SAMPLING AND CONFIDENCE INTERVALS

A STATISTICAL MODELING

A.1 DISTINGUISHING BETWEEN RANDOM VARIABLES AND OBSERVED VALUES

Ex 1: You are about to roll a fair six-sided die.

1. Let D be the result of the roll **before** you throw the die. Is D a random variable or an observed value? What is $P(D = 6)$?
2. You throw the die, and it lands on the number 4. Let d be this result. Is d a random variable or an observed value?
3. Mathematical notation uses uppercase letters for random variables and lowercase for observed values. Write the relationship between D and d using an "equals" sign and words.

Ex 2: A biologist is planning a study to measure the weight of 10 penguins.

1. **Planning Phase:** She writes a formula to calculate the average weight she *expects* to find: $\frac{X_1+X_2+\cdots+X_{10}}{10}$. Explain why she uses uppercase X.
2. **Data Analysis Phase:** She measures the penguins and gets the list: $\{12.5, 11.8, 13.2, \ldots\}$. She calculates the average: $\frac{x_1+x_2+\cdots+x_{10}}{10} = 12.4$. Explain why she uses lowercase x.
3. Which of the two averages is the **Estimator** ($\overline{X}$) and which is the **Estimate** ($\bar{x}$)?

Ex 3: Consider the sentence: "The mean of the random variable X is $\mu = 50$."
Now consider a sample x_1, x_2, x_3 taken from this distribution.
True or False (and explain):

1. The average of the observed values $\bar{x}$ will always be exactly 50.
2. The expected value of the estimator $\overline{X}$ is 50.

A.2 MODELING WITH RANDOM VARIABLES

Ex 4: An orchard harvests apples to be sold in crates. The weight of a single apple varies due to natural conditions.
Based on previous harvests, the weight of an apple has a mean $\mu = 150$ g and a standard deviation $\sigma = 10$ g.
A crate contains $n = 40$ apples. Let X_i be the random variable representing the weight of the i-th apple in the crate.

1. Define the random variable X_i in this context (what does it measure?).
2. State the expected value and standard deviation of a single variable X_i.
3. What does the sum $S_{40} = \sum_{i=1}^{40} X_i$ represent in this context?

Ex 5: A sociologist surveys the daily commute time of workers in a large city.
The population mean commute time is $\mu = 45$ minutes with a standard deviation $\sigma = 12$ minutes.
The sociologist selects a random sample of 100 workers. Let X_k be the random variable representing the commute time of the k-th worker.

1. Define the variable X_k. Is it a discrete (Bernoulli) or continuous variable?
2. Give the mean and standard deviation of a single variable X_k.
3. Express the sample mean $\overline{X}_{100}$ in terms of X_k and explain what it represents.

Ex 6: In a city election, it is known that 45% of the population intends to vote for Candidate A. A pollster surveys a random sample of $n = 100$ voters.
Let X_i be the random variable representing the response of the i-th person surveyed.

1. Define the random variable X_i using the convention 1 for "Success" and 0 for "Failure".
2. State the probability distribution of X_i.
3. What does the sum $S_{100} = \sum_{i=1}^{100} X_i$ represent in this context?

Ex 7: A factory produces light bulbs. Historical data shows that 2% of the bulbs are defective. A quality control manager picks a sample of 50 bulbs to inspect.
Let X_k be the random variable representing the state of the k-th bulb selected.

1. Define the variable X_k to model the proportion of defective bulbs.
2. Give the mean and standard deviation of a single variable X_k.
3. Express the sample proportion $\hat{P}$ in terms of X_k.

A.3 CALCULATING SAMPLE STATISTICS

Ex 8: Consider a random variable X representing a characteristic of a population.
A sample of size $n = 10$ is taken: $x = \{48, 52, 50, 47, 53, 51, 49, 50, 48, 55\}$.

1. Calculate the sample mean $\bar{x}$.
2. Calculate the unbiased sample variance s_n^2.
3. Explain why the sample mean is considered an "unbiased" estimator of the population mean.

Ex 9: A farmer weighs a random sample of $n = 5$ apples from his harvest to estimate the quality of the crop. The weights in grams are: $x = \{120, 125, 118, 122, 130\}$.

1. Calculate the sample mean $\bar{x}$.
2. Calculate the unbiased sample variance s_n^2.
3. Why do we divide by $n-1$ instead of n when calculating the variance s_n^2?

Ex 10: A teacher records the scores of a sample of $n = 6$ students on a difficult mathematics test (out of 50). The scores are: $x = \{32, 40, 28, 35, 42, 27\}$.

1. Calculate the sample mean $\bar{x}$.
2. Calculate the unbiased sample variance s_n^2.
3. Explain what it means for the variance estimator to be "unbiased".

B CONFIDENCE INTERVALS FOR MEANS WITH KNOWN VARIANCE

B.1 FINDING CRITICAL VALUES

Ex 11: A sociologist wants to construct a 90% confidence interval for the average income in a city.
Using your calculator, determine the critical value z required for this interval, rounded to **two decimal places**.

$$z \approx \boxed{}$$

Ex 12: A biologist is analyzing the length of leaves and needs to calculate a 95% confidence interval.
Using your calculator, determine the critical value z required for this interval, rounded to **two decimal places**.

$$z \approx \boxed{}$$

Ex 13: For a high-precision engineering part, a quality control manager requires a 99% confidence interval for the mean diameter.
Using your calculator, determine the critical value z required for this interval, rounded to **two decimal places**.

$$z \approx \boxed{}$$

B.2 CONSTRUCTING CONFIDENCE INTERVALS

Ex 14: A factory produces bottles of water. The volume of water is normally distributed with a standard deviation $\sigma = 4$ ml. A quality control check of 64 bottles shows a mean volume of 503 ml.
Find the 95% confidence interval for the true mean volume μ.

Ex 15: A machine fills bags of sugar. The weight of the bags is normally distributed with standard deviation $\sigma = 15$ g. A sample of 100 bags is weighed, yielding a mean of 1002 g.
Find the 90% confidence interval for the true mean weight μ.

Ex 16: A company manufactures batteries. The battery life is normally distributed with a standard deviation $\sigma = 12$ minutes. A random sample of 36 batteries is tested, yielding a mean life of 240 minutes.
Find the 99% confidence interval for the true mean battery life μ.

C CONFIDENCE INTERVALS FOR MEANS WITH UNKNOWN VARIANCE

C.1 ESTIMATING THE MEAN WITH UNKNOWN VARIANCE

Ex 17: A biologist measures the length of 40 fish of a certain species. The sample mean is 25.4 cm and the sample standard deviation is $s_n = 3.2$ cm.

1. Why can we use the normal approximation for the confidence interval even though the population standard deviation is unknown?
2. Calculate the 95% confidence interval for the mean length.

Ex 18: A consumer group tests the lifespan of 50 randomly selected tires of a specific brand. The sample mean lifespan is $42,000$ km and the sample standard deviation is $s_n = 4,000$ km.

1. Explain why the normal distribution can be used to construct the confidence interval even though the population variance is unknown.
2. Calculate the 95% confidence interval for the mean lifespan of the tires.

Ex 19: An orchard manager selects a random sample of 64 apples to estimate the harvest quality. The sample mean weight is 150 g with a sample standard deviation of $s_n = 12$ g.

1. Justify the use of the z-interval for this estimation.
2. Calculate the 99% confidence interval for the mean weight of an apple.

D CONFIDENCE INTERVALS FOR PROPORTIONS

D.1 ESTIMATING POPULATION PROPORTIONS

Ex 20: A factory produces electronic components. A quality control manager takes a random sample of 500 components and finds that 25 of them are defective.

1. Calculate the sample proportion $\hat{p}$ of defective components.
2. Construct a 95% confidence interval for the true proportion of defective components produced by the factory.

Ex 21: In a recent poll before an election, 1000 voters were surveyed randomly. 520 of them stated they would vote for the incumbent mayor.

1. Calculate the sample proportion $\hat{p}$ of voters supporting the mayor.
2. Construct a 95% confidence interval for the true proportion of voters who support the mayor.

Ex 22: A restaurant chain wants to estimate its customer satisfaction rate. They survey 200 random customers, and 160 report being satisfied with their meal.

1. Calculate the sample proportion $\hat{p}$ of satisfied customers.
2. Construct a 95% confidence interval for the true proportion of satisfied customers.

E HYPOTHESIS TESTING USING CONFIDENCE INTERVALS

E.1 TESTING CLAIMS ABOUT THE POPULATION MEAN

Ex 23: A car manufacturer claims that their new model consumes an average of 5.5 liters of fuel per 100 km.
A consumer group tests 35 cars and finds a sample mean of 5.8 liters/100km with a standard deviation of 0.8.

1. Construct a 95% confidence interval for the true mean consumption.
2. Does the confidence interval support the manufacturer's claim? Explain.

Ex 24: A snack company sells bags of potato chips labeled with a net weight of 200 g. A consumer protection agency suspects the bags are underfilled.
They weigh a random sample of 50 bags and find a mean weight of 196 g with a standard deviation of 8 g.

1. Construct a 95% confidence interval for the true mean weight of the bags.
2. Based on this interval, is the company's claim of 200 g valid? Justify your answer.

Ex 25: The national average score on a mathematics test is known to be 75. A school principal introduces a new teaching method and wants to know if it has changed student performance.
A sample of 40 students taught with the new method achieves a mean score of 78 with a standard deviation of 12.

1. Construct a 95% confidence interval for the mean score under the new method.
2. Can the principal conclude that the new method has significantly changed the average score compared to the national standard?

Ex 26: A laptop manufacturer advertises that its battery lasts an average of 10 hours. A tech reviewer tests 64 laptops and finds a sample mean of 9.6 hours with a standard deviation of 1.2 hours.

1. Construct a 99% confidence interval for the true mean battery life.
2. Does the test support the manufacturer's advertisement at the 1% significance level?

F DETERMINING SAMPLE SIZE

F.1 DETERMINING SAMPLE SIZE

Ex 27: A biologist wants to estimate the mean life span of a certain species of insect. She wants the estimate to be accurate within 5 days of the true mean with 90% confidence. From previous data, the standard deviation of the life span is known to be $\sigma = 28$ days.
Determine the minimum number of insects she needs to study.

Ex 28: A quality control engineer wants to estimate the mean weight of cereal boxes produced by a factory. She wants the estimate to be within 3 grams of the true mean with 95% confidence. The standard deviation of the weight is known to be $\sigma = 15$ grams.
Determine the minimum number of boxes she needs to weigh.

Ex 29: A city planner wants to estimate the average daily commute time for residents. He wants the margin of error to be no more than 2 minutes with 99% confidence. Previous studies suggest the population standard deviation is $\sigma = 12$ minutes.
Calculate the required sample size.