Second Midterm Examination
Econ 103, Statistics for Economists

March 22nd, 2016

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	Total
Points:	25	20	30	35	30	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 and Z be discrete RVs with joint pmf

		Z		
		-1	0	1
X	0	1/6	1/3	0
	1	0	1/6	1/3

5   (a) What kind of RV is X? Specify any and all parameters of its distribution.

5   (b) Calculate $E[X]$.

5   (c) Calculate the conditional pmf of Z given that $X = 0$.

10   (d) Calculate $E[XZ]$.

Name: _____   Student ID #: _____

2. Let $X_1, \ldots X_5 \sim$ iid with pdf $f(x) = (x+1)/2$ and support set $x \in [-1, 1]$.

  5   (a) Calculate the CDF $F(x_0)$ of X_1.

  5   (b) Calculate $P(X_1 > 0)$.

  10   (c) What is the probability that *exactly* four of $X_1, \ldots, X_5$ are greater than zero?

Name: _____   Student ID #: _____

This question is based on 7-17 from the textbook, which I assigned as homework. I have re-worded the problem for clarity but the substance and solution are unchanged.

3. A machine produces perfectly square metal plates. Occasionally, however, this machine malfunctions and the squares it produces are of the wrong area. To make sure the plates are the correct size, we measure the length of each one *twice*. Our measuring device is slightly imperfect: if the true length of a particular square is ℓ inches, our measurement will equal $\ell + Z$ where Z is a random error with mean zero and standard deviation 0.1 inches. Fortunately these errors are independent across measurements: when we measure a square twice our first measurement will be $X_1 = \ell + Z_1$ and our second will be $X_2 = \ell + Z_2$ where $Z_1, Z_2 \sim$ iid ($\mu = 0, \sigma = 0.1$). We want to find a way of combining X_1 and X_2 to estimate the *area* ℓ^2 of a given plate. Three estimators have been proposed: $A, B,$ and C. Calculate the bias of each using the shortcut formula.

10   (a) $A = \dfrac{X_1^2 + X_2^2}{2}$

10   (b) $B = X_1 X_2$

10   (c) $C = \left(\dfrac{X_1 + X_2}{2} \right)^2$

Name: _____   Student ID #: _____

4. This question uses simulations to study what happens when we sample *without* replacement from a population with 100 individuals: 40 of whom are "ones" and the rest of whom are "zeros." The point is to explore the bias and variance of this procedure and compare it to the random sampling procedure (*with* replacement) we studied in class.

|5| (a) Let `popn` be an R vector with 100 elements, the first 40 of which are `1` and the rest of which are `0`. Write an R command that makes 50 random draws from `popn` *without replacement.*

|15| (b) Build your command from the preceding part into a function called `draw.mean` that draws a sample of size `n` from `popn` and returns the sample mean of the draws. Your function should take a single input argument: `n` the number of draws to make.

|5| (c) Write R code that calls the function `draw.mean` 10000 times with `n` equal to 50 and stores the result in a vector called `sims`.

Name: _____   Student ID #: _____

5   (d) After running the code described in the preceding parts I entered `mean(sims)` at the R console and got 0.399912. Based on this result, does it appear that the sample mean remains an unbiased estimator of the population mean when calculated from a sample drawn *without replacement*? Explain briefly.

5   (e) I then entered `sd(sims)` into the R console and got 0.04946187. Based on this result, how does does sampling *without replacement* compare to sampling *with replacement* in terms of the standard error of the sample mean? Explain briefly

5. Let $X_1, \ldots X_{25} \sim$ iid $N(\mu, \sigma^2 = 4)$ and define $Z = 5(\bar{X}_n - \mu)/S$ and $Y = 6S^2$. When asked for a sampling distribution in this question, be sure to name the distribution and specify the values of any and all relevant parameters.

10   (a) Suppose our observed dataset $x_1, \ldots, x_{25}$ has a sample mean of $\bar{x} = 2.3$. Construct a 68% confidence interval for μ using the fact that the population variance is known to be 4. Your answer should *not* include any R commands.

Name: _____   Student ID #: _____

5 (b) What is the sampling distribution of Z? No explanation is needed.

5 (c) Write down the R command to calculate $P(Z \leq 1)$.

5 (d) What is the sampling distribution of Y? No explanation is needed.

5 (e) Write down the R command to calculate 90th percentile of Y.

Name: _____   Student ID #: _____