Exam-like practice questions

1. Consider the following probability space and random variable.

S	s_1	s_2	s_3
p	0.2	0.3	0.5
X	100	110	90
Y	110	90	100

   Compute the expectation and standard deviation of X as well as the coefficient of correlation between X and Y.

2. (Not really an exam question, but a check of your understanding.) If I draw many random samples of size 1,000 from X above, how would the mean of the resulting samples be distributed, roughly speaking?

3. You are planning to start selling a service to a market of $100,000$ potential customers. That will be profitable provided over 5% of the potential market signs up. Offering the service to a random test sample of $1,000$ of these customers leads to 55 sign-ups. Using a one-sided test and these data, can you reject the hypothesis that the sign-up rate in this market will be 5% or less with 95% confidence?

4. Based on the same one-sided test as above, with what confidence can you reject the hypothesis that the sign-up rate will be 5% or less?

5. After downloading monthly adjusted price data for MSFT and the S&P500 over the past 5 years, regress the monthly return on MSFT on the return on the S&P500. Use that regression output to test the hypothesis that there is no relationship between MSFT returns and S&$P500$ returns using a 95% confidence level.

6. Using dataset 1 on my webpage and a two-sided test, test the hypothesis that the average age of male customers is the same as the average age of female customers.

7. Using that same data set, regress $ln(income)$ on age, age^2, and $gender$. Which of these variables are significant at the 5% level (= different from zero with 95% confidence) in this regression?

8. Use that model to forecast the income of a new female customer of age 30.

9. Consider the following data for spending for a representative sample of customers in a particular market during a recent period.

Spending	*Female*	*Male*
$< 50K$	700	601
$[50K - 100K)$	513	557
$[100K - 200K)$	410	518
$\geq 200K$	227	309

   Use these data and a chi-squared test to test the hypothesis that spending is independent of gender in this market.

10. With how much confidence can you reject the hypothesis that spending is independent of gender based on the table above?

11. Consider the following probability space and random variable.

S	s_1	s_2	s_3
p	0.2	0.3	0.5
X	100	110	90

   You are given 200 draws supposedly from the X shown in problem 1 above where the number of 100s is 26, the number of 110s is 68 and the remaining 106 draws are 90s. Based on a Chisquare goodness of fit test, can you reject the hypothesis that these data came from X with 95% confidence?