

Math 115 Statistics Practice First Midterm

You can bring a *handwritten* 4x6 index card to the exam. You should bring a calculator, but you won't be allowed to use any of its statistical functions. I will give you a copy of Table A on the exam. Of course, your actual exam will be shorter than this one. At least one question on your exam will be taken directly off of your homework.

1. Explain in words what the mean, variance, and standard deviation measure about a dataset. Find the mean, variance, and standard deviation of the following dataset: 14, -3, -11, 20.

2. Draw a stemplot of the following data. Give the five-number summary. Without calculating the mean, explain how you can predict whether it will be greater than or less than the median.

0, 1, 1, 2, 3, 4, 4, 5, 7, 10, 40

3. The phone company offers two long-distance calling plans. Plan A charges \$3.50 a month plus \$0.05 (5 cents) per minute for all long-distance calls. Plan B has no monthly fee, but charges \$0.07 (7 cents) per minute for all long distance calls. The number of minutes Jane spends on long-distance calls each month follows a distribution that is approximately Normal, with a mean of 70 minutes and a standard deviation (SD) of 20 minutes. Let X represent Jane's long-distance minutes in a randomly chosen month.

a) Let Y_A and Y_B represent Jane's long distance charges in a randomly chosen month under plan A and under plan B, respectively. Compute the mean and standard deviation for Y_A and Y_B .

b) Give a range of values that has approximate probability .95 of containing the amount of Jane's long-distance bill under Plan A in a randomly chosen month.

c) If she chooses Plan B, in approximately what percentage of months will Jane's long-distance bill be between \$3.60 and \$5.00?

d) According to our approximation, in what percentage of months will Jane spend less than 0 minutes on long-distance calls? Does this make sense? How do we interpret this?

4. True or false?

a) The following is an observational study, not an experiment: Among a group of disabled women aged 65 and older who were tracked for several years, those who had a vitamin B₁₂ deficiency were twice as likely to suffer severe depression as those who did not. (*American Journal of Psychology* 15 (2000): 715)

b) If the correlation between two variables is close to 0, then there is only a weak association between them.

c) The area under a density curve is 1.

d) For any set of data, approximately 68% will be within 1 standard deviation of the mean.

e) The slope of the regression line is always between -1 and 1.

f) The median is resistant to outliers.

5. A small town in Vermont has only one store, the Glove Shack. Each month, they sell some number of pairs of mittens. Also each month, local residents lose some number of fingers to frostbite. The data for a few months are listed below.

mittens sold	0	2	15	11
fingers lost	0	0	4	4

a) Compute the correlation between pairs of mittens sold and fingers lost to frostbite. How do you interpret your answer?

b) Find the equation of the regression line.

c) Evaluate the following statement: “These mittens are terrible! People are MORE likely to get frostbite when they buy them. My hatred for these terrible mittens is as deep as the ocean.” What is a possible lurking variable?

6. A scatterplot of house prices (in thousands of dollars) versus house size (in thousands of square feet) shows a relationship that is straight, with only moderate scatter and no outliers. The correlation between house price Y and house size X is 0.85, and the equation of the regression line is $Y=9.564 + 122.74X$.

a) What does the slope of 122.74 mean?

b) How much can a homeowner expect the value of her house to increase if she builds on an additional 2000 square feet?

c) If a house is 1 SD above the mean size, how many SDs above the mean would you predict its price to be?

d) How much of the total variability in house prices can be explained by the variability in house sizes?

e) If we measured the size in thousands of square meters instead of thousands of square feet, how would the correlation r change? Would the slope of the regression line change?

f) What does the regression line predict will be the price of a house that is only 60 square feet? (That’s 60, not 60 thousand.) Do you trust this prediction? Why or why not?