

Statistics 2000, Section 001, Quiz 1 (200 Points)

February 11, 2000, Dr. Jürgen Symanzik

Your Name: _____

First look at all 4 questions. Then start with the question that looks easiest to you. Continue with a more difficult question. Try to answer as many questions as possible in these 50 minutes.

Note that you will obtain at least partial credit if you indicate a correct formula but your final result is incorrect. If you just rely on your calculator without indicating the formula that should be used and your result is incorrect, you will obtain no credit at all for this part of a question.

Question 1: Numbers and Graphs (90 Points)

As promised, here is one question from our homework assignments:

The Survey of Study Habits and Attitudes (SSHA) is a psychological test that evaluates college student's motivation, study habits, and attitudes toward school. A selective private college gives the SSHA to a sample of 18 of its incoming first-year college women. Their scores are:

154 109 137 115 152 140 154 178 101
103 126 126 137 165 165 129 200 148

The college also administers the test to a sample of 20 first-year college men. Their scores are:

108 140 114 91 180 115 126 92 169 146
109 132 75 88 113 151 70 115 187 104

Please answer the following questions:

1. Make a back-to-back stemplot of the **men's** and **women's** scores. (15 Points)

2. Are there any outliers? If so, circle these outliers on your stemplot. **(5 Points)**

3. Calculate the median \tilde{x} and range R for **men** and the median \tilde{x} and range R for **women**. **(10 Points)**

4. Calculate the arithmetic mean \bar{x} for **women** only. Note that the corresponding value for men is 121.25. **(10 Points)**

5. Calculate the standard deviation for **women** only. Clearly indicate whether you are calculating a **sample standard deviation** or **population standard deviation**. Note that the corresponding value for men is 32.85. **(10 Points)**

6. Find the five-number summaries for **men** and the five-number summaries for **women** and draw side-by-side boxplots based on these numbers. **(10 Points)**

7. Summarize your findings with respect to the SSHA scores for **men** and **women** in three or four sentences and draw an overall conclusion. **(15 Points)**

8. Do you think similar scores would be observed if we sample incoming first-year college women and men here in Logan? Explain your answer. **(15 Points)**

Question 2: Newspaper Graphics (35 Points)

Look at the following graphic which has been published originally in the *Washington Post*, Jan. 11, 1979. It has been republished in Howard Wainer (1997) "Visual Revelations".

Is there something wrong with it? If so, what?

Would you agree that the income growth is almost linear (i.e., data points fall close to a straight line) for both Physicians and Workers, however the slope for Physicians is much larger than the slope for Workers?

If you think the graphic is OK, you can stop here and just write **OK**. If you can think of a better representation of the same data, then go ahead and redraw the graphic (in this case, please use the next page).

Optional space to redraw the graphic from Question 2.

Question 3: Normal Distribution (30 Points)

Let Z be a standard Normal variable, i.e., $Z \sim N(0, 1)$, and X be a Normal variable with mean $\mu = 2$ and variance $\sigma^2 = 4$, i.e., $X \sim N(2, 2^2)$. Determine the following:
(5 Points each)

1. $P(Z < 1.48) =$

2. $P(-1.5 < Z < 1.0) =$

3. $P(X < 1.48) =$

4. $P(-1.5 < X < 1.0) =$

5. Find a number # so that
 $P(Z < \#) = 0.20$

6. Find a number # so that
 $P(X < \#) = 0.20$

Question 4: Scatterplot Matrix & Linked Brushing (45 Points)

The questions on the next page are based on the scatterplot matrix presented in William Cleveland's book "Visualizing Data". This scatterplot matrix has been reprinted below. It displays trivariate data that represents measurements of "Abrasion Loss", "Hardness", and "Tensile Strength" for 30 rubber specimens. "Abrasion Loss" is the dependent variable and "Hardness" and "Tensile Strength" are the independent variables. The goal of this study was to determine conditions that minimize the "Abrasion Loss".

1. Label the (individual) scatterplot that shows the “Abrasion Loss” on the horizontal (x-)axis and the “Tensile Strength” on the vertical (y-)axis with the letter “A”. **(5 Points)**
2. What is the (approximate) range R of “Tensile Strength”? **(5 Points)**

3. Describe the form, direction, and strength of the relationship in the scatterplots that show “Hardness” and “Abrasion Loss”. **(10 Points)**

4. Which of these statements is correct/incorrect/undecidable?
 - (a) The brush is located in the scatterplot that shows the “Tensile Strength” on the vertical (y-)axis and the “Hardness” on the horizontal (x-)axis. **(5 Points)**

 - (b) The low values (i.e., values in the approximate range 120–165) of “Hardness”, have been brushed. **(5 Points)**

 - (c) When “Hardness” is fixed at any given level, lower values of “Tensile Strength” result in a much higher value for “Abrasion Loss” than higher values of “Tensile Strength”. **(5 Points)**

5. Decide on **ONE** of the following options. The best way to minimize “Abrasion Loss” is a combination of **(10 Points)**
 - (a) high “Hardness” and low “Tensile Strength”
 - (b) high “Hardness” and high “Tensile Strength”
 - (c) low “Hardness” and low “Tensile Strength”
 - (d) low “Hardness” and high “Tensile Strength”