

SECOND MIDTERM EXAMINATION  
ECON 103, STATISTICS FOR ECONOMISTS

MARCH 27TH, 2018

**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 | 25 | 40 | 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, a point will be deducted for each page on which you do not write your name and student ID.

1. To answer this question you will need to refer to the following joint pmf:

|   |   | Y   |     |     |
|---|---|-----|-----|-----|
|   |   | 1   | 2   | 3   |
| X | 0 | 0   | 0.2 | 0.4 |
|   | 1 | 0.1 | 0.3 | 0   |

- 5 (a) Are  $X$  and  $Y$  independent? Briefly explain why or why not.

**Solution:** No: if  $Y = 1$  then  $X$  must be 1.

- 5 (b) Write down the support set and marginal pmf of  $Y$  and use them to calculate  $E[Y^3]$ .

**Solution:** The support set of  $Y$  is  $\{1, 2, 3\}$  and its marginal pmf is  $p_Y(1) = 0.1, p_Y(2) = 0.5, p_Y(3) = 0.4$ . We know that  $E[g(X)] = \sum_{\text{all } x} g(x)p(x)$ . Hence,

$$E[Y^3] = 1^3 \times 0.1 + 2^3 \times 0.5 + 3^3 \times 0.4 = 0.1 + 4 + 10.8 = 14.9$$

- 5 (c) Write out the CDF of  $Y$ .

**Solution:**

$$F_Y(y_0) = \begin{cases} 0 & y_0 < 1 \\ 0.1 & 1 \leq y_0 < 2 \\ 0.6 & 2 \leq y_0 < 3 \\ 1 & y_0 \geq 3 \end{cases}$$

- 5 (d) Derive the conditional pmf of  $X$  given  $Y = 2$ .

**Solution:**  $p_{X|Y}(0|2) = 0.2/0.5 = 0.4$  and  $p_{X|Y}(1|2) = 0.3/0.5 = 0.6$

- 5 (e) Calculate the expected value of the *ratio* of  $X$  and  $Y$ , namely  $E\left[\frac{X}{Y}\right]$ .

**Solution:** We know that  $E[g(X, Y)] = \sum_{\text{all } x, y} g(x, y)p_{XY}(x, y)$ , hence

$$E[X/Y] = 0.1 \times (1/1) + 0.3 \times (1/2) = 0.25$$

2. Let  $X$  be a random variable with support set  $[-1, 1]$  and probability density function  $f(x) = c(1 - x^2)$  where  $c$  is a constant.

Name: \_\_\_\_\_

Student ID #: \_\_\_\_\_

- 5 (a) Calculate  $E[X]$ . If needed, your answer may depend on  $c$ .

**Solution:**

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = c \int_{-1}^1 (x - x^3) dx = c \left( \frac{x^2}{2} - \frac{x^4}{4} \right) \Big|_{-1}^1 \\ &= c \left[ \left( \frac{1}{2} - \frac{1}{4} \right) - \left( \frac{1}{2} - \frac{1}{4} \right) \right] = 0 \end{aligned}$$

- 5 (b) Calculate  $\text{Var}(X)$ . If needed, your answer may depend on  $c$ .

**Solution:** Since  $E[X] = 0$ , by the shortcut formula  $\text{Var}(X) = E[X^2]$ :

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = c \int_{-1}^1 (x^2 - x^4) dx = c \left( \frac{x^3}{3} - \frac{x^5}{5} \right) \Big|_{-1}^1 \\ &= c \left[ \left( \frac{1}{3} - \frac{1}{5} \right) - \left( -\frac{1}{3} + \frac{1}{5} \right) \right] = c \left[ \frac{2}{3} - \frac{2}{5} \right] = \frac{4}{15}c \end{aligned}$$

- 5 (c) Calculate the CDF of  $X$ . If needed, your answer may depend on  $c$ .

**Solution:**

$$\begin{aligned} F(x_0) &= \int_{-\infty}^{\infty} f(x) dx = c \int_{-1}^{x_0} (1 - x^2) dx = c \left( x - \frac{x^3}{3} \right) \Big|_{-1}^{x_0} \\ &= c \left[ \left( x_0 - \frac{x_0^3}{3} \right) - \left( -1 + \frac{1}{3} \right) \right] = c \left( \frac{2}{3} + x_0 - \frac{x_0^3}{3} \right) \end{aligned}$$

The preceding expression is valid for  $x_0 \in [-1, 1]$ . For  $x_0 > 1$ ,  $F(x_0) = 1$ . For  $x_0 < -1$ ,  $F(x_0) = 0$ .

- 5 (d) Calculate  $c$ .

**Solution:** We can solve for  $c$  by setting  $F(1) = 1$ :

$$\begin{aligned} c \left( \frac{2}{3} + 1 - \frac{1}{3} \right) &= 1 \\ c(4/3) &= 1 \\ c &= 3/4 \end{aligned}$$

This question is taken from your homework assignment for Lectures 13–14. I have re-worded the question slightly for clarity, but the substance and solutions remain unchanged.

3. Let  $X_1$  and  $X_2$  be independent random variables with  $E[X_1] = E[X_2] = \mu$ ,  $\text{Var}(X_1) = \sigma^2$  and  $\text{Var}(X_2) = 3\sigma^2$ . Define  $\bar{X} = (X_1 + X_2)/2$ .

- 5 (a) Calculate the variance of  $\bar{X}$ .

**Solution:** In this example,

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4} [\text{Var}(X_1) + \text{Var}(X_2)] = \frac{1}{4} (\sigma^2 + 3\sigma^2) = \sigma^2$$

- 5 (b) Let  $\tilde{\mu} = \omega X_1 + (1 - \omega)X_2$  for  $\omega \in [0, 1]$ . Is  $\tilde{\mu}$  an unbiased estimator of  $\mu$ ? Justify your answer.

**Solution:** This estimator is unbiased:

$$E[\tilde{\mu}] = E[\omega X_1 + (1 - \omega)X_2] = \omega\mu + (1 - \omega)\mu = \mu$$

- 5 (c) Define  $\tilde{\mu}$  as in part (b). Calculate the variance of  $\tilde{\mu}$ .

**Solution:**

$$\text{Var}(\tilde{\mu}) = \text{Var}[\omega X_1 + (1 - \omega)X_2] = \omega^2\sigma^2 + 3(1 - \omega)^2\sigma^2$$

- 10 (d) What value of  $\omega$  minimizes  $\text{Var}(\tilde{\mu})$ ? What is the lowest possible variance? You do *not* have to check the second order condition.

**Solution:** We choose  $\omega$  to minimize  $\omega^2\sigma^2 + 3(1 - \omega)^2\sigma^2$ , yielding the FOC:

$$2\omega\sigma^2 - 6(1 - \omega)\sigma^2 = 0$$

$$2\sigma^2 [\omega - 3(1 - \omega)] = 0$$

$$\omega - 3 + 3\omega = 0$$

$$4\omega = 3$$

$$\omega = 3/4$$

Thus, the minimum variance is:

$$\text{Var}\left(\frac{3}{4}X_1 + \frac{1}{4}X_2\right) = \frac{9}{16}\sigma^2 + \frac{1}{16} \times 3\sigma^2 = \frac{12}{16}\sigma^2 = \frac{3}{4}\sigma^2$$

- 5 (e) Which is the more efficient estimator of  $\mu$ : the sample mean, or  $\tilde{\mu}$  using the optimal choice of  $\omega$  from the preceding part? Explain briefly.

**Solution:** In this example the sample mean is NOT efficient because there is another unbiased estimator with a lower variance, namely  $3X_1/4 + X_2/4$ . We know this because we showed in part (d) that

$$\text{Var}\left(\frac{3}{4}X_1 + \frac{1}{4}X_2\right) = \frac{3}{4}\sigma^2$$

whereas, from part (a)

$$\text{Var}(\bar{X}_n) = \sigma^2$$

4. Let  $Z \sim N(0, 1)$ . Answer each of the following parts. No explanations are needed.

- 3 (a) Provide a line of R code to make 10 iid standard normal draws.

**Solution:** `rnorm(10)`

- 3 (b) Provide a line of R code that calculates  $P(Z > 1.5)$ .

**Solution:** `1 - pnorm(1.5)`

- 3 (c) Provide a line of R code to calculate the 90th percentile of  $Z$ .

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

- 3 (d) Provide a line of R code to calculate the pdf of  $Z$  evaluated at 1.

**Solution:** `dnorm(1)`

- 3 (e) Provide a line of R code to find  $c > 0$  such that  $P(-c \leq Z \leq c) = 0.80$ .

**Solution:** `qnorm(0.9)` or `-qnorm(0.1)`

- 10 (f) Let `handspan` be an R vector containing measurements of the handspan of a population of students. Suppose we want to study the sampling distribution of the *sample median* based on a random sample of size  $n = 20$  drawn from this population. Write R code that uses 10,000 simulation replications to create a histogram that approximates this sampling distribution. You may assume that `handspan` has no missing values.

**Solution:**

```
sim_draw <- function(){
  median(sample(handspan, size = 20, replace = TRUE))
}
sample_medians <- replicate(10000, sim_draw())
hist(sample_medians)
```

5. Let  $X_1, \dots, X_n \sim \text{iid } N(\mu, \sigma^2)$  where  $\sigma^2$  is known and define  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ .

- 5 (a) What is the sampling distribution of  $\bar{X}_n$ ? Be sure to specify the values of any and all relevant parameters.

**Solution:**  $\bar{X}_n \sim N(\mu, \sigma^2/n)$

- 5 (b) What is the sampling distribution of  $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ ? Be sure to specify the values of any and all relevant parameters.

**Solution:**  $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0, 1)$

- 5 (c) Let  $\mu = 1$ ,  $\sigma^2 = 4$ , and  $n = 16$ . Give the approximate value of  $P(0 \leq \bar{X}_n \leq 1.5)$ . To be clear: your answer should be a specific numeric value, not an R command.

**Solution:**

$$\begin{aligned} P(0 \leq \bar{X}_n \leq 1.5) &= P\left(\frac{0 - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{1.5 - \mu}{\sigma/\sqrt{n}}\right) \\ &= P(-2 \leq Z \leq 1) = \text{pnorm}(1) - \text{pnorm}(-2) \\ &\approx 0.84 - 0.025 = 0.815 \end{aligned}$$

- 5 (d) Suppose that  $n = 100$ ,  $\sigma^2 = 4$ , and we observe  $\bar{x} = 20$ . Construct a 95% confidence interval for  $\mu$ .

**Solution:** The margin of error is  $2\sigma/\sqrt{n} = 2 \times 2/10 = 0.4$  so the confidence interval is  $20 \pm 0.4$  or equivalently  $(19.6, 20.4)$ .

- 5 (e) Suppose  $\sigma^2 = 9$  and we want to construct a 95% confidence interval for  $\mu$ . How large a sample size  $n$  would we need to get a margin of error equal to 0.1?

**Solution:** The margin of error is  $2 \times \sigma / \sqrt{n} = 6 / \sqrt{n}$ . Setting this equal to  $1/10$  and solving for  $n$ , we obtain

$$1/10 = 6/\sqrt{n}$$

$$\sqrt{n} = 60$$

$$n = 3600$$

- 5 (f) Suppose we want to construct a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mu$ . Write down the expression for the margin of error of this interval. Your answer may involve  $\sigma^2$ ,  $n$ ,  $\alpha$ , and an R command if needed.

**Solution:**  $\sigma / \sqrt{n} \times \text{qnorm}(1 - \alpha/2)$

- 10 (g) Write an R function called `normalCI` that constructs a confidence interval for the mean  $\mu$  of a normal population with known variance  $\sigma^2$  and confidence level  $1 - \alpha$ . Your function should take three input arguments: a vector of data `x`, a probability `alpha`, and the population variance `popvar`. It should return a vector of two elements: the first should be the LCL and the second the UCL of a  $100 \times (1 - \text{alpha})\%$  confidence interval for  $\mu$  constructed from the sample mean of `x`. You may assume that `x` has no missing values.

**Solution:**

```
normalCI <- function(x, alpha, popvar) {  
  xbar <- mean(x)  
  ME <- qnorm(1 - alpha/2) * sqrt(popvar / length(x))  
  LCL <- xbar - ME  
  UCL <- xbar + ME  
  return(c(LCL, UCL))  
}
```