Homework 3
PP 312: Advanced Statistics for Data Analysis I
Instructors: Black, Chen, Chhikara, Matz, and Wyse
Fall 2021

Instructions: Answer all the questions. Show your work for any calculations that you do. For your work in R, please upload the .rmd file and a knitted PDF. For your work in Stata, please submit a .do file and a PDF of a log file. Your computer code should include clear and concise comments. **This assignment is due Friday, October 22 at 11:59pm.**

1. Suppose you have the following pdf:

$$f(x) = \lambda e^{-\lambda x} \qquad \text{for } \ 0 \leq x < \infty \tag{1}$$

A. Find the cdf.

B. Recognizing that $F(x) = u$, where u where u is a uniform random variable on the unit interval, find the inverse of the cdf $x = F^{-1}(u)$.

For problems 2-5, consider the following distribution:

$f(x, y)$

	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$f_x(x)$
$x = 1$	0.15	0.05	0.05	0.05	0.30
$x = 2$	0.05	0.05	0.05	0.25	0.40
$x = 3$	0.05	0.05	0.10	0.10	0.30
$f_y(y)$	0.25	0.15	0.20	0.40	1

2. Calculate $\mathrm{E}(x)$, $\mathrm{Var}(x)$, $\mathrm{E}(y)$, $\mathrm{Var}(y)$, and $\mathrm{Cov}(y, x)$.

3. Suppose $x = 2$. Calculate $\mathrm{E}(y|x = 2)$ and $\mathrm{Var}(y|x = 2)$.

4. Now suppose $y < 3$. Calculate $\mathrm{E}(x|y < 3)$ and $\mathrm{Var}(x|y < 3)$.

5. Simulate the above distribution for a sample of $n = 50$ and $n = 500$ in both R and Stata. Be sure to set your seed for the random number generator so you may replicate your results. For each data set, calculate the empirical counterparts to the moments in problem 2. Which data set does better?

6. Suppose $x \in \{1, 2, 3\}$ and $y \in \{1, 2, 3\}$. Construct a joint distribution where (y, x) are independent.

7. A patient takes a Covid-19 test, which can return a positive result or a negative result. A patient with the disease has an 80% probability of testing positive and a 20% probability of testing negative. (The 80% is called the sensitivity of a test.) A patient without the disease has a 2% probability of testing positive and a 98% probability of testing negative. (The 98% is called the specificity of a test.) Suppose 3% of the population has the disease.

A. For a patient that tests negative, what is the probability of having the disease?

B. For a patient that tests positive, what is the probability of having the disease?