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.
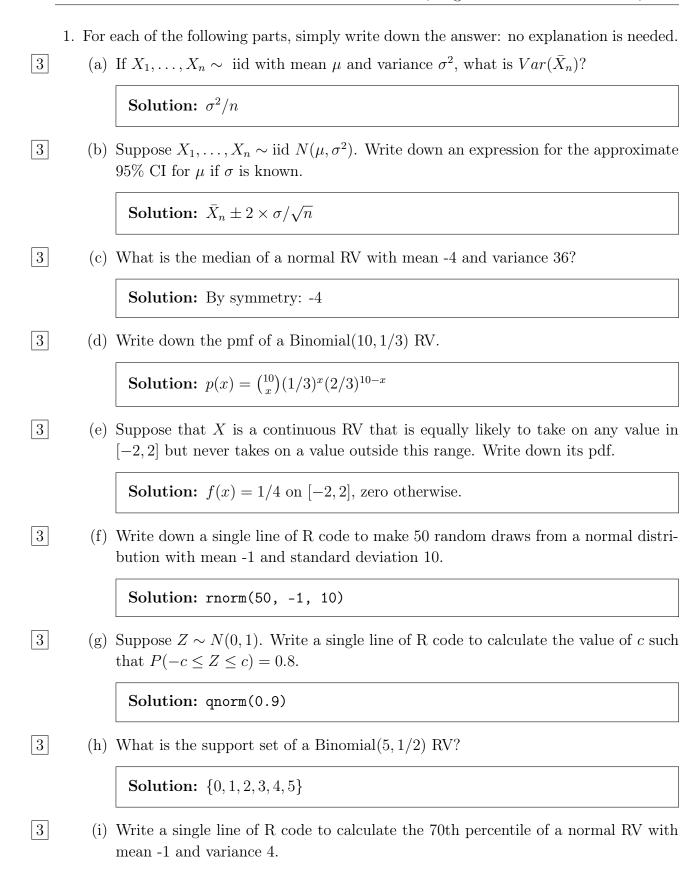
Name: _____

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 μ and variance σ^2, what is $Var(\bar{X}_n)$?

> **Solution:** σ^2/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 μ if σ is known.

> **Solution:** $\bar{X}_n \pm 2 \times \sigma/\sqrt{n}$

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

> **Solution:** By symmetry: -4

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

> **Solution:** $p(x) = \binom{10}{x}(1/3)^x(2/3)^{10-x}$

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.

> **Solution:** $f(x) = 1/4$ on $[-2, 2]$, zero otherwise.

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.

> **Solution:** `rnorm(50, -1, 10)`

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

> **Solution:** `qnorm(0.9)`

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

> **Solution:** $\{0, 1, 2, 3, 4, 5\}$

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

Name: _____   Student ID #: _____

> **Solution:** `qnorm(0.7, mean = -1, sd = 2)`

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

> **Solution:** $P(Z > 3) = P([Z-1]/2 > 1) \approx 0.16$

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.

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?

> **Solution:** Normal with mean zero and variance 3. Its support is $(-\infty, \infty)$.

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?

> **Solution:** χ^2 with 3 degrees of freedom. Its support is $[0, \infty)$ or equivalently $(0, \infty)$ since $P(X_1^2 + X_2^2 + X_3^2 = 0) = 0$.

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

> **Solution:** `qchisq(0.5, df = 3)`

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?

> **Solution:** t with 2 degrees of freedom. Its support is $(-\infty, \infty)$.

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

> **Solution:** `1 - pt(2, df = 2)`

3. Suppose $X \sim$ 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 πr^2.

5　　　(a) Calculate $E[X]$.

> **Solution:** $E[X] = \int_0^1 x \, dx = (x^2/2)|_0^1 = 1/2$

Name: _____　　　　　　Student ID #: _____

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

> **Solution:** $E[A] = E[\pi X^2] = \pi E[X^2] = \pi \int_0^1 x^2 \, dx = \pi/3$

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.

> **Solution:** No: in general the expected value of a function is not equal to the function of the expected value. Here, $E[\pi X^2] \neq \pi E[X]^2$ – the first of these equals $\pi/3$ while the second equals $\pi/4$.

5   (d) Calculate $Var(A)$. You may leave your answer in terms of π.

> **Solution:**
>
> $$
> \begin{aligned}
> Var(A) &= E[A^2] - E[A]^2 = E\left[\pi^2 X^4\right] - (\pi/3)^2 \\
> &= \pi^2 \int_0^1 x^4 \, dx - \pi^2/9 \\
> &= \pi^2 \cdot \left.\frac{x^5}{5}\right|_0^1 - \pi^2/9 = \pi^2 \left(1/5 - 1/9\right) = 4\pi^2/45
> \end{aligned}
> $$

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

4   (a) Is β a random variable or a constant? Explain briefly.

> **Solution:** A constant: it's a function parameters, which are constants.

4   (b) Is Z a random variable or a constant? Explain briefly.

> **Solution:** A random variable: it's a function of the RVs Y and X.

5   (c) Calculate $E[Z]$.

> **Solution:** $E[Z] = E[Y - \beta X] = E[Y] - \beta E[X] = 0$

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

Name: _____   Student ID #: _____

> **Solution:** Since both X and Z have zero mean,
>
> $$\begin{aligned} Cov(X,Z) &= E[XZ] = E\left[X\left(Y - \beta X\right)\right] \\ &= E[XY] - \beta E[X^2] = Cov(X,Y) - \beta Var(X) \\ &= \sigma_{XY} - \frac{\sigma_{XY}}{\sigma_X^2} \cdot \sigma_X^2 = 0 \end{aligned}$$
>
> using the fact that $E[X] = E[Y] = 0$ and the definition of β.

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 ith 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 ith 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.

> **Solution:** We construct the table of joint probabilities using the conditional pmf of Y_i given X_i. We have:
>
> $$\begin{aligned} P(X_i = 0, Y_i = 0) &= P(Y_i = 0 | X_i = 0)P(X_i = 0) = 1 - p \\ P(X_i = 0, Y_i = 1) &= P(Y_i = 1 | X_i = 0)P(X_i = 0) = 0 \\ P(X_i = 1, Y_i = 0) &= P(Y_i = 0 | X_i = 1)P(X_i = 1) = qp \\ P(X_i = 1, Y_i = 1) &= P(Y_i = 1 | X_i = 1)P(X_i = 1) = (1-q)p \end{aligned}$$
>
> | | | Y_i | |
> |-------|-----|---------|----------|
> | | | 0 | 1 |
> | X_i | 0 | $1 - p$ | 0 |
> | | 1 | qp | $(1-q)p$ |

|4| (b) Using your answer to the preceding, calculate the marginal pmf of Y_i.

> **Solution:** Summing over the entries in each *column* of the table, we have, $P(Y_i = 0) = [1 - (p - qp)]$ and $P(Y_i = 1) = (p - qp)$.

|6| (c) Using your answers to the preceding two parts, calculate $Cov(X_i, Y_i)$.

Name: _____   Student ID #: _____

> **Solution:** We proceed using the shortcut formula. First, since $E[X_i] = p$ and $E[Y_i] = (p - qp)$, $E[X_i]E[Y_i] = p^2(1 - q)$. To calculate $E[XY]$, we only need to look at the lower right entry in the joint pmf table: all other terms in the sum are zero. Thus: $Cov(X_i, Y_i) = p(1 - q) - p^2(1 - q) = p(1 - p)(1 - q)$.

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?

> **Solution:** $E\left[(Y_1 + \cdots + Y_n)/n\right] = E[Y_i] = p - qp$. Since the expected value of this estimator is not p, Mallesh's procedure is biased. In particular, the bias is: $(p - qp) - p = -qp$. Unless $q = 0$, Mallesh will tend to *underestimate* the true fraction of smokers. This is because some smokers lie and say that they are non-smokers, but no non-smokers lie and say they are smokers.

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.

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.

> **Solution:**
>
> ```
> cov.sim <- function(n){
>   sim.rows <- sample(1:nrow(StudentData), n, replace = TRUE)
>   sim.data <- StudentData[sim.rows,]
>   return(cov(sim.data$height, sim.data$handspan))
> }
> ```

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 #: _____

Solution:

```
sims <- replicate(1000, cov.sim(20))
hist(sims)
```