

SECOND MIDTERM EXAMINATION  
ECON 103, STATISTICS FOR ECONOMISTS

NOVEMBER 2ND, 2016

**You will have 90 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.

**Please sign the back of your blue book.**

Question:	1	2	3	4	5	6	7	Total
Points:	7	15	22	12	30	14	0	100
Bonus Points:	0	0	0	0	0	0	5	5
Score:								

**Instructions:** Answer all questions in your blue book. 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 in the back of your blue book. Make sure that you have all pages of the exam before starting.

**Warning:** If you continue writing after I call time, even if this is only to fill in your name, twenty-five points will be deducted from your final score. In addition, ten points will be deducted for not signing the back of your blue book.

**CHECKLIST BEFORE CONTINUING:**

- ☐ My name and Penn ID number are on my blue book
- ☐ I have signed the academic integrity statement in the back of the blue book

1. **Short Answer** Be sure to fully explain your answer.

(a) [2 points] What is a random variable? Be precise in your definition.

**Solution:** A random variable is a function that maps from the support set (all basic outcomes) of a random experiment to the real line. In math, a random variable  $X : S \rightarrow \mathbb{R}$ .

(b) [5 points] Prove that  $E[aX + b] = aE[X] + b$  when  $X$  is a continuous random variable.

**Solution:**

$$\begin{aligned} E[aX + b] &= \int_{-\infty}^{\infty} (ax + b)f(x)dx \\ &= \int_{-\infty}^{\infty} axf(x)dx + \int_{-\infty}^{\infty} bf(x)dx \\ &= a \int_{-\infty}^{\infty} xf(x)dx + b \int_{-\infty}^{\infty} f(x)dx \\ &= aE[X] + b \end{aligned}$$

[+3 for properly specifying the expectation definition. +2 for properly rearranging]

2. **Joint Distributions** For the following question,  $X$  and  $Y$  have the following joint distribution:

		Y		
		1	2	3
X	1	$\frac{1}{6}$	$\frac{1}{6}$	0
	2	0	$\frac{1}{6}$	$\frac{1}{6}$
	3	$\frac{1}{6}$	0	$\frac{1}{6}$

(a) [6 points] What is the marginal pmf of  $X$ ? What is the marginal pmf of  $Y$ ?

**Solution:** For  $X$ , we can obtain the marginal distribution by summing across the rows. Similarly for  $Y$ , we sum down the columns. Hence, we get the

following:

$$\begin{array}{ll} p_X(X = 1) = \frac{1}{3} & p_Y(Y = 1) = \frac{1}{3} \\ p_X(X = 2) = \frac{1}{3} & p_Y(Y = 2) = \frac{1}{3} \\ p_X(X = 3) = \frac{1}{3} & p_Y(Y = 3) = \frac{1}{3} \end{array}$$

**[+1 for each part. Partial credit for explanation of what a marginal pmf is and how it is derived from a joint distribution]**

(b) [4 points] What is  $E[E[Y|X]]$ ?

**Solution:**

*Method 1:* From the specification of the joint distribution, we can see that conditional on  $X$ , we know that  $Y$  will take one of two values with equal probability, so computing the expectation given  $X$  can be done as follows. If  $X = 1$ , then  $E[Y|X = 1] = 0.5 * 1 + 0.5 * 2 = 1.5$ . If  $X = 2$ , then  $E[Y|X = 2] = 0.5 * 2 + 0.5 * 3 = 2.5$ . If  $X = 3$ , then  $E[Y|X = 3] = 0.5 * 1 + 0.5 * 3 = 2$ . Then, to get the conditional expectation, we simply take these values and multiply them by the probability that  $X$  takes on each of those values. Therefore,  $E[E[Y|X]] = \frac{1}{3} * 1.5 + \frac{1}{3} * 2.5 + \frac{1}{3} * 2 = 2$ .

*Method 2:* Use the Law of Iterated Expectations and recognize that

$$E[E[Y|X]] = E[Y]$$

Computing  $E[Y]$  can be done from the marginal pmf obtained in an earlier part.

$$E[Y] = \frac{1}{3} * 1 + \frac{1}{3} * 2 + \frac{1}{3} * 3 = 2$$

**[+2 for using Law of Iterated Expectations or computing the individual conditional expectations. +2 for appropriately applying the results of either approach]**

(c) [5 points] Are  $X$  and  $Y$  independent? Why or why not?

**Solution:** The definition of independence specifies that two random variables

are independent if

$$p_{XY}(x, y) = p_X(x)p_Y(y)$$

In this problem, we can test that for the outcome  $(1, 2)$ .

$$p_{XY}(1, 2) = \frac{1}{6} \neq \frac{1}{3} * \frac{1}{3} = p_X(X = 1)p_Y(Y = 2)$$

Since  $p_{XY}(x, y) \neq p_X(x)p_Y(y)$ ,  $X$  and  $Y$  cannot be independent.

**+3 for using proper definition. +2 for proper evaluation**

3. **Continuous Random Variables:** Let  $X$  be a random variable with pdf  $f(x) = (x + 2)/4$  and support set  $x \in [-2, 0]$ .

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

**Solution:** Given a pdf, the CDF of a random variable is the integral up to  $x_0$ . Therefore, we can get the CDF as follows:

$$\begin{aligned} F(x_0) &= \int_{-\infty}^{x_0} f(x)dx = \int_{-2}^{x_0} \frac{x+2}{4}dx \\ &= \frac{1}{4} \int_{-2}^{x_0} (x+2)dx \\ &= \frac{1}{4} \left( \frac{x^2}{2} + 2x \right) \Big|_{-2}^{x_0} \\ &= \frac{1}{4} \left( \frac{x_0^2}{2} + 2x_0 - 2 + 4 \right) \\ &= \frac{1}{4} \left( \frac{x_0^2}{2} + 2x_0 + 2 \right) \end{aligned}$$

$$F(x_0) = \begin{cases} 0 & x_0 < -2 \\ \frac{1}{4} \left( \frac{x_0^2}{2} + 2x_0 + 2 \right) & -2 \leq x_0 \leq 0 \\ 1 & 0 < x \end{cases}$$

**[+3 for proper integral. +2 for proper evaluation]**

**Errata:** The pdf is incorrectly specified. The correct pdf should be  $f(x) = (x + 2)/2$  over the same support set. Doing the same method as above, the correct CDF is

$$F(x_0) = \begin{cases} 0 & x_0 < -2 \\ \frac{1}{2} \left( \frac{x_0^2}{2} + 2x_0 + 2 \right) & -2 \leq x_0 \leq 0 \\ 1 & 0 < x \end{cases}$$

(b) [5 points] Calculate  $P(X > -1)$ .

**Solution:** This is the chance that our random variable takes on values between -1 and 0. This corresponds to the probability that our random value does not take values less than -1. Hence, using what we found in part (a), we can compute this as follows:

$$\begin{aligned} P(X > -1) &= 1 - P(X \leq -1) = 1 - \frac{1}{4} \left( \frac{1}{2} - 2 + 2 \right) \\ &= 1 - \frac{1}{4} * \frac{1}{2} = \frac{7}{8} \end{aligned}$$

The same answer could be reached by taking the integral of the pdf from -1 to 0 or by using geometric properties of the pdf.

**Errata:** With the properly specified pdf,  $P(X > -1) = 0.75$

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

**Solution:** Using the definition of expectation, this can be computed as follows:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-2}^0 x \frac{x+2}{4} dx \\ &= \frac{1}{4} \int_{-2}^0 (x^2 + 2x) dx \\ &= \frac{1}{4} \left( \frac{x^3}{3} + x^2 \right) \Big|_{-2}^0 \\ &= \frac{1}{4} \left( 0 + 0 + \frac{8}{3} - 4 \right) \\ &= -\frac{1}{3} \end{aligned}$$

[+3 for proper integral. +2 for proper evaluation]

**Errata:** With the proper pdf,  $E[X] = -\frac{2}{3}$

(d) [7 points] Calculate  $Var(X)$ .

**Solution:** Using the shortcut formula, we know that  $Var(X) = E[X^2] - (E[X])^2$  [+2]. First, we compute  $E[X^2]$ :

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-2}^0 x^2 \frac{x+2}{4} dx \\ &= \frac{1}{4} \int_{-2}^0 (x^3 + 2x^2) dx \\ &= \frac{1}{4} \left( \frac{x^4}{4} + \frac{2x^3}{3} \right) \Big|_{-2}^0 \\ &= \frac{1}{4} \left( 0 + 0 - 4 + \frac{16}{3} \right) \\ &= \frac{1}{3} \end{aligned}$$

[+3]

Combining this with part (c), we get the following:

$$Var(X) = \frac{1}{3} - \left( \frac{1}{3} \right)^2 = \frac{2}{9}$$

[+2]

**Errata:** With the proper pdf,  $E[X^2] = \frac{2}{3}$  and  $Var(X) = \frac{2}{9}$

4. **Distributions** Be sure to fully specify the following distributions. Let  $X_1$ ,  $X_2$ , and  $X_3 \sim iidN(0, 1)$ .

(a) [3 points] What is the distribution of  $X_1 + X_2$ ?

**Solution:** This is the sum of two normals, which is also normal. Hence we need to find the mean and variance. Since expectation is linear, the new mean will be the sum of the old means, which is 0. Furthermore, since  $X_1$  and  $X_2$  are iid, then the new variance will also just be the sum of the two variances, which is 2. Therefore, the final distribution is  $N(0, 2)$

(b) [3 points] What is the distribution of  $X_1^2 + X_2^2 + X_3^2$ ?

**Solution:** This is the sum of squares of standard normals, which is a chi-squared distribution. The chi-squared distribution only has one parameter, which specifies the number of squared standard normals that are added together. In this case, there are 3 that are being added together. Therefore, the full specification of this is  $\chi^2(3)$

(c) [3 points] What is the distribution of

$$\frac{(X_1^2 + X_2^2)/2}{(X_2^2 + X_3^2)/2} ?$$

**Solution:** This is the quotient of two chi-squared random variables, which takes an F-distribution. the F-distribution has two parameters, which depend on the degrees of freedom of the numerator and the denominator. In this case, the degrees of freedom of both are 2, so the distribution is  $F(2, 2)$ .

(d) [3 points] What is the distribution of

$$\frac{X_1}{\sqrt{(X_1^2 + X_2^2 + X_3^2)/3}} ?$$

**Solution:** This is the quotient of a standard normal and the square root of a chi-squared. This gives a t-distribution. The t-distribution has only one parameter, which depends upon the degrees of freedom of the chi-squared variable in the denominator. In this case, the chi-squared has 3 degrees of freedom so the full specification of this is  $t(3)$

5. **Estimators** Let  $X_1, \dots, X_{10}$  be iid draws from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Joe is trying to estimate  $\mu$  and he proposed two estimators. His first estimator is  $\hat{\mu}_1 = \frac{1}{n} \sum_i X_i$ . His second estimator is  $\hat{\mu}_2 = X_7$ , where he reports the 7th observation as his estimate of the sample mean.

(a) [3 points] What does it mean for an estimator to be unbiased? Be precise in your definition.

**Solution:** Given an estimator  $\hat{\theta}$  that is trying to estimate some true parameter  $\theta_0$ , its bias can be computed using the following:  $Bias = E[\hat{\theta}] - \theta_0$  and  $\hat{\theta}$  will be unbiased if the bias is 0.

- (b) [3 points] What does efficiency mean? Be precise in your definition.

**Solution:** Given two **unbiased** estimators,  $\hat{\theta}_1$  and  $\hat{\theta}_2$ , of a parameter  $\theta_0$ ,  $\hat{\theta}_1$  is said to be *more efficient* if  $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ . **[-1 for not specifying unbiased estimators]**

- (c) [2 points] What is mean-squared error (MSE) of an estimator? Be precise in your definition.

**Solution:** Given an estimator  $\hat{\theta}$ , the mean-squared error of the estimator is defined as  $MSE(\hat{\theta}) = Bias(\hat{\theta})^2 + Variance(\hat{\theta})$

- (d) [3 points] What does MSE consistency mean? Be precise in your definition. (**Note:** this is the only kind of consistency we have talked about in class)

**Solution:** An estimator  $\hat{\theta}_n$  is consistent if  $\lim_{n \rightarrow \infty} MSE(\hat{\theta}_n) = 0$ ,

- (e) [5 points] Is  $\hat{\mu}_1$  unbiased? Show your work

**Solution:** Evaluating the expectation of this estimator gives us the following:

$$\begin{aligned} E[\hat{\mu}_1] &= E\left[\frac{1}{n} \sum_i X_i\right] \\ &= \frac{1}{n} E\left[\sum_i X_i\right] \quad \text{by linearity of expectation} \\ &= \frac{1}{n} \sum_i E[X_i] \quad \text{by linearity of expectation} \\ &= \frac{1}{n} \sum_i \mu = \frac{1}{n} n\mu = \mu \end{aligned}$$

**[+3]**



Now we can evaluate the bias of the estimator:

$$Bias = E[\hat{\mu}_1] - \mu = \mu - \mu = 0$$

Therefore our estimator is unbiased. **[+2]**

- (f) [3 points] Is  $\hat{\mu}_2$  unbiased? Show your work.

**Solution:** Evaluating the bias of our estimator gives the following:

$$Bias = E[\hat{\mu}_2] - \mu = E[X_7] - \mu = \mu - \mu = 0$$

Since the bias is zero, this estimator is unbiased. **[+3]**

- (g) [8 points] Which estimator is more efficient? Clearly explain how you reached your conclusion. Show your work.

**Solution:** First, in parts (e) and (f), we established that both estimators are unbiased, so the question of efficiency is well-specified. Recall that efficiency requires both estimators to be unbiased. The next step is to see which one has a lower variance, so we must compute their variances. The variance of the first estimator can be computed as follows:

$$\begin{aligned} Var(\hat{\mu}_1) &= Var\left(\frac{1}{n} \sum_i X_i\right) \\ &= \frac{1}{n^2} Var\left(\sum_i X_i\right) \quad \text{by properties of variance} \\ &= \frac{1}{n^2} \sum_i Var(X_i) \quad \text{because the draws are iid} \\ &= \frac{1}{n^2} \sum_i \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} = \frac{\sigma^2}{10} \end{aligned}$$

**[+4]**

The variance of the second estimator can be computed as follows:

$$Var(\hat{\mu}_2) = Var(X_7) = \sigma^2$$

[+2]

Since  $\frac{\sigma^2}{10} < \sigma^2$ , then  $\hat{\mu}_1$  is more efficient than  $\hat{\mu}_2$  [+2]

(h) [3 points] Which estimator (if any) is consistent?

**Solution:** Consistency is an asymptotic property and it requires that the MSE goes to 0 as the population gets larger. In this case, both estimators are unbiased, so their MSE is simply their variances. Taking limits of the variances, we get:

$$\lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

$$\lim_{n \rightarrow \infty} \sigma^2 = \sigma^2 \neq 0$$

Therefore, only  $\hat{\mu}_1$  is consistent.

[+3]

6. **Confidence Intervals** Jane is trying to estimate the average height of American men. She has collected data on the heights of a random sample of 100 men. The average height of her sample is 70 inches. For this question, assume heights are normally distributed with mean  $\mu$  and variance  $\sigma^2$ .

(a) [3 points] Jane knows the true population variance,  $\sigma^2$ , is 25 inches. What is the 95% confidence interval for the population mean  $\mu$ ?

**Solution:** The 95% confidence interval will be  $70 \pm 2 * \frac{5}{10} = [69, 71]$ . This can be alternatively stated as  $P(69 \leq \mu \leq 71) = 0.95$

[+3]

(b) [3 points] Jane actually wants to find the 98% confidence interval for the population mean  $\mu$ . Once again, she knows the true population variance,  $\sigma^2$ , is 25 inches. What is the confidence interval? Use R commands where necessary.

**Solution:** In this case, the 98% confidence interval will be  $70 \pm \text{qnorm}(0.99) * \frac{5}{10}$

[+3]. -1 for not using standard normal.

- (c) [3 points] Jane actually does not know the population variance  $\sigma^2$ , but she knows the sample variance  $S^2$  is 25 inches and she wants to construct the 95% confidence interval for the population mean  $\mu$ . What is the 95% confidence interval? Use R commands where necessary.

**Solution:** Since the population variance is unknown, Jane must use a different distribution to compute the confidence. The distribution that she must use is the t-distribution. The computation of the confidence interval follows similarly with that modification. The confidence interval will be  $70 \pm \text{qt}(0.975, \text{df} = 99) * \frac{5}{10}$

- (d) [5 points] Why was Jane's confidence interval different between parts (a) and (c)? There is no need to re-derive any results from the lecture slides, but you should write out the relevant expressions for confidence intervals and discuss what changed and how that changed the result. (**Note:** If you got the same answer for (a) and (c), you might want to double check your work)

**Solution:** The critical difference between parts (a) and (c) was that in part (a), Jane knew the true population variance ( $\sigma^2$ ) while in part (c), she did not and had to use an estimator of the population variance ( $S^2$ ). In part (a), the confidence interval rested on the property that

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

[+2]

However, in part (c), this expression has to be modified to be

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n - 1)$$

because the estimator  $S$  is a random variable instead of a constant, as it was in part (a).

[+2]

While tedious, the above expression can be rearranged into the form:

$$\frac{Z}{\sqrt{\chi^2(n-1)/(n-1)}}$$

which takes a  $t(n - 1)$  distribution.

[+1]

7. [5 points (bonus)] **Bonus Question: PDFs and PMFs** Compare and contrast a probability mass function (pmf) and a probability density function (pdf). Be sure to reference their particular properties where possible. Remember, “compare” means to discuss similarities while “contrast” means to discuss differences.

**Solution:** A probability mass function (pmf) is related to discrete random variables. It is a function that gives the probability of a particular outcome occurring. Specifically,  $p : S \rightarrow [0, 1]$ . A pmf must satisfy 2 properties:

- $\sum_x p(x) = 1$
- $0 \leq p(x) \leq 1$  for all  $x \in S$  where  $S$  is the support of the random variable

[+2]

A probability density function (pdf) is related to continuous random variables. Similar to a pmf, it can be used to compute probabilities, but the crucial difference is that its values are not the probabilities of particular events occurring since for continuous random variables, each outcome has a zero probability of occurring.

Instead, the integral of the pdf over intervals gives the probability of the random variable falling within those intervals. A pdf must satisfy the following properties:

- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $f(x) \geq 0$  for all  $x$

[+2]

The first property is analagous to the discrete case. However, the second property is more relaxed than the second property in the discrete case. This is because the pdf does not describe probabilities of points. Instead, it describes probabilities over intervals (by taking integrals over those areas). As a result, the pdf does not necessarily have to be less than 1. [+1]