


DS 1000

Deadline: **Monday Dec 6 at 23:55 EST**

Assignment 4 – Chapters 8, 9, 13, 15, 16 and 32

This assignment is worth 10% of your final grade

- Questions with the computer symbol  must be answered using Python and Jupyter notebook. All codes must be provided.
- Submission must be made via Gradescope. **You must carefully assign questions to their corresponding pages.** Questions with no pages assigned to them will receive **zero** marks.
- We strongly encourage you to write all answers directly in the Jupyter Notebook and then download the notebook as a pdf file for submission.
- **Each student must submit their own work.** Scholastic offences are taken seriously, and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site:
http://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_undergrad.pdf

Question 1 (10 pts)

To test the effects of drugs and alcohol use on driving performance, 20 drivers were each asked to take a driving test under three conditions: sober, after two drinks, and after smoking marijuana. The order in which they drove the three conditions was randomized. An evaluator watched them drive on a test course and rated their accuracy on a scale from 1 to 10, without knowing which condition they were under each time.

- a) (2 pts) What was the explanatory variable in this experiment?
- b) (2 pts) What was the response variable in this experiment?
- c) (3 pts) Was this experiment single-blind, double-blind, or neither? Explain.
- d) (3 pts) Did this experiment use matched pairs, blocks, or neither? Explain.

Question 2 (5 pts)

Suppose you want to determine whether taking vitamins every day helps people lose weight. You survey 150 people who had been on a weight loss program for six months, and ask them whether or not they take vitamins every day and how much weight they lost. Suppose you found that the people who lost the most weight were more likely to have taken vitamins every day. Does this mean vitamins caused the additional weight loss? Explain why or why not.

Question 3 (10 pts)


For each of the following situations, state which type of sampling plan was used. Explain whether you think the sampling plan would result in a biased sample.

- a) (3 pts) You randomly sample 1500 students across all 50 elementary schools in your city.

- b) (3 pts) Out of the 50 schools in your city, 25 are from school district A and the other 25 from school district B. You randomly sample 750 students across schools from district A and 750 students across schools from district B.
- c) (4 pts) You obtain the list of students from all 50 schools in alphabetical order and select the first 1500 students in that list to be your sample.

Question 4 (15 pts)

Retail stores overflowing with merchandise can make consumers anxious, and minimally stocked spaces can have the same effect. Researchers investigated whether the use of ambient scents can reduce anxiety by creating feelings of openness in a crowded environment or coziness in a minimally stocked environment. Participants were invited to a lab that simulated a retail environment that was either jam-packed or nearly empty. For each of these two product densities, the lab was infused with one of three scents: (1) a scent associated with spaciousness, such as the seashore, (2) a scent associated with an enclosed space, like the smell of firewood, and (3) no scent at all. Consumers evaluated several products, and their level of anxiety was measured.

- a) (5 pts) Use a diagram like [Figure 9.2](#) (page 228) to display the treatments in a design with two factors: “product density” and “ambient scent.” Then outline the design of a completely randomized experiment to compare these treatments.
- b) (5 pts) There are 30 subjects available for the experiment, and they are to be randomly assigned to the treatments, an equal number of subjects in each treatment. Explain how you would number subjects and then randomly assign the subjects to the treatments.
- c) (5 pts)  Use Python code to generate the sample you have described in part b).

Question 5 (10 pts)

As suburban gardeners know, deer will eat almost anything green. In a study of pine seedlings at an environmental center in Ohio, researchers noted how deer damage varied with how much of the seedling was covered by thorny undergrowth:

Thorny Cover	Deer Damage	
	Yes	No
None	60	151
< 1/3	76	158
1/3 to 2/3	44	177
> 2/3	29	176

- a) (3 pts) What is the probability that a randomly selected seedling was damaged by deer?
- b) (3 pts) What are the conditional probabilities that a randomly selected seedling was damaged, given each level of cover?

c) (4 pts) Are cover and damage independent? That is, does knowing about the amount of thorny cover on a seedling change the probability of deer damage? Explain.

Question 6 (10 pts)

Although cigarette smoking has declined among U.S. youth in recent years, the use of some other tobacco products has increased. When high school students were asked which of several tobacco products they had used in the past 30 days, more than 40% of those who had used any tobacco product had used multiple tobacco products. Let A, B, and C be the events corresponding to the use of the following types of tobacco products in the past 30 days:

A = cigarettes

B = electronic cigarettes


C = other tobacco products (including cigars, pipes, smokeless tobacco, and hookahs)

Here are the probabilities that a randomly selected high school student used these different tobacco products:

$$P(A)=0.08 \quad P(B)=0.21 \quad P(C)=0.19$$

$$P(A \text{ and } B)=0.06 \quad P(A \text{ and } C)=0.03 \quad P(B \text{ and } C)=0.06$$

$$P(A \text{ and } B \text{ and } C) = 0.02$$

- a) (4 pts)  Make a Venn diagram of the events A, B, and C. As in [Figure 13.4](#) (page 297), mark the probabilities of every intersection involving these events. Use this diagram for parts b) through d).
- b) (2 pts) What is the probability that a randomly selected high school student did not use any tobacco product? Show all your work.
- c) (2 pts) What is the probability that a randomly selected high school student used electronic cigarettes but no other tobacco products?
- d) (2 pts) If a student smokes other tobacco products, what is the conditional probability that he or she also smokes cigarettes? If a student smokes cigarettes, what is the conditional probability that he or she also smokes other tobacco products? Comment your findings.

Question 7 (10 pts)

In a study of exercise, a large group of male runners walk on a treadmill for six minutes. After this exercise, their heart rates vary with a mean of 8.8 beats per five seconds and a standard deviation of 1.0 beats per five seconds. The researcher records the number of heartbeats per five seconds for each runner over a period of time. This distribution takes only whole-number values, so it is certainly not Normal.

- a) (3 pts) Let \bar{x} be the mean number of beats per five seconds after measuring heart rate for 24 five-second intervals (two minutes). What is the approximate distribution of \bar{x} according to the central limit theorem?
- b) (3 pts) What is the approximate probability that \bar{x} is less than 8?

- c) (4 pts) What is the approximate probability that the heart rate of a runner is less than 100 beats per minute? (Hint: Restate this event in terms of \bar{x} .)

Question 8 (15 pts)

Breast-feeding mothers secrete calcium into their milk. Some of the calcium may come from their bones, so mothers may lose bone mineral. Researchers measured the percentage change in mineral content of the spines of 47 mothers during three months of breast-feeding. Here are the data:

-4.7	-2.5	-4.9	-2.7	-0.8	-5.3	-8.3	-2.1	-6.8	-4.3
2.2	-7.8	-3.1	-1.0	-6.5	-1.8	-5.2	-5.7	-7.0	-2.2
-6.5	-1.0	-3.0	-3.6	-5.2	-2.0	-2.1	-5.6	-4.4	-3.3
-4.0	-4.9	-4.7	-3.8	-5.9	-2.5	-0.3	-6.2	-6.8	1.7
0.3	-2.3	0.4	-5.3	0.2	-2.2	-5.1			


- a) (5 pts) The researchers are willing to consider these 47 women to be a simple random sample from the population of all nursing mothers. Suppose that the percentage change in this population has standard deviation $\sigma = 2.5\%$. Make a stemplot of the data to verify that the data follow a Normal distribution quite closely. (Don't forget that you need both a 0 and a -0 stem because there are both positive and negative values.)
- b) (5 pts) Use a 99% confidence interval to estimate the mean percentage change in the population.
- c) (5 pts) Would it be correct to say that the probability is 99% that the mean percentage change in the population lies in the interval you computed in part (b)? Explain your answer

Question 9 (15 pts)

The composition of the earth's atmosphere may have changed over time. To try to discover the nature of the atmosphere long ago, we can examine the gas in bubbles inside ancient amber. Amber is tree resin that has hardened and been trapped in rocks. The gas in bubbles within amber should be a sample of the atmosphere at the time the amber was formed. Measurements on specimens of amber from the late Cretaceous era (75 million to 95 million years ago) give these percentages of nitrogen:

63.4 65.0 64.4 63.3 54.8 64.5 60.8 49.1 51.0

Assume that these observations are a simple random sample from the late Cretaceous atmosphere.

- a) (5 pts)  Construct a percentile-based 95% bootstrap confidence interval for the mean percentage of nitrogen in ancient air (the population). Use 1000 bootstrap samples.

- b) (5 pts) We wonder if ancient air differs significantly from the present atmosphere, which is 78.1% nitrogen. Based on your confidence interval in part (a), what do you conclude?