**Econ 140** 

Summer 2022

Instructor: Fernando Hoces de la Guardia

GSIs: Elena Stacy & Yige Wang

Midterm Exam 1

Thursday July 7, 2022

## **Exam Instructions:**

- You have 80 minutes to answer this exam
- This exams has a total of 80 points (suggesting the length of time to spent in each question). Each question indicates the number of points, and indicates a maximum length for its answer
- Most questions ask for short answers (from a couple of words, to one or two sentence maximum)
- Explanation in black or blue ink is recommended as these often scan the best.
- You must submit your solutions using the exam packet provided.
- Do not write your solutions on pages that say "Do not write solutions on this page". Answers written on these pages will not be graded. You may use these pages as scratch paper.
- When time is called, STOP writing, immediately CLOSE your exam packet and hold it up until it is collected.
- Show your work. Credit will only be awarded on the basis of what is written on the exam.
- Sign the academic honesty pledge. Cheating will be punished.

Student ID Number:
Affirm the academic honesty pledge below. For those writing on a non-printed copy,
please just write "Academic Honesty Pledge as on exam", and sign your name.
If you do not affirm this pledge, your exam will be marked invalid.
0. ACADEMIC HONESTY PLEDGE
I confirm that I have abided by all academic honesty rules for UC Berkeley and Economics 140.
I confirm that I did not see this exam before my official exam start time. I confirm that I have
not shared and will not share this exam with anyone else. I confirm that I haven't copied from
anybody else's exam.
Signature:

Student Name:

1. Based on our class discussion of the concept of bullshit: what is the key characteristic that distinguishes bullshit from lies (or the bullshitier from the liar) [2pt, 1-2 sentences].

2. How can we compute the probability of an event for the case of a continuous random variable? [2pt, 1 sentence]

3. For the case of a continuous random variable: which distribution is a good representation of not knowing much about a phenomenon (represented by the random variable)? How about for a discrete random variable (assume 2 events for simplicity)? [2pt, 2 sentences]

4. The probability that two random variables, X, Y take the values of x, and y can always be described as: [2pt]

a) 
$$Pr(X = x) + (Y = y)$$

b) 
$$Pr(X = x, Y = y)$$
.

c) 
$$Pr(X = x) * (Y = y)$$

d) 
$$Pr(X = x)/(Y = y)$$

5. Using the concept of conditional probabilities, the same term as in the question above (the probability that two random variables, X, Y take the values of x, and y) can be described as. [2pt]

a) 
$$Pr(X = x | Y = y)^* Pr(X = x)$$

b) 
$$Pr(Y = y | X = x)^* Pr(Y = y)$$

- c)  $Pr(X = x \mid Y = y)^* Pr(Y = y)$
- d)  $Pr(Y = y \mid X = X)$
- 6. Define the concept of independence in plain English [4pt, 1-2 sentences]

7. Define the concept of independence in terms of conditional probabilities. [2pt, 1-2 sentences equation]

8. In which sense is the variance, an average. What is an average of? [2pt, 1 sentence]

9. Compute the variance for a variable with values 2,3, and 4. Show your calculations. [4pts, 2-5 lines]

10.	An econometrics class has 75 students, and the mean student weight is 155 lb. A random sample of four students is selected from the class, and their average weight is calculated.
	(a) Will the average weight of the students in the sample equal 155 lb? [1pt, Yes/No]
	(b) Explain why you answered yes or no to the above question. [1pt, 1 sentence]
	(c) Is the sample average, $\overline{Y}$ , a random variable? [1pt, 1 sentence]
11.	In a rectangular data set: what is represented by rows and what by columns. [1pt, 1 sentence]
12.	Assume that you have received a data set with millions of observations on the income of all the American households. Which concept(s) discussed in class can you use to communicate some initial insights about these millions of observations? [2pt, 1 sentence]
13.	What is the interpretation of the mean of a binary variable? [2pt, 1 sentence]

14.	What is the main reason we prefer to report (and read) standard deviations instead of variances? [2pt, 1 sentence]
15.	Given a collection of random variables, Y1, Y2,, Yn let $\overline{Y}$ represent its sample mean. What happens with its expected value as the number of random (n) variables increases. What happens to its variance? [2pt, 1-2 sentences]
16.	Given a data set with information on the age of each individual in the US (360 million), choose which of the following could be a <b>plausible</b> value for the standard deviation <b>of the sample mean</b> . [2pt]  (a) -32 (b) 32 (c) 320 (d) 0.32
17.	Explain your answer to the above question. [2pt, 1 sentence]
18.	For the simulation performed in class and sections to explore the Central Limit Theorem:  (a) What happens with the distribution of the sample mean as we increase the sample size? [4pt, 1-2 sentences]

(b)	What happens with the distribution as	s we increase	the number	of simulations, $\frac{1}{2}$	draws?
	[2pt, 1 sentence]				

19. Name what is the key assumption that is violated when we observe a spurious correlation. [3pt, 1 sentence]

20. Let's look again at the table used to explain the intuition behind conditional probabilities. Add two new variables "Pass|S=1" and "Pass & S" and fill-in the corresponding value of each observation [2 columns, 6pts].

i	Pass	S
1	1	1
2	1	0
3	1	1
4	0	0
5	1	1
6	0	1

Figure 1:

21. (Modified "Conditionities"). Annie lives in a town of 3000 individuals where a disease called "Conditionities" is scaring the population. The disease affects 2% of the population and there is a test that detects the disease with 90% success (meaning that, when a person has the disease, the tests comes positive 90% of the time; and when the person does not have the disease, it comes negative 90% of the time). Annie just took a test and it came up positive, what is the probability that the has the disease? [6pt, 2-6 lines]

[Suggestion: getting stuck in this question might consume a significant portion of your time, if the answer doesn't come to you immediately, leave it for the end]

[Help, in addition to Bayes formula printed on question 27, you might also need the law of total probabilities. This states that for two random variables X and Y, the probability of X taking a specific value can be broken down into the following conditional probabilities:]

$$P(X = x) = \sum_{i} P(X = x | Y = y_i) P(Y = y_i)$$
 (1)

22. Describe the intuition behind balance in pre-treatment variables. [4pt, 1-3 sentences]

23. Provide a one-line interpretation for all the numbers highlighted in the 2rd row of the following table (one line per box)? [8pt, 3 sentences]

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
A. Health-care use					
Face-to-face visits	2.78	.19	.48	1.66	.90
	[5.50]	(.25)	(.24)	(.25)	(.20)
Outpatient expenses	248	42	60	169	101
	[488]	(21)	(21)	(20)	(17)
Hospital admissions	.099	.016	.002	.029	.017
	[.379]	(.011)	(.011)	(.010)	(.009)
Inpatient expenses	388	72	93	116	97
	[2,308]	(69)	(73)	(60)	(53)
Total expenses	636	114	152	285	198
	[2,535]	(79)	(85)	(72)	(63)

Figure 2:

24. For the exercise of potential outcomes done in class with a fictional data of 10 individuals, cross with an x the potential outcomes that are missing from the data. [6pts, write x's on table]

i	$Y^1$	$Y^0$	D	$\kappa$
1	5	2	1	3
2	1	4	0	-3
3	3	1	1	2
4	2	1	1	1
5	4	4	0	0
6	5	4	1	1
7	1	2	1	-1
8	2	3	0	-1
9	4	1	0	3
10	3	1	0	2

Figure 3:

25. 
$$\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y \mid X)) + \mathbb{V}(\mathbb{E}(Y \mid X))$$

Ev(v)es Law, the expression above, tells us that the unconditional variance will always be [3pt]:

- (a) Larger or equal than the conditional variance
- (b) Smaller or equal than the conditional variance
- (c) Larger when the Y is independent of X
- (d) Smaller when Y is independent of X

26. [Bonus Question. Suggestion: focus on this question only after you are done with the rest of the exam]

Monty Hall Problem (slightly modified): Suppose you're on a game show, and you're given the choice of **four** doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, **who knows what's behind the doors**, opens another door, say No. 4, which has a goat. He then says to you, "Do you want to pick door No. 4?" Is it to your advantage to switch your choice?

Explain the solution to the Monty Hall problem in plain English. [3pts, 1-3 sentences]

27. [Bonus] Derive the solution to the Monty Hall problem using Bayes rule [3pts, 2-6 lines]. Reminder of Bayes rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

[Do not write on this page.]

[Do not write on this page.]