Spring 2021 Midterm II SOLUTIONS

March 19, 2021

The exam is open book. This means you can use electronic or hard copies of all class materials and can use datahub if you wish. You may not use the internet to search for the answers or to inform your answers. Using the internet is strictly prohibited and any evidence of this may result in a 0 on the exam.

While you take the exam, you are prohibited from discussing the test with anyone. If you are taking the test after your classmates, you are also prohibited from talking to them about the test before you take it. Evidence of cheating may result in a 0 on the exam and be reported to the Student Conduct Board.

Berkeley's code of conduct is here: https://sa.berkeley.edu/code-of-conduct. See Section V and Appendix II for information about how UC Berkeley defines academic misconduct. In particular note the sections on cheating and plagiarism.

UC Berkeley Honor Code "As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others." Please carefully read the statements below, and indicate your understanding and intent to adhere to the UC Berkeley Honor code by typing your name in the space below. I agree not to engage in any of the following behaviors: - Copying or attempting to copy from others during an exam or on an assignment. - Communicating answers with another person during an exam. - Pre-programming a calculator or other personal electronic device to contain answers, or using other unauthorized information for exams. - Using unauthorized materials, i.e. prepared answers. - Allowing others to do an assignment or a portion of an assignment for you, including the use of a commercial term-paper service. - Submitting the same assignment for more than one course without prior approval of all the instructors involved. - Collaborating on an exam or assignment with any other person without prior approval from an instructor. - Taking an exam for another person or having someone take an exam for you. - Altering a previously graded exam or assignment for the purpose of a grade appeal or of gaining points in a re-grading process. - Submitting an electronic file the student knows to be unreadable or corrupted instead of a completed assignment.

Type your name and SID below [1 point]:

Name:

Enter your name:

Enter your SID:

INSTRUCTIONS:

1.Use Adobe Reader or Acrobat as a stand-alone application (NOT in a browser) to complete this assignment. (this software can b e accessed for free for UCB studentshttps://software.berkeley.edu/adobe-creative-cloud)

2.Give your responses ONLY in the space provided. Do NOT add any additional textboxes.

3.Please rename the file LASTNAME_FIRSTNAME_Midterm2_Spring2021.pdf

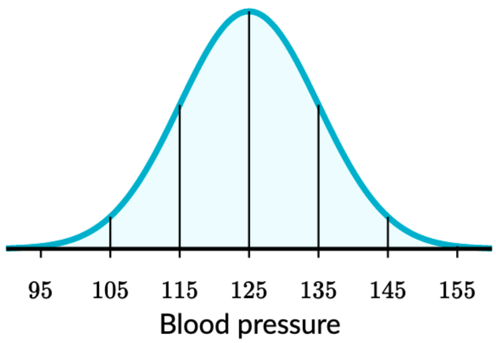
Unless otherwise specified in the question, format your answers according to the following guidelines: - present your answers rounded to two decimal places - present proportions as % values (40.50% rather than .405)

MAKE SURE YOU ARE WORKING WITH THIS DOCUMENT IN ADOBE AND YOU ARE NOT IN A BROWSER WINDOW

Question 1 [6 points total]

Use the following figure to answer questions 1.1 through 1.5

The following curve represents the systolic blood pressure of 5,000 females enrolled in a hypothetical study at UC Berkeley. The mean blood pressure was 125 mmHg. The solid vertical lines on the plot represent standard deviation-sized intervals. Answer the following questions about the curve.



1.1 [2 points] What is the variance of the systolic blood pressure? What are the units for this variance?

```
# Solution: 1 SD = 10 mmHg, variance = 100 mmHg^2
```

1.2 [1 point] What is the probability that a female has a systolic blood pressure of exactly 135 mmHg?

```
# Solution: 0
```

1.3 [1 point] What is the probability that a female has a systolic blood pressure greater than 115 mmHg? Round to two decimal places.

```
# Solution: P(x > mu - 1sigma) = .34 + .50 = .84 = 84%
```

1.4 [1 point] Write R code to calculate the probability that a female's systolic blood pressure is between 106 and 126 mmHg.

```
# Solution: pnorm(q = 126, mean = 125, sd = 10) - pnorm(q = 106, mean = 125, sd = 10)
```

1.5 [1 point] Write R code to find the 25th percentile of systolic blood pressure for females in the study.

```
# Solution: qnorm(0.25, mean = 125, sd = 10)
```

Question 2 [8 points total]

A new experimental drug (Drug A) has been developed to treat hospitalized COVID-19 patients over 70 years of age. Researchers are interested in modeling the probability that Drug A will provide these patients relief from COVID-19 symptoms (a dichotomous variable defined as a 70% reduction of symptoms). Let X be the number of hospitalized COVID-19 patients age 70 years+ who are enrolled in a clinical trial and take Drug A.

$$P(X = k) = \binom{n}{k}(0.73)^k(0.27)^{n-k}$$

2.1 [2 points] What is the name of this distribution? What does n represent?

```
# Binomial Distribution (1 pt), n is the total number of hospitalized COVID-19 patients
# age 70 years+ who are enrolled in the clinical trial and taking Drug A (1 pt)
```

2.2 [1 point] According to the distribution, what is the probability that Drug A will provide patients in the trial relief from COVID-19 symptoms?

```
# Solution: p = 0.73 or 73%
```

2.3 [1 point] What did the researchers assume when making the probability model? Check all that apply.

- ☐ Dependence between hospitalized COVID-19 patients in the trial.
- ☐ The event of relief from COVID-19 symptoms is a binary outcome.
- ☐ The number of COVID-19 patients enrolled in the trial is finite and equal to n.
- ☐ The chance that a patient in the trial experiences symptom relief is equal amongst all patients enrolled in the trial.

```
# Solution: Options 2, 3, 4
# option 1 is incorrect - we assume independence between individuals
```

2.4 [1 point] Choose the correct probability statement for the exact probability that three-quarters of the patients in the trial experienced symptom relief. Assume that the total number of individuals in the trial is 100,000.

- ○ $$P(X = k) = \binom{100,000}{25,000}(0.25)^{25,000}(0.75)^{75,000}$$
- ○ $$P(X = k) = \binom{100,000}{75,000}(0.73)^{75,000}(0.27)^{25,000}$$

○
$$P(X = k) = \binom{100,000}{75,000}(0.75)^{75,000}(0.25)^{25,000}$$

○
$$P(X = k) = \binom{100,000}{25,000}(0.73)^{25,000}(0.27)^{75,000}$$

```
# Solution: B. $$P(X = k) = {100,000 \choose 75,000}(0.73)^{75,000}(0.27)^{25,000}$$
```

2.5 [1 point] Use an R function to calculate the probability that more than 73,000 of the COVID-19 patients in the trial experience symptom relief after taking Drug A. You don't need to perform the calculation.

```
# Solution: 1 - pbinom(q = 73000, size = 100000, prob = 0.73)
# or pbinom(q = 73000, size = 100000, prob = 0.73, lower.tail = FALSE)
```

2.6 [2 points] Calculate the expected value and standard deviation of the probability distribution from the COVID-19 drug trial. Show your work.

```
# Solution: mu = n*p = (100,000)(0.73) = 73,000
#sigma = sqrt(n*p*1-p) = sqrt(100,000*.73*.27) = 140.39
```

Question 3 [5 points total]

A country's mortality data shows that 5% of deaths occur among individuals less than 24 years old (excluding neonatal deaths), 10% of deaths occur among individuals between 25-54 years old, 38% of deaths occur among individuals between 55-84 years old, and 47% of deaths occur among those over 85 years of age.

Among deaths in the <24 years old category, 13% are accidental. Among deaths in the 25-54 years old category, 31% are accidental. Among deaths in the 55-84 years old category, 22% are accidental. Among deaths in the 85+ years category, 15% are accidental.

3.1 [3 points] Pretend you have data from 100,000 death certificates. Fill out the following table given the information in the problem.

Deaths	<24 years	25-54 years	55-84 years	>85 years	Total
Accidental	A	B	C	D	E
Natural	F	G	H	I	J
Total	K	L	M	N	O

Row 1:

A.

```
# Solution: 650 deaths
```

B.

```
# Solution: 3100 deaths
```

C.

```
# Solution: 8360 deaths
```

D.

```
# Solution: 7050 deaths
```

E.

```
# Solution: 19160 deaths
```

Row 2:

F.

```
# Solution: 4,350 deaths
```

G.

```
# Solution: 6,900 deaths
```

H.

```
# Solution: 29,640 deaths
```

I.

```
# Solution: 39,950 deaths
```

J.

```
# Solution: 80,840 deaths
```

Row 3:

K.

```
# Solution: 5,000 deaths
```

L.

```
# Solution: 10,000 deaths
```

M.

```
# Solution: 38,000 deaths
```

N.

```
# Solution: 47,000 deaths
```

O.

```
# Solution: 100,000 deaths
```

3.2 [1 point] What is the probability that any given death in this country is accidental? Provide your answer as a percentage rounded to two decimal places.

```
# Solution: 19.16%
```

3.3 [1 point] What is the probability that any given death in this country is natural? Provide your answer as a percentage rounded to two decimal places.

```
# Solution: 80.84%
```

Question 4 [1 point total]

A doctor is examining a sick patient. 87% of sick patients in the doctor's city have the flu, while the other 13% are sick with COVID-19. Let F stand for the event of a patient being sick with flu and C stand for the event of a patient being sick with COVID-19. Assume that there are no other illness spreading in this city.

A common symptom of COVID-19 is a sore throat (the event of having which we denote as S). Assume the probability of having a sore throat if one has COVID-19 is 0.95. However, patients with the flu sometimes also develop a sore throat, and the probability of having a sore throat if one has the flu is 0.77.

Upon examining the patient, the doctor finds a sore throat. What is the probability that the patient has COVID-19? Provide your answer as a percentage rounded to two decimal places.

```
# Solution: 15.57%
```

Question 5 [5 points total]

A research paper in the UK found that there were 2000 heart-beating organ donors, who are severely sick patients requiring a ventilator to provide oxygen to their blood, across the UK within two years. Due to the low probability of such incidence, we assume the number of organ donors *per day* in UK follows a Poisson distribution, with 365 days in one year.

5.1 [2 points] **Please write the distribution of the number of organ donors *per day* in UK, including the value of its parameter, using notation learned in lecture. Round the calculated parameter to two decimal places.**

```
# Solution: mu = 2000/(365*2) = 2.74
# X ~ Poisson (2.74)
```

5.2 [2 points] **What is the probability that there are exactly two people who donate organs in the last day of the year? Show your work (not R code) for full credit. Write your answer as a percentage between 0 and 100 rounded to 2 decimal places.**

```
# Solution: P(X = 2) = (2.74)^2 * e^(-2.74)/2! = 24.24%
```

5.3 [1 point] **Write R code to calculate the probability that more than 10 people donate organs in the same day. You do not need to calculate the probability.**

```
# Solution: ppois(10, lambda = 2.74, lower.tail = FALSE)
```

Question 6 [8 points total]

Observational studies have shown that older adults who report low physical activity levels are at elevated risk of mortality compared with those who report moderate or high levels of activity. The free-living activity energy expenditure in one day has been measured and we define the physical activity to be high if the free-living activity energy expenditure is larger than 521 kcal/day. The true population distribution of the energy expenditure per day of the elderly in the US follows a normal distribution with a standard deviation of 80 kcal. We took a simple random sample of 100 older adults representative of the population and find that their mean energy expenditure per day is 550 kcal. We would like to perform a hypothesis testing with $\alpha = 0.05$ to see if there is strong evidence that the sample has a mean energy expenditure larger than 521 kcal.

6.1 [1 point] Is it a one-sided or two-sided hypothesis testing?

- ○ One-sided
- ○ Two-sided

```
# Solution: one-sided
```

6.2 [2 points] State your null and alternate hypotheses.

Null hypothesis:

```
# Solution: mu = 521
```

Alternative hypothesis:

```
# Solution: mu > 521
```

6.3 [1 point] Calculate the 95% confidence interval for μ and list the lower and upper bounds below. Round each to two decimal places.

Lower bound:

Upper bound:

```
# Solution: lower bound = 550 - 1.96 * 80/sqrt(100) = 534.32;
# upper bound = 550 + 1.96 * 80/sqrt(100) = 565.68
```

6.4 [1 point] Calculate the z-score using the values provided in the question and round it to two decimal places.

```
# Solution: (550-521)/(80/sqrt(100)) = 3.63
```

6.5 [1 point] Write one line of R code to calculate the p-value. You do not need to calculate the probability.

```
# Solution: pnorm(3.63, lower.tail = FALSE)
```

6.6 [2 points] Assume the p-value calculated is 0.0001 for the test. Interpret the p-value.

```
# Solution: The p-value of 0.0001  is less than our cut-off value of alpha = 0.05.
# So we have evidence to reject the null hypothesis that the mean energy expenditure of the sample
# is equal to 521 kcal/day and have evidence to support the alternative hypothesis
# that the mean energy expenditure of the sample is larger than 521 kcal/day.
```

Question 7 [5 points total]

7.1 [2 points] Continuing with Question 6, we would like to increase the power of our test. In which of the following scenarios, the power of the test is increased? Check all that apply.

- [ ] Take a simple random sample of 200 older adults instead of 100.
- [ ] Decrease α from 0.05 to 0.01.
- [ ] Another SRS is taken with mean energy expenditure per day 660 kcal.
- [ ] None of the above.

```
# Solution: A (increase the sample size), C (increase the effect size)
```

7.2 [1 point] If we would like the test to have a power of 95%, what will be the value of Type II error?

```
# Solution: beta = 0.05
```

7.3 [2 points] How many individuals do we need to include in our sample in order to estimate the mean energy expenditure for the elderly within 10 kcal with 95% confidence? Show your calculation and conclusion.

```
# Solution: z = 1.96, sigma = 80, m = 10, n = (1.96 * 80/10)^2 = 245.8624
# round up to 246
```

END