Exam 2

EC 303: Empirical Economic Analysis

November 15, 2019

This exam is worth **100 points**, and is timed to take 45 minutes (you have **50 minutes** to complete it). You may use a one-sided, hand-written formula sheet (letter size) and a simple calculator. You **may not** use your cell phone or any other electronic device. If necessary, round your final answer (not the numbers in intermediate steps) to the third decimal place.

Please keep answers neat and organized. If you must separate parts of an answer, indicate this in the blue book so that a grader can clearly find your work. Any suspected academic misconduct will be reported to the Deans office.

1 True/False Questions

(6 minutes). Answer each question true (T) or false (F) with no more than 3 sentences justifying your answers.[1] This section is worth a total of 18 points.

1. (3 pts) $\text{Corr}(X-1,Y)=\text{Corr}(X,Y)$.
2. (3 pts) If X, Y, and Z are random variables, then
$$\mathbb{V}\left[2X-6Y-Z\right]=4\mathbb{V}[X]+36\mathbb{V}[Y]-\mathbb{V}[Z]-24\text{Cov}(X,Y)-4\text{Cov}(X,Z)+12\text{Cov}(Y,Z).$$
3. (3 pts) If $X\sim F(30,20)$ and $Y\sim t(25)$, then the 10th percentile of X is greater than the 5th percentile of Y.
4. (3 pts) If $\hat{\theta}_1$ is an unbiased estimator of θ and is more efficient than a consistent estimator $\hat{\theta}_2$, $\hat{\theta}_1$ is also consistent.
5. (3 pts) The statement, "the 95% confidence interval for the population mean is $(100,125)$", is equivalent to the statement, "there is a 95% probability that the population mean is between 100 and 125".
6. (3 pts) If the probability that $\overline{x}_n$ differs from μ by more than ϵ converges to 0 as $n\to 0$ for any $\epsilon>0$, then
$$\lim_{n\to\infty}\mathbb{E}[(\overline{x}_n-\mu)^2]=0.$$

2 Short Problems

(24 minutes). This section contains 4 questions and is worth a total of 50 points.

Problem 1. Education and Work Time. We are interested in the relationship between education and hours worked, so we interview 1,000 people about their work habits and training. Suppose that when we bin responses, we are left with the following results:

[1] Justifications may take the form of counterexamples, proofs, or intuition.

Education:	Hours Worked/Week < 30	30–39	40–49	50–59	60+	*Marginal:*
Did not Graduate High School	12	--	63	11	8	0.110
High School Diploma	29	110	--	17	9	0.200
Bachelor's Degree	60	212	--	58	27	0.400
Master's Degree	--	57	61	21	9	0.150
Ph.D. or Higher	13	54	59	--	10	0.140

a. (3 pts) Using the marginal probabilities of education status, complete the table and turn it into a valid pmf. That is, find $p(x, y)$ for each value x of education and y of hours worked/week.

b. (2 pts) To your table, add the marginal probabilities of weekly hours worked. Verify that your marginal probabilities sum to 1.

c. (3 pts) Calculate the expected value of weekly hours worked $\mathbb{E}[y]$. For the value of each column, use the lowest reported value of hours worked. Is this higher or lower than what you expected? Why do you think that is?

d. (4 pts) Find and interpret the correlation between the two random variables.

   - Assign a value of 1 through 5 to each of the education values (1 for the first row, 5 for Ph.D., etc.) to calculate this.
   - You may assume that $\mathbb{E}[XY] \approx 100.87$, $\sigma_X = 2.7$ and $\sigma_Y = 36$ to shorten calculations.

e. (3 pts) Are the two variables independent?

Problem 2: Confidence Intervals.. An entrepreneur wants to provide pet care while her clients are on vacation. As part of her market research, she interviews 80 people, chosen randomly from among the pet owners in her community, to ask whether they travel frequently. Twenty-eight respondents say that they travel frequently, while the other 52 say they do not travel frequently.

- (10 pts) Construct a two-sided 95% confidence interval for the fraction of pet owners that travel frequently.

Problem 3: Maximum Likelihood Estimation. Suppose that an unknown fraction p of jobs offer health insurance. In a random sample of n jobs, the number y that offer health insurance can be viewed as a single draw from a binomial distribution, with parameters n and p.

a. (10 pts) Find an estimator for p, using the method of maximum likelihood.

b. (5 pts) Interpret your estimator, and show that you've found a maximum of the likelihood function.

   - *Hint:* Use C_y^n as shorthand for the number of ways y objects can be chosen from n, rather than worrying about expressing the full formula. Also note our sample size here.

Problem 4: Sampling Distributions. We are the leaders of a developing country attempting to reign in corruption, so we are conducting an audit of precinct leaders. For each leader, we count the number of ways they fail an audit, which has the following pmf:

x	0	1	2	3+
$p(x)$	0.3	0.4	0.3	0.00

a. (5 pts) Suppose that we sample 2 officials. List each possible outcome, its probability, and the corresponding sample mean $\overline{x}$ and sample variance s^2.

b. (3 pts) List the probability distributions for the sample mean and variance. Are the distributions symmetric?

c. (2 pts) Verify that the expected value of the sampling distribution is equal to the population mean.

3 Proofs

(16 minutes). This section contains 2 questions and is worth a total of 30 points.

Problem 5: Method of Moments Evaluation. Suppose that $\{X_1, X_2, ..., X_n\}$ is an i.i.d. sample drawn from a uniform distribution over the interval $[a, 0]$, for some $a < 0$.

a. (3 pts) Find an estimator of a using the method of moments.

b. (5 pts) Prove that your estimator is unbiased.

c. (5 pts) Prove that your estimator is consistent.

d. (2 pts) What is the MSE of your estimator?

e. (3 pts) Consider the alternative estimator:

$$\tilde{a} = \frac{n+1}{n} \min\{X_1, ..., X_n\}.$$

Using the fact that $\mathbb{V}[\min\{X_1, ..., X_n\}] = \frac{a^2 n}{(n+1)^2(n+2)}$, can you determine which of these estimators has minimized the MSE? Does your answer depend on n?

- You may assume that $\tilde{a}$ is unbiased as well.

Problem 6: Factoring Theorem. Suppose that X and Y are two *continuous* random variables with a joint probability distribution $f(x, y)$ that can be written as

$$f(x, y) = u(x)v(y)$$

Prove that:

a. (5 pts) There exists a constant $c > 0$ such that X and Y have marginal distributions given by

$$g(x) = cu(x)$$
$$h(y) = \frac{1}{c}v(y).$$

b (5 pts) The variables X and Y are independent.