You have 70 minutes. That's 14 minutes per page. (There are 5 pages of questions.) You may use a calculator (scientific or graphing) and a table of z-scores.

The questions are not in any particular order. Don't get stuck too long on one question; you may find later questions easier. In multi-part questions, the later parts often do not depend on the earlier parts; attempt them all.

Show your work and explain your answers. In fact, I will largely be grading your explanations, rather than the numbers that show up in your explanations. If you cannot answer a question, then show as much work as you can; good work often earns partial credit.

Good luck.

1. Adults in a mall were asked about their educations and whether they smoked. The table below shows the data collected. Is there an association between smoking and educational attainment?

	Smoker	Non-smoker	Total
High school	32	61	93
Two-year college	5	17	22
Four-year college	13	72	85
Total	50	150	200

2. When this exam is returned to you after grading, it will bear a numerical score between 0% and 100%. It will not bear a letter grade. However, so that you have some idea of how you did relative to your classmates, I will report some summary statistics about the distribution of scores. What statistics should I report?

3. Here's a scatterplot of total mortgages in the U.S. (in millions of 2005 dollars) vs. the interest rate, in various years over the past 26 years. The correlation is -0.84.
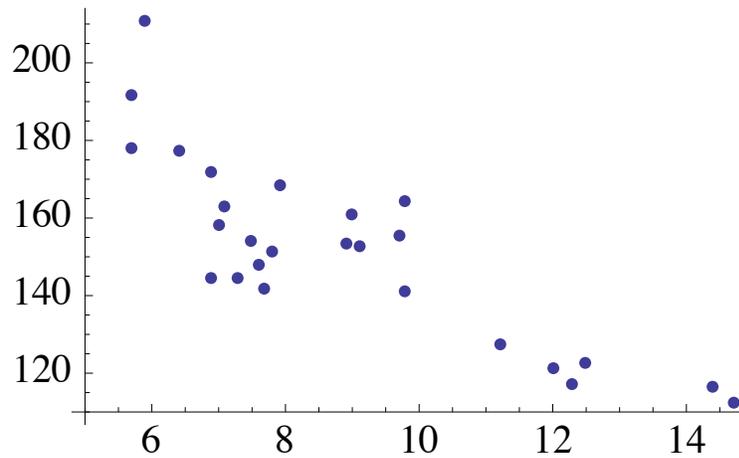
A. Describe the relationship between total mortgages and interest rate.

B. If both variables were standardized, what would the correlation between them be?

C. Suppose that, in another year, the interest rate was 11% and total mortgages were $250 million. How would including that year with these data change the correlation?

D. Do these data prove that lowering interest rates will lead to more people taking out mortgages?

4. A city has 682 total police: 522 police officers, 55 detectives, and 79 higher-ranking personnel. In an effort to keep informed about morale, the chief of police institutes job satisfaction interviews with a sample of police every month.

A. Do you see any danger of bias in this plan?

B. Propose a sampling strategy.

C. Propose one good question about job satisfaction. Explain why your question is good.

5. A lumber company intends to purchase logging rights in one of two managed forests. The two forests have roughly the same number of trees. In the first forest, the tree heights (in meters) are normally distributed according to $N(17.8, 3.4)$. In the second forest, they are distributed according to $N(16.1, 0.9)$. The company wants to cut down all trees that are 14 meters or taller; it will leave shorter trees untouched, because they are unprofitable.

A. Which forest has more trees that are desirable to the lumber company?

B. In that forest, how tall must a tree be, to be taller than 95% of the other trees in the forest?

C. As precisely as you can, draw the distribution of trees in that forest, after the lumber company has cut down the trees that it wants. Also describe the distribution.

6. The first plot below shows data about the prices of computers (in 2010 dollars) vs. speed of computers (measured on a 1-10 scale). The second plot shows the logarithm of price against the logarithm of speed, with the best-fit line superimposed. The best-fit line has intercept 1.69028 and slope 1.77665. The R^2 value is 0.961079. The third plot shows the residuals of the linear model vs. the logarithm of speed.
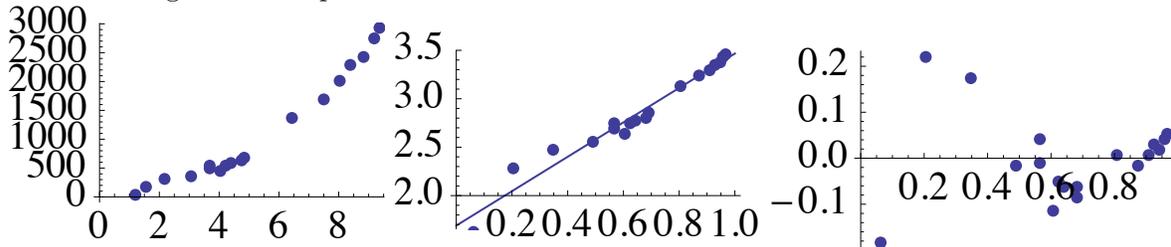
A. What is the equation that gives the predicted price in terms of the speed? Simplify.

B. What is the predicted price for a computer with speed rating 8.0?

C. Are any of the data influential?

D. Is the line an appropriate model for the relationship between log-price and log-speed?