

SECOND MIDTERM EXAMINATION
ECON 103, STATISTICS FOR ECONOMISTS
MARCH 22ND, 2016

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	35	30	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 and Z be discrete RVs with joint pmf

		Z		
		-1	0	1
X	0	1/6	1/3	0
	1	0	1/6	1/3

- 5 (a) What kind of RV is X ? Specify any and all parameters of its distribution.

Solution: The marginal distribution of X is $p_X(0) = 1/2$, $p_X(1) = 1/2$ so $X \sim \text{Bernoulli}(1/2)$.

- 5 (b) Calculate $E[X]$.

Solution: The expectation is of a Bernoulli is p , in this case $1/2$.

- 5 (c) Calculate the conditional pmf of Z given that $X = 0$.

Solution: The marginal probability that $X = 0$ is $1/6 + 1/3 = 1/2$. Hence, dividing the joint probabilities by this quantity:

$$p_{Z|X}(-1|0) = 1/3, \quad p_{Z|X}(0|0) = 2/3, \quad p_{Z|X}(1|0) = 0$$

- 10 (d) Calculate $E[XZ]$.

Solution: The only non-zero element in the sum that we use to calculate $E[XZ]$ is the one corresponding to $X = 1, Z = 1$ so $E[XZ] = 1 \times 1 \times 1/3 = 1/3$.

2. Let $X_1, \dots, X_5 \sim \text{iid}$ with pdf $f(x) = (x+1)/2$ and support set $x \in [-1, 1]$.

- 5 (a) Calculate the CDF $F(x_0)$ of X_1 .

Solution: For $x_0 \in (-1, 1)$ we have

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

For $x < -1$, on the other hand, $F(x_0) = 0$ and for $x_0 > 1$ it equals 1.

- 5 (b) Calculate $P(X_1 > 0)$.

Solution: $P(X > 0) = 1 - F(0) = 1 - \frac{1}{4} (0^2 + 2 \times 0 + 1) = 3/4$

- 10 (c) What is the probability that *exactly* four of X_1, \dots, X_5 are greater than zero?

Solution: Let Z denote the number of random variables among X_1, X_2, X_3, X_4 that take on a positive value. E.g. if all of them are positive then Z is 4 and if none are positive then Z is zero. Since X_1, X_2, X_3, X_4 are iid and $P(X_1 > 0) = 3/4$, it follows that $Z \sim \text{Binomial}(4, 3/4)$ and we need to calculate $p(4)$ where $p(z) = \binom{n}{z} p^z (1-p)^{n-z}$. We have:

$$p(4) = \binom{5}{4} (3/4)^4 (1/4)^1 = \frac{5!}{4!1!} (3/4)^4 1/4 = 5/4 \times 81/256 \approx 0.40$$

This question is based on 7-17 from the textbook, which I assigned as homework. I have re-worded the problem for clarity but the substance and solution are unchanged.

3. A machine produces perfectly square metal plates. Occasionally, however, this machine malfunctions and the squares it produces are of the wrong area. To make sure the plates are the correct size, we measure the length of each one *twice*. Our measuring device is slightly imperfect: if the true length of a particular square is ℓ inches, our measurement will equal $\ell + Z$ where Z is a random error with mean zero and standard deviation 0.1 inches. Fortunately these errors are independent across measurements: when we measure a square twice our first measurement will be $X_1 = \ell + Z_1$ and our second will be $X_2 = \ell + Z_2$ where $Z_1, Z_2 \sim \text{iid}$ ($\mu = 0, \sigma = 0.1$). We want to find a way of combining X_1 and X_2 to estimate the *area* ℓ^2 of a given plate. Three estimators have been proposed: A, B , and C . Calculate the bias of each using the shortcut formula.

- 10 (a) $A = \frac{X_1^2 + X_2^2}{2}$

Solution:

$$\begin{aligned} E[A] &= \frac{1}{2} [E(X_1^2) + E(X_2^2)] \\ &= E(X_1^2) = \text{Var}(X_1) + E(X_1)^2 = 0.01 + \ell^2 \end{aligned}$$

by the shortcut formula for variance, using the fact that X_1, X_2 are identically distributed, so the bias is 0.01 square inches.

- 10 (b) $B = X_1 X_2$

Solution:

$$\begin{aligned} E[B] &= E[X_1 X_2] = \text{Cov}(X_1, X_2) + E(X_1)E(X_2) \\ &= 0 + \ell \times \ell = \ell^2 \end{aligned}$$

by the shortcut formula for covariance, using the fact that X_1 and X_2 are independent hence uncorrelated. Therefore, B is an unbiased estimator.

10 (c) $C = \left(\frac{X_1 + X_2}{2} \right)^2$

Solution: Using our calculations for $E[X_1^2]$ and $E[X_1 X_2]$ from above,

$$\begin{aligned} E[C] &= \frac{1}{4} E[X_1^2 + 2X_1 X_2 + X_2^2] = \frac{1}{4} [2E(X_1^2) + 2E(X_1 X_2)] \\ &= \frac{1}{2} [(0.01 + \ell^2) + \ell^2] = 0.005 + \ell^2 \end{aligned}$$

Therefore the bias of C is 0.005 square inches.

4. This question uses simulations to study what happens when we sample *without* replacement from a population with 100 individuals: 40 of whom are “ones” and the rest of whom are “zeros.” The point is to explore the bias and variance of this procedure and compare it to the random sampling procedure (*with* replacement) we studied in class.

- 5 (a) Let `popn` be an R vector with 100 elements, the first 40 of which are 1 and the rest of which are 0. Write an R command that makes 50 random draws from `popn` *without replacement*.

Solution:

```
sample(popn, size = 50, replace = FALSE)
```

- 15 (b) Build your command from the preceding part into a function called `draw.mean` that draws a sample of size `n` from `popn` and returns the sample mean of the draws. Your function should take a single input argument: `n` the number of draws to make.

Solution:

```
draw.mean <- function(n){
  draws <- sample(popn, size = n, replace = FALSE)
  return(mean(draws))
}
```

}

- 5 (c) Write R code that calls the function `draw.mean` 10000 times with `n` equal to 50 and stores the result in a vector called `sims`.

Solution:

```
sims <- replicate(10000, draw.mean(50))
```

- 5 (d) After running the code described in the preceding parts I entered `mean(sims)` at the R console and got 0.399912. Based on this result, does it appear that the sample mean remains an unbiased estimator of the population mean when calculated from a sample drawn *without replacement*? Explain briefly.

Solution: Sampling once from `popn` is equivalent to making one `Bernoulli(0.4)` draw. Hence the population mean is 0.4 which is extremely close to the mean of the sampling distribution that we calculated by simulation. The sample mean appears to remain unbiased even when sampling without replacement. Or if there is a bias, it is extremely tiny.

- 5 (e) I then entered `sd(sims)` into the R console and got 0.04946187. Based on this result, how does does sampling *without replacement* compare to sampling *with replacement* in terms of the standard error of the sample mean? Explain briefly

Solution: As explained in the solution to the preceding part, this population is essentially a `Bernoulli(0.4)` RV. Hence the population standard deviation is $\sqrt{0.4 * 0.6} \approx 0.49$. When sampling *with replacement* (a random sample) we know that the standard error of the mean is $\sigma/\sqrt{n} = \sqrt{(0.4 \times 0.6)/50} \approx 0.07$. This is *much larger* than the standard error of the mean for the simulations made without replacement. It appears that sampling without replacement reduces sampling error compared to sampling with replacement (random sampling), at least when drawing a fairly large sample from a fairly small population.

5. Let $X_1, \dots, X_{25} \sim \text{iid } N(\mu, \sigma^2 = 4)$ and define $Z = 5(\bar{X}_n - \mu)/S$ and $Y = 6S^2$. When asked for a sampling distribution in this question, be sure to name the distribution and specify the values of any and all relevant parameters.

- 10 (a) Suppose our observed dataset x_1, \dots, x_{25} has a sample mean of $\bar{x} = 2.3$. Construct a 68% confidence interval for μ using the fact that the population variance is known

to be 4. Your answer should *not* include any R commands.

Solution: The standard error is $\sigma/\sqrt{n} = 2/5 = 0.4$ so the confidence interval is 2.3 ± 0.4 or equivalently $(1.9, 2.7)$.

- 5 (b) What is the sampling distribution of Z ? No explanation is needed.

Solution: $t(24)$

- 5 (c) Write down the R command to calculate $P(Z \leq 1)$.

Solution: `pt(1, df = 24)`

- 5 (d) What is the sampling distribution of Y ? No explanation is needed.

Solution: $\chi^2(24)$

- 5 (e) Write down the R command to calculate 90th percentile of Y .

Solution: `qchisq(0.9, df = 24)`