Second Midterm Examination

Econ 103, Statistics for Economists

November 2nd, 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	7	Total
Points:	25	20	20	20	15	20	20	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, write down R code to carry out the requested operation.

3 (a) Make 100 iid draws from a standard normal distribution.

Solution: `rnorm(100)`

3 (b) Calculate c such that $P(-c \leq Z \leq c) = 0.80$ if $Z \sim N(0, 1)$.

Solution: `qnorm(0.9)`

3 (c) Calculate the 75th percentile of a $t(5)$ RV.

Solution: `qt(0.75, df = 5)`

3 (d) Plot $(x^3 - x)$ on $[-2, 2]$ with a step size of 0.01.

Solution:

```
x <- seq(from = -2, to = 2, by = 0.01)
plot(x, x^3 - x)
```

3 (e) Find the probability that a Binomial(10, 0.4) RV takes on a value in $\{3, 4, \ldots, 8\}$.

Solution: `sum(dbinom(3:8, 10, 0.4))`

10 (f) Write an R function `exact95CI` to calculate an *exact* 95% confidence interval for the mean of a normal population when the population variance is unknown. The function should take one input, `x` a vector of data, and return a vector with two elements: the lower and upper confidence limits of the interval.

Solution:

```
exact95CI <- function(x){
  n <- length(x)
  SE <- sd(x) / sqrt(n)
  ME <- qt(1 - 0.05/2, df = n-1) * SE
  LCL <- mean(x) + ME
  UCL <- mean(x) - ME
  return(c(LCL, UCL))
}
```

2. Let $X, Y, Z \sim$ iid $N(0, 1)$. No explanation is needed for any of the following.

Name: ______________________ Student ID #: ______________________

4 (a) What kind of RV is $3X + 2$? Specify all parameters of its distribution.

> **Solution:** $N(2, \sigma^2 = 9)$

4 (b) What kind of RV is $X^2 + Y^2 + Z^2$? Specify all parameters of its distribution.

> **Solution:** $\chi^2(3)$

4 (c) What kind of RV is $(X + Y + Z)/3$? Specify all parameters of its distribution.

> **Solution:** $N(0, \sigma^2 = 1/3)$

4 (d) What kind of RV is X^2/Y^2? Specify all parameters of its distribution.

> **Solution:** $F(1, 1)$

4 (e) Calculate $P(X < Y)$.

> **Solution:** $P(X < Y) = P(X - Y < 0) = P\left[N(0, \sigma^2 = 2) < 0\right] = 1/2$

3. Let X be a Binomial$(2, 1/3)$ random variable (RV). Define the RV Y as follows. Given that $X = 0$, Y can only take on the values $-1, 0$ or 1, each with probability $1/3$. Given that $X = 1$, Y can only take on the values $0, 1$ or 2 each with probability $1/3$. Finally given that $X = 2$, Y can only take on the values $1, 2$ or 3, each with probability $1/3$.

4 (a) Write down the marginal pmf of X. Be sure to specify the support set.

> **Solution:** $p_X(x) = \binom{2}{x} (1/3)^x (2/3)^{2-x}$ with support set $\{0, 1, 2\}$

12 (b) Write out the joint pmf of X, Y in a table. To make this problem easier to grade, please put the values of X in the *rows* of your table.

> **Solution:** We are given the marginal distribution of X and the conditional distribution of Y given X. By the multiplication rule, $p_{XY}(x, y) = p_{Y|X}(y|x)p_X(x)$. From the problem statement, we know that the only possible values Y can take on given that $X = x$ are $\{x - 1, x, x + 1\}$ and that these each occur with probability $1/3$. Using the result of the preceding part, we calculate the marginal probabilities for X as follows:
>
> $$p_X(0) = 4/9, \quad p_X(1) = 4/9, \quad p_X(2) = 1/9$$

Thus, we find that:

		Y				
		-1	0	1	2	3
	0	4/27	4/27	4/27	0	0
X	1	0	4/27	4/27	4/27	0
	2	0	0	1/27	1/27	1/27

4 (c) Calculate the marginal pmf of Y. Make sure to specify its support set.

Solution: The support is $\{-1, 0, 1, 2, 3\}$ and $p_Y(-1) = 4/27$, $p_Y(0) = 8/27$, $p_Y(1) = 9/27$, $p_Y(2) = 5/27$, and $p_Y(3) = 1/27$.

4. Let X be a continuous random variable with support set $[0, \infty)$ and pdf $f(x) = e^{-x}$. This is a special case of the so-called "Exponential" random variable, an example we did not cover in class. This question asks you to apply what you know about continuous random variables in general to this new example.

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

   Solution: $\int_{-\infty}^{x_0} f(x)\, dx = \int_0^{x_0} e^{-x}\, dx = -e^{-x}\big|_0^{x_0} = -e^{-x^0} - (-e^0) = 1 - e^{-x_0}$

   5 (b) Calculate $P(0 \leq X \leq 1)$.

   Solution: $P(0 \leq X \leq 1) = F(1) - F(0) = F(1) = 1 - 1/e \approx 0.63$

   5 (c) Derive the quantile function $Q(p)$ of X.

   Solution: Since $Q(p) = F^{-1}(x)$, we simply solve $p = F(x) = 1 - e^{-x}$ for x, yielding: $Q(p) = -\log(1-p)$ where log denotes the natural logarithm.

   5 (d) The "Survival Function" of a continuous RV Y is defined as $S(t) = P(Y > t)$ where t is some threshold. Calculate the survival function of X.

   Solution: By the complement rule:

   $$S(t) = P(X > t) = 1 - P(X \leq t) = 1 - F(t) = e^{-t}$$

5. The "Moment Generating Function" of a random variable X is defined as $M_X(t) = E[e^{tX}]$ where t is a constant. Don't let the constant t confuse you: this is simply the expected value of a function g of X, namely $g(x) = e^{tx}$.

   3 (a) Suppose that X is a discrete RV. Use the general formula for $E[g(X)]$ from class to express $M_X(t)$ as a sum involving the pmf $p(x)$ of X.

   Solution: $M_X(t) = \sum_{\text{all } x} e^{tx} p(x)$

   3 (b) Suppose that X is a continuous RV. Use the general formula for $E[g(X)]$ from class to express $M_X(t)$ as an integral involving the pdf $f(x)$ of X.

   Solution: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$

Name: ____________________ Student ID #: ____________________

4 (c) Use your answer to part (a) to calculate the moment generating function of the Bernoulli(p) random variable. *Hint:* we worked out a special case of this calculation in class, namely $t = 1$.

Solution: $M_X(t) = pe^{t\times 1} + (1-p)e^{t\times 0} = 1 - p + pe^t$

5 (d) Use your answer to part (b) to calculate the moment generating function of the Uniform$(0,1)$ random variable.

Solution: $M_X(t) = \int_0^1 e^{tx} \times 1\, dx = \frac{1}{t}\, e^{tx}\Big|_0^1 = \frac{1}{t}(e^t - 1)$

6. Dr. Evil teaches Reverse Psychology at Minion State University. Grades in his course are based solely on a midterm and final exam but students are allowed to choose their *own weights* at the start of the semester. The only restrictions are that you cannot give an exam negative weight, and your weights must sum to one. You *are* allowed to give an exam zero weight. By testing both exams on a very large number of students, Dr. Evil has determined that the population mean score on the midterm is 75 points out of 100 with a standard deviation of 10 points. In contrast, the population mean score on his final is 85 points out of 100 with a standard deviation of 15 points. Scores on the midterm and final are *not* independent: the population correlation equals 0.5.

2 (a) Let w be the weight you elect to place on the midterm, and let (X, Y, Z) be three RVs representing your grades: X is your midterm exam score, Y your final exam score, and Z your overall course grade. Express Z as a function of X, Y and w.

Solution: $Z = wX + (1-w)Y$

3 (b) Express the expected value of your course grade as a function of w. Simplify your result: you'll need this expression below.

Solution: By the Linearity of Expectation,

$$E[Z] = wE[X] + (1-w)E[Y] = 75w + 85(1-w) = 85 - 10w$$

3 (c) Suppose you wanted to maximize your expected grade in Dr. Evil's course using the population information given above. What weights should you choose?

Name: ____________________ Student ID #: ____________________

> **Solution:** To maximize the function from the preceding part we need to make w as small as possible. Since you cannot give an exam negative weight, you should set $w = 0$ which corresponds to giving 100% of the weight to the final.

2 (d) Calculate the covariance between scores on Dr. Evil's midterm and final exams.

> **Solution:** $Cov(X,Y) = SD(X) \times SD(Y) \times Corr(X,Y) = 10 \times 15 \times 0.5 = 75$

6 (e) Express the variance of your course grade as a function of w. Simplify your result: you'll need this expression below.

> **Solution:**
>
> $$\begin{aligned} Var(Z) &= w^2 Var(X) + (1-w)^2 Var(Y) + 2w(1-w)Cov(X,Y) \\ &= 100w^2 + 225(1 - 2w + w^2) + 150(w - w^2) \\ &= 225 - 300w + 175w^2 \end{aligned}$$

4 (f) Suppose you wanted to minimize the variance of your grade in Dr. Evil's course using the population information given above. What weights should you choose? You do not need to check the second order condition.

> **Solution:** Differentiating the solution from the previous part with respect to w yields the first order condition $350w - 300 = 0$. Solving, $w^* = 6/7 \approx 0.86$.

7. Aimee wants to find out the fraction of Democratic primary voters who support Bernie Sanders, so she carries out a poll. Her dataset contains responses from 500 individuals, of whom 200 say they support Sanders. For parts (a) and (b), assume that Aimee's 500 individuals constitute a random sample from the population of all Democratic primary voters. Part (c) asks you to explore the possible consequences of non-random sampling.

6 (a) Calculate Aimee's estimate $\widehat{p}$ of the true population proportion p of voters who support Sanders, along with its associated standard error.

> **Solution:** $\widehat{p} = 200/500 = 0.4$ and
>
> $$\widehat{SE}(\widehat{p}) = \sqrt{\widehat{p}(1-\widehat{p})/n} = \sqrt{0.4 \times 0.6/500} \approx 0.02$$

8 (b) Continuing from the preceding parts, construct an approximate 95% confidence interval for p based on the Central Limit Theorem.

Name: ____________________ Student ID #: ____________________

Solution: The margin of error for an approximate 95% CI based on the Central Limit Theorem is roughly $2 \times \widehat{SE}(\widehat{p}) \approx 0.04$. Thus, the confidence interval is 0.4 ± 0.04 or equivalently (0.36, 0.44).

6 (c) It turns out that the 500 individuals in Aimee's poll might *not* represent a random sample: although she sent the poll to 800 people, only 500 of them responded. Suppose we know nothing about the 300 individuals who *failed* to respond to Aimee's poll. What is the range of possible values for $\widehat{p}$ that she *could have gotten* if all of these individuals had responded? Explain in no more than two sentences.

Solution: We calculate the range by considering the two most extreme possibilities. The first is that all of the 300 non-respondents are Sanders supporters. If this were the case, and all of them had responded then Aimee's estimate would have been $(200 + 300)/800 = 5/8 = 87.5\%$. If, on the other hand, *none* of the 300 non-respondents are Sanders supporters, then Aimee's estimate would have been $200/800 = 25\%$ if they had all responded. Thus the overall range of possible estimates is $[25\%, 62.5\%]$.