

Math 110
Test 1 Sample
September 26, 2008

Be sure to provide explanations for your answers as indicated.

1. (20 pts.) The article “Do We Really Know What Makes Us Healthy,” by Gary Taubes in the New York Times Magazine of September 16, 2007, concerns the difficulties of relying on observational studies to set treatment policy. In particular, it points out four confounding factors whose effects are difficult to determine.
 - (a) Define two of these four confounding factors.
 - (b) For the factors you defined in (a), explain why they might have been confounding factors in the Nurses’ Health Study observational study of hormone replacement therapy.
2. (20 pts.)
 - (a) Sketch a football shaped data cloud with $r \approx .6$.
 - (b) Sketch the SD-line on your plot. How do you know this is (roughly) the SD-line?
 - (c) Sketch the regression line on your plot. How do you know this is (roughly) the regression line?
3. (20 pts.) The following data on people without health insurance by family income level for 2006 is from the Current Population Survey.

Family Income	No Insurance	All People
	N= 44,815	N=293,834
Less than \$25,000	18,590	70,478
\$25,000 to \$49,999	13,620	72,963
\$50,000 to \$74,999	6,445	55,258
\$75,000 to \$200,000	6,160	95,136

Notice N= at the top of each column is the total number of people in the four class intervals for the column. Also, the number of people is in units of 1000.

- (a) For each of the four income class intervals, compute the percentage of No Insurance people in the class interval and the percentage of all people in the interval.
- (b) Use your percentages from (a) to construct two histograms, one for people without insurance and one for all people using the class intervals in the table. (Use the graph paper.)
- (c) By comparing your histograms, what can you say about the distribution of people without health insurance by family income in comparison to the distribution of all people by family income?

4. (20 pts.) The grades in a large university statistics class are normally distributed with a mean of 60 (out of a possible 100 points) and an SD of 12.
- (a) Draw a smooth histogram for the data.
 - (b) If 200 students took the test, estimate how many scored between 70 and 80.
5. (20 pts.) (Hypothetical) A town near Worcester administers a spelling test to all students in fourth grade and all students in fifth grade. The test consists 100 words. For one group of fourth graders, the average for the number of words correctly spelled was 50 with a standard deviation of 10. When the same group took the test again in the fifth grade, the average for the number of words correctly spelled was 60 with a standard deviation of 15. The correlation coefficient for the data is $r = .7$. Suppose in the fourth grade a student was in the 70th percentile. What would you estimate for the student's percentile in the fifth grade? (Be sure to show your calculations and draw normal approximations as needed.)