

MIDTERM EXAMINATION II  
ECON 103, STATISTICS FOR ECONOMISTS

MARCH 25TH, 2014

**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: \_\_\_\_\_

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

Signature: \_\_\_\_\_

Question:	1	2	3	4	5	6	7	Total
Points:	25	20	15	30	20	15	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. Let  $X_1 \sim \text{Bernoulli}(1/2)$  independently of  $X_2 \sim \text{Bernoulli}(2/3)$  and define  $Y = 2X_1$  and  $Z = X_2 - X_1$ .
- (a) (15 points) Express the joint pmf of  $Y$  and  $Z$  in tabular form. Please put the realizations of  $Y$  in the *rows* of the table and realizations of  $Z$  in the *columns*.

**Solution:**

		Z		
		-1	0	1
Y	0	0	1/6	1/3
	2	1/6	1/3	0

$p_{YZ}(0, -1) = 0$  since  $Y = 0$  implies  $X_1 = 0$  so  $Z$  is at least 0.

$p_{YZ}(0, 0) = 1/2 \times 1/3 = 1/6$  since this corresponds to  $X_1 = X_2 = 0$ .

$p_{YZ}(0, 1) = 1/2 \times 2/3 = 1/3$  since this corresponds to  $X_1 = 0, X_2 = 1$ .

$p_{YZ}(2, -1) = 1/2 \times 1/3 = 1/6$  since this corresponds to  $X_1 = 1, X_2 = 0$ .

$p_{YZ}(2, 0) = 1/2 \times 2/3 = 1/3$  since this corresponds to  $X_1 = X_2 = 1$ .

$p_{YZ}(2, 1) = 0$  since  $Y = 2$  implies  $X_1 = 1$  so  $Z$  is at most 0.

- (b) (3 points) What is the marginal pmf of  $Y$ ?

**Solution:** Since  $X_1$  is a Bernoulli(1/2) random variable and  $Y = 2X_1$ , the support of  $Y$  is  $\{0, 2\}$ ,  $p_Y(0) = 1/2$  and  $p_Y(2) = 1/2$ . This agrees with the values we obtain by summing over the rows of the table in the previous part.

- (c) (3 points) What is the marginal pmf of  $Z$ ?

**Solution:** The support of  $Z$  is  $\{-1, 0, 1\}$ . Summing over the columns of the table we find that  $p_Z(-1) = 1/6$ ,  $p_Z(0) = 1/2$  and  $p_Z(1) = 1/3$ .

- (d) (4 points) Calculate the covariance between  $Y$  and  $Z$ . Are  $Y$  and  $Z$  independent?

**Solution:** We proceed via the shortcut formula:

$$\text{Cov}(Y, Z) = E[YZ] - E[Y]E[Z]$$

From the marginal pmfs,

$$E[Y] = 0 \times 1/2 + 2 \times 1/2 = 1$$

$$E[Z] = -1 \times 1/6 + 0 \times 1/2 + 1 \times 1/3 = 1/6$$

Calculating  $E[YZ]$  is easy since there is only *one* non-zero term in the sum:  $2 \times -1 \times 1/6 = -1/3$ . Thus, we have

$$\text{Cov}(Y, Z) = -1/3 - 1/6 \times 1 = -1/2$$

Because they have a nonzero covariance,  $Y$  and  $Z$  cannot be independent.

2. Let  $X$  be a continuous random variable with support  $[1, c]$  and pdf  $f(x) = 1/x$ .

(a) (5 points) What is the value of the constant  $c$ ?

**Solution:**

$$\int_1^c \frac{dx}{x} = \log(c)|_1^c = \log(c) - \log(1) = \log(c)$$

Since a pdf must integrate to one, we have  $1 = \log(c)$ . Exponentiating,  $c = e$ .

(b) (5 points) Calculate the CDF of  $X$ .

**Solution:**

$$F(x_0) = \int_{-\infty}^{x_0} f(x) dx = \int_1^{x_0} \frac{dx}{x} = \log(x)|_1^{x_0} = \log(x_0) - \log(1) = \log(x_0)$$

(c) (5 points) Calculate  $E[X]$ .

**Solution:**

$$E[X] = \int_1^e x \frac{dx}{x} = \int_1^e dx = x|_1^e = e - 1$$

(d) (5 points) Calculate  $E[X^2]$ .

**Solution:**

$$E[X^2] = \int_1^e x^2 \frac{dx}{x} = \int_1^e x dx = \frac{x^2}{2} \Big|_1^e = \frac{e^2 - 1}{2}$$

3. Suppose that  $Y \sim N(\mu = 2, \sigma^2 = 3)$  independently of  $X \sim N(\mu = -1, \sigma^2 = 6)$ .
- (a) (5 points) Let  $Q = X - Y$ . What kind of random variable is  $Q$ ? Be sure to specify its parameters.

**Solution:**  $Q \sim N(\mu = -3, \sigma^2 = 9)$

- (b) (10 points) Approximately what is the probability that  $X > Y$ ? Explain.

**Solution:**

$$\begin{aligned} P(X > Y) &= P(X - Y > 0) = P(Q > 0) \\ &= P\left(\frac{Q - (-3)}{3} > \frac{0 - (-3)}{3}\right) \\ &= P(Z > 1) \approx 0.16 \end{aligned}$$

where  $Z \sim N(0, 1)$ . To approximate the probability we used the fact that a standard normal has a probability of approximately 0.68 of falling in the interval  $(-1, 1)$  along with the symmetry of the normal distribution.

4. Garth wants to learn how much taller NBA players are than Penn Undergraduates, on average. To answer this question, he's recruited volunteers to make up two independent random samples. The first sample contains 10 NBA players:  $X_1, \dots, X_{10} \sim \text{iid}$  with mean  $\mu_X$  and variance  $\sigma^2$ . The second sample is independent of the first and contains 15 Penn Undergrads:  $Y_1, \dots, Y_{15} \sim \text{iid}$  with mean  $\mu_Y$  and variance  $\sigma^2$ . Just to be completely clear, in this question we are assuming that the variance is identical for Penn Students and NBA players to make the calculations simpler.
- (a) (4 points) To answer his question, Garth needs to estimate  $\mu_X - \mu_Y$ . There is an obvious unbiased estimator of this quantity. What is it? Prove that it is unbiased.

**Solution:** The obvious choice is the difference of sample means:  $\bar{X} - \bar{Y}$ . We know that this is unbiased because  $\bar{X}$  is an unbiased estimator of  $\mu_X$ ,  $\bar{Y}$  is an unbiased estimator of  $\mu_Y$  and by the linearity of expectation

$$E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu_X - \mu_Y$$

- (b) (4 points) Calculate the variance of the estimator you proposed in part (a).

**Solution:**

$$\text{Var}(\bar{X} - \bar{Y}) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) = \frac{\sigma^2}{10} + \frac{\sigma^2}{15} = \frac{\sigma^2}{6}$$

- (c) (8 points) When measuring the players and students Garth makes a mistake: although he accurately records each of the 25 heights, he forgets to note which correspond to Penn students and which correspond to NBA players. Fortunately Garth remembers that, among the first 10 heights on his list, there were exactly 5 students and 5 NBA players. In other words, the first 10 heights on his list are  $X_1, \dots, X_5$  and  $Y_1, \dots, Y_5$  in some unknown order and the last 15 are  $X_6, \dots, X_{10}$  and  $Y_6, \dots, Y_{15}$  in some unknown order. Let  $\bar{Z}_1$  be the sample mean of the first 10 heights on Garth's list and  $\bar{Z}_2$  be the sample mean of the last 15. Prove that

$$E[\bar{Z}_1] = \frac{\mu_X + \mu_Y}{2} \quad \text{and} \quad E[\bar{Z}_2] = \frac{1}{3}\mu_X + \frac{2}{3}\mu_Y$$

For full credit, provide a careful explanation and cite any results you have used.

**Solution:** We have

$$\bar{Z}_1 = \frac{1}{10} (X_1 + X_2 + \dots + X_5 + Y_1 + Y_2 + \dots + Y_5)$$

Hence, by the linearity of expectation,

$$E[\bar{Z}_1] = \frac{1}{10} (5\mu_X + 5\mu_Y) = \frac{\mu_X + \mu_Y}{2}$$

Similarly,

$$\bar{Z}_2 = \frac{1}{15} (X_6 + X_7 + \dots + X_{10} + Y_6 + Y_7 + \dots + Y_{15})$$

and, again, by the linearity of expectation

$$E[\bar{Z}_2] = \frac{1}{15} (5\mu_X + 10\mu_Y) = \frac{1}{3}\mu_X + \frac{2}{3}\mu_Y$$

- (d) (4 points) Let  $\alpha$  and  $\beta$  be two arbitrary constants. Using part (c), calculate  $E[\alpha\bar{Z}_1 + \beta\bar{Z}_2]$ . Simplify your answer so that it takes the form  $c_1\mu_X + c_2\mu_Y$  where  $c_1$  and  $c_2$  are two constants that will depend on  $\alpha$  and  $\beta$ .

**Solution:** By the linearity of expectation:

$$\begin{aligned} E[\alpha\bar{Z}_1 + \beta\bar{Z}_2] &= \alpha E[\bar{Z}_1] + \beta E[\bar{Z}_2] \\ &= \alpha \left( \frac{\mu_X + \mu_Y}{2} \right) + \beta \left( \frac{1}{3}\mu_X + \frac{2}{3}\mu_Y \right) \\ &= \mu_X \left( \frac{\alpha}{2} + \frac{\beta}{3} \right) + \mu_Y \left( \frac{\alpha}{2} + \frac{2\beta}{3} \right) \end{aligned}$$

- (e) (6 points) Back to Garth's problem: *amazingly* it's still possible to construct an unbiased estimator of  $\mu_X - \mu_Y$  *without knowing* which observations corresponded to NBA players and which corresponded to Penn students. The trick is to take a particular linear combination of  $\bar{Z}_1$  and  $\bar{Z}_2$ . Use your answer to the previous part to find the values of  $\alpha$  and  $\beta$  that give  $E[\alpha\bar{Z}_1 + \beta\bar{Z}_2] = \mu_X - \mu_Y$ .

**Solution:** We have two linearly independent equations in two unknowns:

$$\begin{aligned} \alpha/2 + \beta/3 &= 1 \\ \alpha/2 + 2\beta/3 &= -1 \end{aligned}$$

Solving the first equation gives  $\alpha/2 = 1 - \beta/3$ . Substituting this into the second:

$$\begin{aligned} (1 - \beta/3) + 2\beta/3 &= -1 \\ \beta &= -6 \end{aligned}$$

Hence  $\alpha = 2(1 - \beta/3) = 6$ . Plugging these values into the result of the previous part, we can verify that these values indeed give an unbiased estimator:

$$E[6\bar{Z}_1 - 6\bar{Z}_2] = 3(\mu_X + \mu_Y) - (2\mu_X + 4\mu_Y) = \mu_X - \mu_Y$$

- (f) (4 points) Although it's unbiased, there's a clear downside to the estimator from the previous part: it has a very high variance. Calculate its variance and compare it to that of the estimator from part (a). Explain the intuition behind the difference.

**Solution:** We have:

$$\text{Var}(6\bar{Z}_1 - 6\bar{Z}_2) = 36 [\text{Var}(\bar{Z}_1) + \text{Var}(\bar{Z}_2)] = 36(\sigma^2/10 + \sigma^2/15) = 6\sigma^2$$

which is 36 times as large as  $\text{Var}(\bar{X} - \bar{Y})$ ! Although  $\text{Var}(\bar{X} - \bar{Y}) = \text{Var}(\bar{Z}_1 -$

$\bar{Z}_2$ ), when we multiplied  $\bar{Z}_1$  and  $\bar{Z}_2$  by six to get an unbiased estimator, this multiplied the variance by  $6^2 = 36$ .

5. Suppose  $X_1, \dots, X_n \sim \text{iid}$  with mean  $\mu$  and variance  $\sigma^2$ . Let  $\bar{X}_n$  be the sample mean and define  $\hat{\mu}_n = n\bar{X}_n/(n+1)$ .

- (a) (15 points) Calculate the MSE of  $\hat{\mu}_n$  as an estimator of  $\mu$ .

**Solution:** First we calculate the bias:

$$\begin{aligned} E[\hat{\mu} - \mu] &= E\left[\left(\frac{n}{n+1}\right)\bar{X}_n - \mu\right] = \left(\frac{n}{n+1}\right)E[\bar{X}_n] - \mu \\ &= \left(\frac{n}{n+1}\right)\mu - \mu = \mu\left[\frac{n}{n+1} - 1\right] = \frac{-\mu}{n+1} \end{aligned}$$

Next we calculate the variance:

$$\text{Var}(\hat{\mu}_n) = \left(\frac{n}{n+1}\right)^2 \text{Var}(\bar{X}_n) = \left(\frac{n}{n+1}\right)^2 \frac{\sigma^2}{n} = \frac{n\sigma^2}{(n+1)^2}$$

Thus, MSE is given by

$$\text{MSE}(\hat{\mu}_n) = \frac{\mu^2 + n\sigma^2}{(n+1)^2}$$

- (b) (5 points) Is  $\hat{\mu}_n$  a consistent estimator of  $\mu$ ? Briefly explain your answer.

**Solution:** Yes: in the limit as  $n \rightarrow \infty$  both the bias and the variance converge to zero so  $\hat{\mu}_n$  is a consistent estimator of  $\mu$ .

6. This question concerns the standard normal RV and related R functions.

- (a) (5 points) Write R code to plot the standard normal pdf between -3 and 3 using a step size of 1/100. Be sure to plot a *smooth curve* rather than isolated points.

**Solution:**

```
x <- seq(from = -3, to = 3, by = 0.01)
f.x <- dnorm(x)
plot(x, f.x, type = 'l')
```

- (b) (5 points) Write R code to approximate the probability that a standard normal random variable takes on a value greater than 0.5 using a Monte Carlo experiment with 10,000 draws.

**Solution:**

```
sims <- rnorm(10000)
sum(sims > 0.5)/length(sims)
```

- (c) (5 points) There is an R command that gives the *exact* answer to the preceding part without using simulation. What is it?

**Solution:** `1 - pnorm(0.5)`

7. This question concerns the mean of a  $\chi^2(5)$  random variable.

- (a) (10 points) Write R code to approximate the mean of a  $\chi^2(5)$  random variable by making 10,000 simulation draws. You may use any R functions you like in your answer *except* `rchisq`.

**Solution:**

```
rchisq5 <- function(){
  out <- rnorm(5)
  return(sum(out^2))
}
sims <- replicate(10000, rchisq5())
mean(sims)
```

- (b) (5 points) Use the facts we studied in class to *derive* the mean of a  $\chi^2(5)$  random variable. Hint: recall how we derived the mean of the Binomial RV in class.

**Solution:** Let  $Y$  be a  $\chi^2(5)$  random variable. Then,

$$Y = Z_1^2 + \cdots + Z_5^2$$

where  $Z_1, \dots, Z_5$  are independent standard normal random variables. Let  $Z$  be one of these. By the shortcut formula

$$E[Z^2] = \text{Var}(Z) + E[Z]^2 = 1 + 0^2 = 1$$

Therefore, by the linearity of expectation,

$$E[Y] = E[Z_1^2] + \cdots + E[Z_5^2] = 1 + \cdots + 1 = 5$$