Midterm Examination II

Econ 103, Statistics for Economists

March 25th, 2014

You will have 70 minutes to complete this exam. Graphing calculators, notes, and textbooks are not permitted.

I pledge that, in taking and preparing for this exam, I have abided by the University of Pennsylvania's Code of Academic Integrity. I am aware that any violations of the code will result in a failing grade for this course.

Name: ______________________________

Student ID #: ______________________________

Signature: ______________________________

Question:	1	2	3	4	5	6	7	Total
Points:	25	20	15	30	20	15	15	140
Score:								

Instructions: Answer all questions in the space provided, continuing on the back of the page if you run out of space. Show your work for full credit but be aware that writing down irrelevant information will not gain you points. Be sure to sign the academic integrity statement above and to write your name and student ID number on *each page* in the space provided. Make sure that you have all pages of the exam before starting.

Warning: If you continue writing after we call time, even if this is only to fill in your name, twenty-five points will be deducted from your final score. In addition, two points will be deducted for each page on which you do not write your name and student ID.

1. Let $X_1 \sim \text{Bernoulli}(1/2)$ independently of $X_2 \sim \text{Bernoulli}(2/3)$ and define $Y = 2X_1$ and $Z = X_2 - X_1$.

   (a) (15 points) Express the joint pmf of Y and Z in tabular form. Please put the realizations of Y in the *rows* of the table and realizations of Z in the *columns*.

   (b) (3 points) What is the marginal pmf of Y?

   (c) (3 points) What is the marginal pmf of Z?

(d) (4 points) Calculate the covariance between Y and Z. Are Y and Z independent?

2. Let X be a continuous random variable with support $[1, c]$ and pdf $f(x) = 1/x$.

   (a) (5 points) What is the value of the constant c?

   (b) (5 points) Calculate the CDF of X.

   (c) (5 points) Calculate $E[X]$.

   (d) (5 points) Calculate $E[X^2]$.

3. Suppose that $Y \sim N(\mu = 2, \sigma^2 = 3)$ independently of $X \sim N(\mu = -1, \sigma^2 = 6)$.
   (a) (5 points) Let $Q = X - Y$. What kind of random variable is Q? Be sure to specify its parameters.

   (b) (10 points) Approximately what is the probability that $X > Y$? Explain.

4. Garth wants to learn how much taller NBA players are than Penn Undergraduates, on average. To answer this question, he's recruited volunteers to make up two independent random samples. The first sample contains 10 NBA players: $X_1, \ldots, X_{10} \sim$ iid with mean μ_X and variance σ^2. The second sample is independent of the first and contains 15 Penn Undergrads: $Y_1, \ldots, Y_{15} \sim$ iid with mean μ_Y and variance σ^2. Just to be completely clear, in this question we are assuming that the variance is identical for Penn Students and NBA players to make the calculations simpler.
   (a) (4 points) To answer his question, Garth needs to estimate $\mu_X - \mu_Y$. There is an obvious unbiased estimator of this quantity. What is it? Prove that it is unbiased.

   (b) (4 points) Calculate the variance of the estimator you proposed in part (a).

(c) (8 points) When measuring the players and students Garth makes a mistake: although he accurately records each of the 25 heights, he forgets to note which correspond to Penn students and which correspond to NBA players. Fortunately Garth remembers that, among the first 10 heights on his list, there were exactly 5 students and 5 NBA players. In other words, the first 10 heights on his list are $X_1, \ldots, X_5$ and $Y_1, \ldots, Y_5$ *in some unknown order* and the last 15 are $X_6, \ldots, X_{10}$ and $Y_6, \ldots, Y_{15}$ *in some unknown order*. Let $\bar{Z}_1$ be the sample mean of the first 10 heights on Garth's list and $\bar{Z}_2$ be the sample mean of the last 15. Prove that

$$E[\bar{Z}_1] = \frac{\mu_X + \mu_Y}{2} \qquad \text{and} \qquad E[\bar{Z}_2] = \frac{1}{3}\mu_X + \frac{2}{3}\mu_Y$$

For full credit, provide a careful explanation and cite any results you have used.

(d) (4 points) Let α and β be two arbitrary constants. Using part (c), calculate $E[\alpha\bar{Z}_1 + \beta\bar{Z}_2]$. Simplify your answer so that it takes the form $c_1\mu_X + c_2\mu_Y$ where c_1 and c_2 are two constants that will depend on α and β.

(e) (6 points) Back to Garth's problem: *amazingly* it's still possible to construct an unbiased estimator of $\mu_X - \mu_Y$ *without knowning* which observations corresponded to NBA players and which corresponded to Penn students. The trick is to take a particular linear combination of $\bar{Z}_1$ and $\bar{Z}_2$. Use your answer to the previous part to find the values of α and β that give $E[\alpha \bar{Z}_1 + \beta \bar{Z}_2] = \mu_X - \mu_Y$.

(f) (4 points) Although it's unbiased, there's a clear downside to the estimator from the previous part: it has a very high variance. Calculate its variance and compare it to that of the estimator from part (a). Explain the intuition behind the difference.

5. Suppose $X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2. Let $\bar{X}_n$ be the sample mean and define $\widehat{\mu}_n = n\bar{X}_n/(n+1)$.
   (a) (15 points) Calculate the MSE of $\widehat{\mu}_n$ as an estimator of μ.

   (b) (5 points) Is $\widehat{\mu}_n$ a consistent estimator of μ? Briefly explain your answer.

6. This question concerns the standard normal RV and related R functions.
   (a) (5 points) Write R code to plot the standard normal pdf between -3 and 3 using a step size of 1/100. Be sure to plot a *smooth curve* rather than isolated points.

(b) (5 points) Write R code to approximate the probability that a standard normal random variable takes on a value greater than 0.5 using a Monte Carlo experiment with 10,000 draws.

(c) (5 points) There is an R command that gives the *exact* answer to the preceding part without using simulation. What is it?

7. This question concerns the mean of a $\chi^2(5)$ random variable.

   (a) (10 points) Write R code to approximate the mean of a $\chi^2(5)$ random variable by making 10,000 simulation draws. You may use any R functions you like in your answer *except* `rchisq`.

   (b) (5 points) Use the facts we studied in class to *derive* the mean of a $\chi^2(5)$ random variable. Hint: recall how we derived the mean of the Binomial RV in class.

Name: ____________________ Student ID #: ____________________