT scores ordered from lowest to highest. Determine the number of students N who took the SAT in 2006.

    First, since the median was an actual data value, not the average of two data values, we know N should be odd. There were 756,154 students to the left of the median and 756,154 students to the right of the median.

    So overall there were $756154(2) + 1 = 1,512,309$ students.