

SECOND MIDTERM EXAMINATION
ECON 103, STATISTICS FOR ECONOMISTS

NOVEMBER 7, 2012

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 | 8 | Total |
| Points: | 20 | 15 | 20 | 15 | 20 | 15 | 15 | 20 | 140 |
| Score: | | | | | | | | | |

Instructions: Answer all questions in the space provided. Should you run out of space, continue on the back of the page. 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.

1. Suppose that X is a random variable with support $\{1, 2\}$ and Y is a random variable with support $\{0, 1\}$ where X and Y have the following joint distribution:

$$p_{XY}(1, 0) = 0.30$$

$$p_{XY}(1, 1) = 0.25$$

$$p_{XY}(2, 0) = 0.20$$

$$p_{XY}(2, 1) = 0.25$$

- (a) (2 points) Express the joint probability mass function (pmf) in a 2×2 table.

Solution:

| | | X | |
|---|---|------|------|
| | | 1 | 2 |
| Y | 0 | 0.30 | 0.20 |
| | 1 | 0.25 | 0.25 |

- (b) (3 points) Using the table, calculate the marginal pmfs of X and Y .

Solution:

$$p_X(1) = p_{XY}(1, 0) + p_{XY}(1, 1) = 0.30 + 0.25 = 0.55$$

$$p_X(2) = p_{XY}(2, 0) + p_{XY}(2, 1) = 0.20 + 0.25 = 0.45$$

$$p_Y(0) = p_{XY}(1, 0) + p_{XY}(2, 0) = 0.30 + 0.20 = 0.50$$

$$p_Y(1) = p_{XY}(1, 1) + p_{XY}(2, 1) = 0.25 + 0.25 = 0.50$$

- (c) (5 points) Calculate the conditional pmfs of $Y|X = 1$ and $Y|X = 2$.

Solution: The distribution of $Y|X = 1$ is

$$P(Y = 0|X = 1) = \frac{p_{XY}(1, 0)}{p_X(1)} = \frac{0.30}{0.55} = \frac{6}{11}$$

$$P(Y = 1|X = 1) = \frac{p_{XY}(1, 1)}{p_X(1)} = \frac{0.25}{0.55} = \frac{5}{11}$$

while the distribution of $Y|X = 2$ is

$$P(Y = 0|X = 2) = \frac{p_{XY}(2, 0)}{p_X(2)} = \frac{0.20}{0.45} = \frac{4}{9}$$

$$P(Y = 1|X = 2) = \frac{p_{XY}(2, 1)}{p_X(2)} = \frac{0.25}{0.45} = \frac{5}{9}$$

- (d) (3 points) Calculate $\mathbb{E}[Y|X = 1]$ and $\mathbb{E}[Y|X = 2]$.

Solution:

$$\mathbb{E}[Y|X = 1] = 0 \cdot 6/11 + 1 \cdot 5/11 = 5/11$$

$$\mathbb{E}[Y|X = 2] = 0 \cdot 4/9 + 1 \cdot 5/9 = 5/9$$

- (e) (7 points) Calculate the covariance between X and Y .

Solution: First, from the marginal distributions, $\mathbb{E}[Y] = 0.5$ and $E[X] = 1 \cdot 0.55 + 2 \cdot 0.45 = 1.45$. Hence $\mathbb{E}[X]\mathbb{E}[Y] = 0.725$. Second,

$$\begin{aligned} \mathbb{E}[XY] &= (0 \cdot 1) \cdot 0.3 + (0 \cdot 2) \cdot 0.2 + (1 \cdot 1) \cdot 0.25 + (1 \cdot 2) \cdot 0.25 \\ &= 0.25 + 0.50 = 0.75 \end{aligned}$$

Finally $Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.750 - 0.725 = 0.025$

2. The random variables X_1 and X_2 correspond to the annual returns of Stock 1 and Stock 2. Suppose that $\mathbb{E}[X_1] = 0.1$, $\mathbb{E}[X_2] = 0.3$, $Var(X_1) = Var(X_2) = 1$, and $\rho = Corr(X_1, X_2)$. A portfolio $\Pi(\omega)$ is defined by the proportion ω of Stock # 1 that it contains. That is, $\Pi(\omega) = \omega X_1 + (1 - \omega)X_2$ where $0 \leq \omega \leq 1$.

- (a) (3 points) What value of ω gives a portfolio with expected return 0.15?

Solution:

$$\mathbb{E}[\Pi(\omega)] = \omega\mathbb{E}[X_1] + (1 - \omega)\mathbb{E}[X_2] = 0.15$$

$$\omega/10 + 3(1 - \omega)/10 = 0.15$$

$$\omega + 3 - 3\omega = 3/2$$

$$-4\omega + 6 = 3$$

$$\omega = 3/4$$

- (b) (6 points) Suppose that $\omega = 1/4$. In terms of ρ , what is the portfolio variance?

Solution: First, we have

$$Var[\Pi(\omega)] = \omega^2 Var(X_1) + (1 - \omega)^2 Var(X_2) + 2\omega(1 - \omega)Cov(X_1, X_2)$$

but since $Var(X_1) = Var(X_2) = 1$ and $\rho = Corr(X_1, X_2)$,

$$Var[\Pi(\omega)] = \omega^2 + (1 - \omega)^2 + 2\omega(1 - \omega)\rho$$

Substituting $\omega = 1/4$,

$$\begin{aligned} Var[\Pi(1/4)] &= 1/16 + 9/16 + 2(1/4)(3/4)\rho \\ &= 10/16 + 6\rho/16 \\ &= (3\rho + 5)/8 \end{aligned}$$

- (c) (3 points) Again, suppose that $\omega = 1/4$. What are the maximum and minimum values of the portfolio variance? What are the corresponding values of ρ ?

Solution: From the previous part, $Var[\Pi(\omega)] = (3\rho + 5)/8$. By inspection this variance takes on a maximum value of 1 when $\rho = 1$. It takes on a minimum value of 0.25 when $\rho = -1$.

- (d) (3 points) If we assume that variance is a reasonable measure of risk, what does your answer to part (c) suggest about the benefits of constructing a portfolio rather than holding only one stock? Explain briefly.

Solution: We see that the variance of this portfolio cannot exceed the variance of the individual assets that make it up. Unless $\rho = 1$, the portfolio is less risky than either of the individual stocks.

3. Suppose that X is a continuous random variable with probability density function

$$f(x) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

- (a) (5 points) Calculate the cumulative distribution function, $F(x_0)$, of X .

Solution: $\int_{-\infty}^{x_0} f(x) dx = \int_0^{x_0} 2x dx = x^2|_0^{x_0} = x_0^2$. Hence,

$$F(x_0) = \begin{cases} 0 & \text{for } x_0 < 0 \\ x_0^2 & \text{for } 0 \leq x_0 \leq 1 \\ 1 & \text{for } x_0 > 1 \end{cases}$$

- (b) (3 points) Calculate the median of X using your answer to part (a).

Solution: We want $x \in [0, 1]$ such that $F(x) = 0.5$. From part (a), $F(x) = x^2$ for $x \in [0, 1]$. Hence,

$$x = \sqrt{0.5} = \sqrt{1/2} = 1/\sqrt{2} \approx 0.71$$

(We take the positive square root since this lies on the support of X .)

- (c) (5 points) Calculate $\mathbb{E}[X]$.

Solution:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}x^3 \Big|_0^1 = 2/3$$

- (d) (5 points) Calculate $\mathbb{E}[X^2]$

Solution:

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2}x^4 \Big|_0^1 = 1/2$$

- (e) (2 points) Using your answers to (c) and (d) along with the shortcut formula for variance, calculate $\text{Var}(X)$.

Solution: $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1/2 - (2/3)^2 = 1/2 - 4/9 = 1/18$

4. Suppose that $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$. Define $S_n = \sum_{i=1}^n X_i$.

- (a) (3 points) Write down the pmf and support of X_1 .

Solution: The support of a Bernoulli random variable is $\{0, 1\}$. Since 1 is defined as “success” and p defined as the probability of success, the pmf is $p_X(1) = p$, $p_X(0) = 1 - p$.

- (b) (2 points) Calculate $\mathbb{E}[X_1]$.

Solution: $\mathbb{E}[X_1] = 0 \cdot (1 - p) + 1 \cdot p = p$

- (c) (2 points) Calculate $\mathbb{E}[X_1^2]$.

$$\textbf{Solution: } \mathbb{E}[X_1^2] = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

- (d) (3 points) Calculate $\text{Var}(X_1)$ using the shortcut formula.

$$\textbf{Solution: } \text{Var}(X_1) = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = p - p^2 = p(1 - p).$$

- (e) (5 points) What kind of random variable is S_n ? Write down its pmf and support.

Solution: Since it is the sum of n independent, identically distribution Bernoulli random variables, each with probability of success p , $S_n \sim \text{Binomial}(n, p)$. The support of this random variable is $\{0, 1, \dots, n\}$ and its pmf is

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

5. Suppose that $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$ and define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

- (a) (3 points) Calculate $\mathbb{E}[\bar{X}_n]$.

Solution: By the linearity of expectation, and identical distribution:

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^n p = (np)/n = p$$

- (b) (5 points) Calculate $\text{Var}(\bar{X}_n)$.

Solution: By independence, identical distribution, and the fact that the variance of a Bernoulli random variable is $p(1 - p)$,

$$\begin{aligned} \text{Var}[\bar{X}_n] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] \\ &= \frac{1}{n^2} \sum_{i=1}^n p(1 - p) = np(1 - p)/n^2 = p(1 - p)/n \end{aligned}$$

- (c) (4 points) Is \bar{X}_n an unbiased estimator of p ? What about $\hat{p} = \frac{3}{4}X_1 + \frac{1}{4}X_2$?

Solution:

$$\mathbb{E}[\hat{p}] = \mathbb{E}[3X_1/4 + X_2/4] = 3p/4 + p/4 = p$$

Therefore, \hat{p} is unbiased. We know that \bar{X}_n is unbiased from part (a).

- (d) (5 points) Suppose that $n = 2$. Which estimator has a lower variance: \hat{p} or \bar{X}_2 ? Prove your answer.

Solution: We showed above that $\text{Var}(\bar{X}_n) = p(1-p)/n$. Hence, $\text{Var}(\bar{X}_2) = p(1-p)/2$. We calculate the variance of \hat{p} as follows, using independence and identical distribution

$$\begin{aligned} \text{Var}(\hat{p}) &= \text{Var}\left(\frac{3}{4}X_1 + \frac{1}{4}X_2\right) \\ &= \frac{9}{16}p(1-p) + \frac{1}{16}p(1-p) \\ &= \frac{10}{16}p(1-p) = \frac{5}{8}p(1-p) \end{aligned}$$

We see that \bar{X}_2 has a lower variance than \hat{p} since $1/2 < 5/8$.

- (e) (3 points) Provide some intuition for your answers to parts (c) and (d).

Solution: The sample mean gives each observation equal weight. We see from part (c) that this is not necessary for unbiasedness: all we need is that the weights sum to one. However, unequal weighting increases the variance of our estimator.

6. Suppose we carry out a sequence of independent Bernoulli trials, each with probability of success p , and stop as soon as we get the first success.

- (a) (2 points) What is the probability that we get a success on our first trial?

Solution: p

- (b) (3 points) What is the probability that we get our first success on the *second* trial? (That is, what is the probability of a Failure followed by a Success?)

Solution: $p(1 - p)$

- (c) (5 points) What is the probability that we get our first success on the n th trial?

Solution: $p(1 - p)^{n-1}$

- (d) (5 points) Suppose that we define a random variable X that equals the trial number of the first success in a sequence of independent Bernoulli trials, each with probability p of success. This is the definition of a Geometric(p) random variable. What is the probability mass function $p(x) = \mathbb{P}(X = x)$ of X ? What is the support of this random variable?

Solution: The support of X is \mathbf{N} , i.e. $\{1, 2, 3, \dots\}$ and the pmf is

$$p(x) = \begin{cases} p(1 - p)^{x-1} & \text{for } x \in \mathbf{N} \\ 0 & \text{elsewhere} \end{cases}$$

7. Let X and Z be independent random variables with $\mathbb{E}[X] = \mathbb{E}[Z] = 0$, $Var(X) = \sigma_X^2$, and $Var(Z) = \sigma_Z^2$. Define $Y = aX + Z$ where a is a constant.

- (a) (2 points) What is $\mathbb{E}[Y]$?

Solution: $\mathbb{E}[Y] = a\mathbb{E}[X] + \mathbb{E}[Z] = 0$

- (b) (3 points) What is $Var(Y)$?

Solution: $Var(Y) = a^2Var(X) + Var(Z) = a^2\sigma_X^2 + \sigma_Z^2$ since X and Z are independent.

- (c) (5 points) What is $Cov(X, Z)$?

Solution: Since X and Z are independent, their covariance is zero.

- (d) (5 points) What is $Cov(X, Y)$?

Solution: Since $\mu_X = \mu_Z = \mu_Y = 0$,

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] = \mathbb{E}[X(aX + Z)] \\ &= \mathbb{E}[aX^2 + XZ] = a\mathbb{E}[X^2] + \mathbb{E}[XZ] \\ &= a\mathbb{E}[(X - \mu_X)^2] + \mathbb{E}[(X - \mu_X)(Z - \mu_Z)] \\ &= a\text{Var}(X) + \text{Cov}(X, Z) \\ &= a\sigma_X^2 \end{aligned}$$

8. Let X, Y and Z be independent normal random variables where $\mathbb{E}[X] = \mathbb{E}[Z] = 0$, $\mathbb{E}[Y] = 3$, $\text{Var}(X) = 9$, $\text{Var}(Y) = 4$ and $\text{Var}(Z) = 1$. For each of the following, unless I specifically ask you to provide an R command, please give a *numeric* answer.

- (a) (5 points) What is the approximate value of $\mathbb{P}(-1 \leq X/3 \leq 1)$?

Solution: This is simply the probability that a standard normal takes a value in the interval $[-1, 1]$ which we know is 0.68.

- (b) (5 points) What value of k ensures that $\mathbb{P}(3 - k \leq Y \leq 3 + k) \approx 0.95$?

Solution: Rearranging, we need to find k such that

$$\begin{aligned} \mathbb{P}(3 - k \leq Y \leq 3 + k) &= \mathbb{P}(-k \leq Y - 3 \leq k) \\ &= \mathbb{P}\left(-k/2 \leq \frac{Y - 3}{2} \leq k/2\right) \approx 0.95 \end{aligned}$$

Since $(Y - 3)/2$ is a standard normal, we need $k/2 = 2$, hence $k = 4$.

- (c) (5 points) Suppose I want to find the value of d such that $\mathbb{P}(-d \leq Z \leq d) = 0.5$. What R command should I use?

Solution: The command `qnorm(0.25, mean = 0, sd = 1)`, or equivalently `qnorm(0.25)`, gives $-d$. The command `qnorm(0.75, mean = 0, sd = 1)`, or equivalently `qnorm(0.75)`, gives d .

- (d) (5 points) What R command should I use to calculate the probability that the random variable Z^2 is greater than or equal to 5?

Solution: Since $Z^2 \sim \chi^2(1)$, the command is `1 - pchisq(5, df = 1)`.