1. (5 points) Which of the following is a true statement about correlation?

A. The correlation r is a measure of the effect of lurking variables.

B. Correlation implies causation.

C. The correlation r is a measure of the likelihood that a change in the explanatory variable causes a change in the response variable.

D. The correlation r is a measure of the predictive strength of a linear relationship between an explanatory and response variable.

2. (5 points) The use of a regression line to predict outside the range of your data set that you used to obtain the line is called

A. intrapolation

B. interpretation

C. mediation

D. extrapolation

E. interpolation

3. (5 points) If a distribution is skewed to the left,

A. the mean and median are equal.

B. the mean is less than the median.

C. the mean is greater than the median.

4. (5 points) Which of the following would be the most appropriate numerical summary of a symmetric distibution?

A. the mean $\bar{x}$ and the median

B. the five number summary

C. $\bar{x}$ and the standard deviation s

D. the median and quartiles

E. the mean $\bar{x}$

5. (20 points) Our brains don't like losses. Most people dislike losses more than they like gains. In money terms, people are about as sensitive to a loss of $10 as to a gain of $20. To discover what parts of the brain are active in decisions about gain and loss, psychologists presented subjects with a series of gambles with different odds and different amounts of winnings and losses. From a subject's choices, they constructed a measure of "behavioral loss aversion." Higher scores show greater sensitivity to losses. Observing brain activity while subjects made their decisions pointed to specific brain regions. Here are data for 16 subjects on behavioral loss aversion and neural loss aversion, a measure of activity in one region of the brain

Neural	−50.0	−39.1	−25.9	−26.7	−28.6	−19.8	−17.6	5.5
Behavioral	0.08	0.81	0.01	0.12	0.68	0.11	0.36	0.34
Neural	2.6	20.7	12.1	15.5	28.8	41.7	55.3	155.2
Behavior	0.53	0.68	0.99	1.04	0.66	0.86	1.29	1.94

(a) Make a scatterplot with neural loss aversion as x and behavioral loss aversion as y. One point is a high outlier in both the x and y directions.

(b) Find the least-squares line for predicting y from x, leaving out the outlier, and add the line to your plot.

(c) The outlier lies very close to your regression line. Looking at the plot, you now expect that adding the outlier will increase the correlation but will have little effect on the least-squares line. Explain why.

(d) Find the correlation and the equation of the least-squares line with and without the outlier. Do your results verify your expectation?

6. (20 points) Your body mass index (BMI) is your weight in kilograms divided by the square of your height in meters. Many online BMI calculators allow you to enter weight in pounds and height in inches. High BMI is a common but controversial indicator of overweight or obesity. A study by the National Center for Health Statistics found that the BMI of American young women (ages 20 to 29) is approximately Normal with mean 26.5 and standard deviation 6.4.

(a) People with BMI less than 18.5 are often classified as underweight. What percent of young women are underweight by this criterion?

(b) People with BMI over 30 are often classified as obese. What percent of young women are obese by this criterion?

7. (20 points) Births are not, as you might think, evenly distributed across the days of the week. Here are the average numbers of babies born on each day of the week in 2008.

Day	Births
Sunday	7,534
Monday	12,371
Tuesday	13,415
Wednesday	13,171
Thursday	13,147
Friday	12,919
Saturday	8,617

(a) Present these data in a well-labeled bar graph.

(b) Would it also be correct to make a pie chart?

(c) Suggest some possible reasons why there are fewer births on weekends.

8. (20 points) According to the Census Bureau's 2010 Current Population Survey, the mean and median 2009 income of people at least 25 years old who had a bachelor's degree but no higher degree were $46,931 and $58,762.

Which of these numbers is the mean and which is the median? Explain your reasoning.