Final Exam – Math 243 on-campus classes

The exam contains 20 questions spanning textbook sections 1.1 through 6.1, with emphasis on the concepts from Section 4.3 through Section 6.1. Some questions are multiple-choice, some are short answer types, and others are free-response (support your answers with work shown in the space provided). Many of the questions have multiple parts; altogether there are about 40 responses, where many of the responses have a point value of 8 points. Total points: 200.

Three things are allowed on the desk during the exam (submit the sheet of notes and the scratch sheet):

- an 8.5 x 11 inch sheet of notes (both sides), hand-written by you (all other types of notes will be disallowed). The notes cannot be word-processed and they cannot be an electronically reproduced copy. Notes may contain worked-out examples and exercises, and calculator keystrokes.

- calculator with the advanced statistics program installed. All other electronic devices must be put away. Statistical tables, such as Table A, are not allowed.

- an 8.5 x 11 inch sheet of blank paper to use as scratch paper. Submit sheet of notes and scratch.

Your score on the Final Exam is a function of the exercises-for-practice you've worked and the insights you've gained from doing the projects.

Example question from Section 5.2:
The idea of insurance is that we all face risks that are unlikely but carry high cost. Think of a fire destroying your home. So we form a group to share the risk: we all pay a small amount, and the policy pays a large amount to those few of us whose homes burn down. An insurance company looks at the records for millions of homeowners and sees that the mean loss from fire in a year is $\mu = \$250$ per policy and that the standard deviation of the loss is $\sigma = \$1000$. The distribution of losses is extremely right-skewed: most people have $0$ loss, but a few have
large losses. The company plans to sell fire insurance for $250 plus enough to cover its costs and profit. If the company sells 10,000 policies, what is the approximate probability that the average loss in a year will be greater than $275?

a. In one sentence, what does the random variable represent in the context of the question? Give an appropriate symbol for this random variable.

b. Draw a sketch of the distribution of the random variable. Show the mean, the point $275, and shade the area that corresponds to the probability of interest.

c. Express the event of interest in terms of the random variable and complete the mathematical sentence that begins with:

\[ P( \) 

d. Interpret the probability value in the context of the scenario.

**Key:**

a. The random variable represents the average (or mean) loss in a year for a sample of 10,000 policies. An appropriate symbol for the random variable is: \( \bar{x} \)

b. A sketch of the applicable distribution of the statistic \( \bar{x} \) is:

(z-axis optional)

\[ \mu = \$250 \]
\[ SE_{\bar{x}} = \frac{\$1000}{\sqrt{10,000}} \]

Note: The Central Limit Theorem tells us that although the population distribution of an individual loss is right-skewed with a mean loss of $250, the sampling distribution of the average loss, \( \bar{x} \), measured on a sample of 10,000 policies is approximately normal with a mean loss of $250 per policy. Because the standard deviation of the sampling distribution of \( \bar{x} \) is $1000 divided by the square root of 10,000, it is much smaller than the population standard deviation of $1000, meaning that
average losses are much closer than individual losses to the true average loss. The probability that an average loss will be greater than $275 is represented by the shaded upper tail above $275.

c. An appropriate mathematical sentence is:

\[ P(\bar{x} > 275) = \text{shaded area} = \text{normalcdf}\left(275, \infty, 250, \frac{1000}{\sqrt{10,000}}\right) \approx 0.0062 \]

OR: \[ P(\bar{x} > 275) = \text{shaded area} = P(z > 2.50) = \text{normalcdf}(2.50, \infty, 0, 1) \approx 0.0062 \]

d. There is a less than 1% (0.62%) chance that the average loss to the insurance company per year on 10,000 policies will exceed $275.

Example question from Section 6.1.

The proportion of intervals of the form \( \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \) that capture the true mean \( \mu \) is the confidence level, or capture rate, C. A 95% confidence interval (C.I.) will have a 95% capture rate, for example. What are the conditions for building a C.I. for estimating the true mean \( \mu \)?

Key:

The conditions are:

- Your sample is random (or experimental treatments, such as weight loss plans, are randomly assigned).
- The underlying population is normal. The capture rate will be approximately correct for non-normal populations as long as the sample size is large (generally \( n > 30 \)).
- The standard deviation population standard deviation \( \sigma \) is known.

Note to students: the requirement that \( \sigma \) is known is unrealistic. In Math 244 we learn what substitution we can make for \( \sigma \).

Example question Section 6.1:

A 95% confidence interval for (to estimate) a population mean \( \mu \) is calculated from a sample of weights, and the resulting confidence interval is: 42 to 48 pounds. Which of the statements below are true
interceptions of a confidence interval? **Explain** why the false statements are false.

a. Ninety-five percent of the weights in the population are between 42 and 48 pounds.
b. Ninety-five percent of the weights in the sample are between 42 and 48 pounds.
c. The probability that the interval (42, 48) includes the true population mean $\mu$ is 95%.
d. The sample mean $\bar{x}$ may not be in the confidence interval.
e. If 200 confidence intervals were generated using the same process, about 10 of the confidence intervals would not capture the true population $\mu$.

**Key:**

a. False. The “ninety-five percent” refers to the capture rate of the mean $\mu$, a fixed but unknown parameter of the population. The capture rate of ninety-five percent tells us nothing about the range of possible weights in the population.
b. False. The capture rate of ninety-five percent tells us nothing about the range of possible weights in the sample.
c. False. Once you replace the random variable $\bar{x}$ in the recipe for a confidence interval, no randomness remains. The probability that the interval (42, 48) has captured the true population $\mu$ is now either 0% or 100%; i.e., (42, 48) either includes $\mu$ or it does not.
d. False. The recipe of the confidence interval always starts with the sample mean $\bar{x}$. The sample mean is in the center of the confidence interval.
e. True. The capture rate is 95%. If 200 confidence intervals were generated using the same process, about 95% of 200 or 190 of the intervals would, in the long run, include the true population mean $\mu$, and about 5% of 200 or 10 of the intervals would not. We just wouldn’t know which ones did and which ones did not.

More examples are available in a separate link on this sidebar.