# Second Midterm Examination

Econ 103, Statistics for Economists

March 24th, 2015

**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: ______________________________

Signature: ______________________________

Student ID #: ______________ Recitation #: ______________

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| Points: | 30 | 20 | 20 | 25 | 30 | 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. For each of the following parts, simply write down the answer: no explanation is needed.

   3 (a) If $X_1, \ldots, X_n \sim$ iid with mean $\mu$ and variance $\sigma^2$, what is $Var(\bar{X}_n)$?

   3 (b) Suppose $X_1, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$. Write down an expression for the approximate 95% CI for $\mu$ if $\sigma$ is known.

   3 (c) What is the median of a normal RV with mean -4 and variance 36?

   3 (d) Write down the pmf of a Binomial$(10, 1/3)$ RV.

   3 (e) Suppose that $X$ is a continuous RV that is equally likely to take on any value in $[-2, 2]$ but never takes on a value outside this range. Write down its pdf.

   3 (f) Write down a single line of R code to make 50 random draws from a normal distribution with mean -1 and standard deviation 10.

   3 (g) Suppose $Z \sim N(0, 1)$. Write a single line of R code to calculate the value of $c$ such that $P(-c \leq Z \leq c) = 0.8$.

Name: ______________________ Student ID #: ______________________

$\boxed{3}$ (h) What is the support set of a Binomial$(5, 1/2)$ RV?

$\boxed{3}$ (i) Write a single line of R code to calculate the 70th percentile of a normal RV with mean -1 and variance 4.

$\boxed{3}$ (j) Calculate the approximate value of $P(Z > 3)$ if $Z \sim N(\mu = 1, \sigma^2 = 4)$.

2. Suppose $X_1, X_2, X_3 \sim$ iid $N(0, 1)$. For each of the following parts, simply write down the answer: no explanation is needed.

$\boxed{4}$ (a) What kind of random variable is $X_1 + X_2 + X_3$? What is its support and what are the values of its parameters?

$\boxed{4}$ (b) What kind of random variable is $X_1^2 + X_2^2 + X_3^2$? What is its support and what are the values of its parameters?

$\boxed{4}$ (c) Write a single line of R code to calculate the median of $X_1^2 + X_2^2 + X_3^2$.

$\boxed{4}$ (d) What kind of random variable is $X_1/\sqrt{(X_2^2 + X_3^2)/2}$? What is its support and what are the values of its parameters?

$\boxed{4}$ (e) Write a single line of R code to calculate $P\left[X_1/\sqrt{\left(X_2^2+X_3^2\right)/2}>2\right]$.

3. Suppose $X \sim \text{Uniform}(0,1)$ and let $A$ be the *area* of a circle with radius equal to $X$. Recall that the area of a circle with radius $r$ is $\pi r^2$.

$\boxed{5}$ (a) Calculate $E[X]$.

$\boxed{5}$ (b) Calculate $E[A]$.

$\boxed{5}$ (c) Suppose I constructed a circle with radius equal to $E[X]$. Would its area equal $E[A]$? Why or why not? Explain briefly.

$\boxed{5}$ (d) Calculate $Var(A)$. You may leave your answer in terms of $\pi$.

4. Let $X$ and $Y$ be two random variables with covariance $\sigma_{XY}$ and variances $\sigma_X^2$ and $\sigma_Y^2$ where $E[X] = E[Y] = 0$. Define $\beta = \sigma_{XY}/\sigma_X^2$ and $Z = Y - \beta X$.

$\boxed{4}$ (a) Is $\beta$ a random variable or a constant? Explain briefly.

$\boxed{4}$ (b) Is $Z$ a random variable or a constant? Explain briefly.

$\boxed{5}$ (c) Calculate $E[Z]$.

12 (d) Calculate $Cov(X, Z)$.

5. Mallesh wants to know the proportion $p$ of Penn undergraduates who smoke cigarettes, so he interviews a random sample of $n$ students. Let $X_i$ be a random variable that indicates whether the $i^{th}$ person in the sample is *truly* a smoker: $0 =$ No, $1 =$ Yes. Because of social stigma, conditional on being a smoker, person $i$ will *lie* and claim to be a non-smoker with probability $q$ when interviewed. Conditional on being a non-smoker, person $i$ always tells the truth when interviewed. Let $Y_i$ be a random variable that indicates whether the $i^{th}$ person in the sample *claims* to be a smoker: $0 =$ No, $1 =$ Yes.

12 (a) For a single individual $i$, write the joint distribution of $X_i$ and $Y_i$ in tabular form, arranging $X_i$ in the *rows* and $Y_i$ in the *columns* of your table. Your answer should be expressed in terms of $p$ and $q$.

$\boxed{4}$ (b) Using your answer to the preceding, calculate the marginal pmf of $Y_i$.

$\boxed{6}$ (c) Using your answers to the preceding two parts, calculate $Cov(X_i, Y_i)$.

$\boxed{8}$ (d) Since some students may be lying to him, Mallesh does *not* observe $X_1, \ldots, X_n$. He only observes $Y_1, \ldots, Y_n$. If Mallesh uses the sample mean of $Y_1, \ldots, Y_n$ to estimate $p$, will this procedure be unbiased? If so, prove it. If not, will this procedure produce an overestimate or an underestimate? Why?

6. Suppose we have a dataframe called `StudentData` with two columns: `height` and `handspan`. Each row in the dataset corresponds to a student who took Econ 103 at Penn over the past three years: `height` gives her height in inches while `handspan` gives her handspan in centimeters. In the spirit of the sampling experiment I discussed in class, which you replicated on Problem Set #7, suppose we want to examine the sampling distribution of the *sample covariance* between height and handspan using a Monte Carlo simulation in which we treat `StudentData` as a *population* from which we repeatedly draw random samples.

$\boxed{10}$ (a) Write an R function called `cov.sim` that takes a random sample of `n` *rows* of `StudentData` and returns the sample correlation between `height` and `handspan` based on the sampled rows. Your function should take a single input argument: `n`. You may assume that there are no missing values.

$\boxed{5}$ (b) Using your function from the previous part, write R code to plot a histogram of the sampling distribution of the sample covariance with a sample size of 20, based on 1000 simulation replications.

Name: ______________________ Student ID #: ______________________