Name: ______________________________

Review Quiz of Statistics Background Needed for 36-303: Surveys, Sampling and Society

- You have 30 minutes to do this quiz. Plan on spending about 5 minutes per page on the quiz. It's OK if you don't finish a page, just move on to the next one; return to the unfinished ones at the end if you have time.
- Some problems require a little arithmetic, but a calculator will not be needed.
- Don't forget to put your name at the top of this page.

1. Please list any statistics courses you've had (whether or not at Carnegie Mellon), starting with the most recent:

____

____

____

2. Figure 1 on the next page shows results of a midterm exam in an introductory statistics class, for the students who finished the test.

   (a) What is the name of the graph in Figure 1?

   (b) What shape does the distribution of this variable have, approximately?

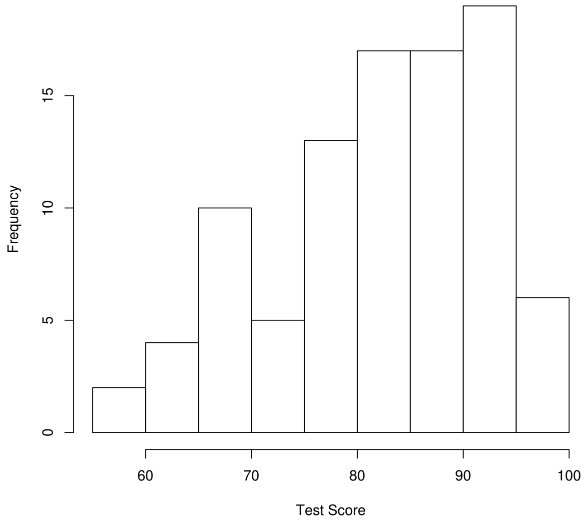

Figure 1: Distribution of midterm scores in an introductory statistics class.

(c) Approximately what is the mode on this test?

(d) How many people scored below 65 on this test? *(Assume no one scored exactly 65.)*

3. Here is some summary information on the same test:

N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
93	82.61	1.06	10.21	58.00	77.00	84.00	91.00	99.00

(a) In the summary, what does the first column ("N" with a "93" under it) mean?

(b) The summary tells us that the median of these test scores is 84. How was this number calculated?

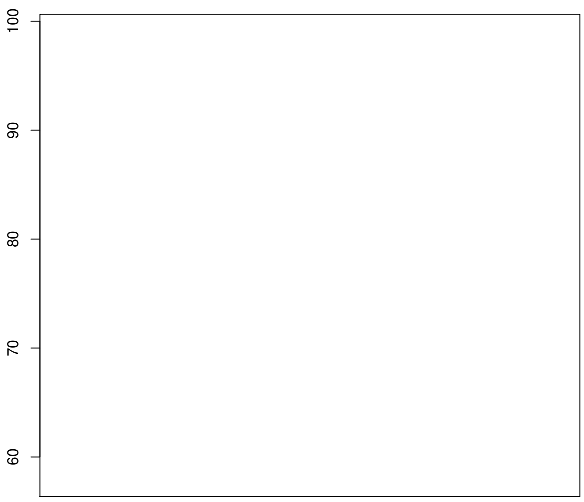

Figure 2: Boxplot of midterm scores introductory statistics.

(c) In the space provided in Figure 2, make a boxplot of the test scores, using the information provided in the summary. *(Assume there are no outliers.)*

(d) What is an outlier?

(e) How would you check for outliers using information in the summary?

(f) Using the information in the summary, calculate an approximate 95% confidence interval for the true mean score on this test.

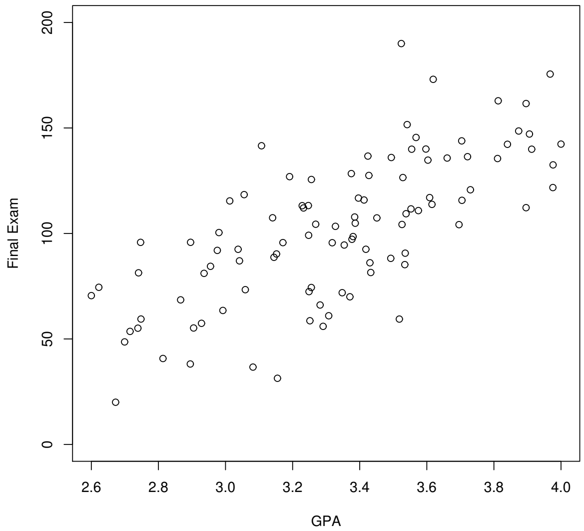

Figure 3: Relationship between students' GPAs and Final Exam Scores.

4. Figure 3 shows the relationship between students' GPA and their Final Exam Score in the introductory statistics class.

   (a) What is this graph called?

   (b) What numerical information does each dot in Figure 3 represent?

   (c) What kind of relationship is there between the two variables in Figure 3?

   (d) What statistic (single number) would you ask MINITAB or R to compute from this data, to quantify the relationship between variables like these?

5. Summation notation.

   (a) Which of the following is the formula for the sample mean? (Circle one.)

      i. $\sum_{i=1}^{n} x_i$

      ii. $\frac{1}{n} \sum_{i=1}^{n} x_i$

      iii. $\frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\overline{x})^2$

   (b) In the answer you chose, what is the role of the symbol i?

   (c) In the answer you chose, what is the role of the symbol x_i?

   (d) Please describe, in words, a step-by-step procedure to carry out the computation in this formula:

$$\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad (*)$$

   (e) What quantity does the formula in equation $(*)$ calculate?

6. Miscellaneous.

   (a) What is the difference between a *population* and a *sample*?

   (b) Suppose I have a box with 3 blue tickets and 2 red tickets, and you take a ticket out at random. What is the probability that you will get a red ticket?

   (c) Now suppose I have to pay you \$1 if you get a blue ticket and \$2 if you get a red one. Which of the following is a valid formula for the *expected value* of my payout to you (circle one)?

      i. $(1 + 2)/5$
      ii. $(3 \cdot 1 + 2 \cdot 2)/5$
      iii. $(1 \cdot 3 + 2 \cdot 2)/5$
      iv. $(3 + 2)/5$

   (d) Now suppose that we run the experiment in part (b) 100 times, each time putting the ticket back in the box, and we tally how many red tickets were drawn. Let the total number of red tickets be X. What is the name of the distribution of X?

   (e) The usual parameters for the distribution of X in this case are n and p. Give the values of n and p.

   (f) Which of the following is a valid formula for the expected value of X?

      i. $E[X] = np(1 - p)$
      ii. $E[X] = \sqrt{p(1 - p)/n}$
      iii. $E[X] = np$

   (g) What does $E[X]$ tell us about X?