5. According to a 2017 national survey in Country B, the mean number of bedrooms in newly built houses was 2.9. Rodney, a researcher, believes the mean number of bedrooms in newly built houses in the country was different in 2024 than it was in 2017. To investigate his belief, he took a large random sample of newly built houses in Country B in 2024 and recorded the number of bedrooms in each house. The distribution of the number of bedrooms for the sampled houses is summarized in the table.

Distribution of the Number of Bedrooms for the Houses Sampled in 2024

Number of Bedrooms	1	2	3	4	5	6
Proportion of Houses	0.12	0.22	0.28	0.22	0.14	0.02

A.

i. A house from the sample will be selected at random. What is the probability that the house had fewer than 3 bedrooms? Show your work.

ii. What is the mean number of bedrooms for the sample of newly built houses in 2024? Show your work.

B. Rodney will use a one-sample t-test for a population mean to test his belief.

i. In the context of Rodney's investigation, state the hypotheses for the test.

ii. Explain, in context, what a Type I error would be for Rodney's hypothesis test.

C. A different researcher, Keisha, suggests using a confidence interval to investigate whether the mean number of bedrooms in newly built houses in 2024 in Country B was different from 2.9.

Assume the conditions for inference have been met. Using Rodney's data, Keisha calculated a one-sample 97 percent confidence interval to estimate the population mean as $(3.01, 3.19)$. Based on the confidence interval, what conclusion can be made for Rodney's hypothesis test in part B at $\alpha = 0.03$? Justify your answer.

4. In an online game, players move through a virtual world collecting geodes, a type of hollow rock. When broken open, these geodes contain crystals of different colors that are useful in the game. A red crystal is the most useful crystal in the game. The color of the crystal in each geode is independent and the probability that a geode contains a red crystal is 0.08.

(a) Sarah, a player, will collect and open geodes until a red crystal is found.

(i) Calculate the mean of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

(ii) Calculate the standard deviation of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

(b) Another player, Conrad, decides to play the game and will stop opening geodes after finding a red crystal or when 4 geodes have been opened, whichever comes first. Let $Y =$ the number of geodes Conrad will open. The table shows the partially completed probability distribution for the random variable Y.

Number of geodes Conrad will open, y	1	2	3	4
Probability, $P(Y = y)$	0.08	0.0736		

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 4** on this page.

(i) Calculate $P(Y = 3)$. Show your work.

(ii) Calculate $P(Y = 4)$. Show your work.

(c) Consider the table and your results from part (b).

(i) Calculate the mean of the distribution of the number of geodes Conrad will open. Show your work.

(ii) Interpret the mean of the distribution of the number of geodes Conrad will open, which was calculated in part (c-i).

GO ON TO THE NEXT PAGE.

Descriptive Statistics

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n} \qquad s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$\hat{y} = a + bx \qquad \bar{y} = a + b\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) \qquad b = r\frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum(x_i - \mu_X)^2 P(x_i)}$
If X has a **binomial** distribution with parameters n and p, then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \ldots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a **geometric** distribution with parameter p, then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \ldots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: statistic $\pm$ (critical value)(standard error of statistic)

Chi-square statistic: $\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ where $\hat{p}_c = \dfrac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\overline{X}$	$\mu_{\overline{X}} = \mu$	$\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$	$s_{\overline{X}} = \dfrac{s}{\sqrt{n}}$
For two populations: $\overline{X}_1 - \overline{X}_2$	$\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$	$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \dfrac{\sigma}{\sigma_x\sqrt{n}}$, where $\sigma_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n}}$	$s_b = \dfrac{s}{s_x\sqrt{n-1}}$, where $s = \sqrt{\dfrac{\sum(y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}}$

andard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the ndard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

(b) Consider two rules for identifying outliers, method A and method B. Let method A represent the $1.5 \times \text{IQR}$ rule, and let method B represent the 2 standard deviations rule.

(i) Using method A, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(ii) The mean length of stay for the sample is 7.42 days with a standard deviation of 2.37 days. Using method B, determine any data points that are potential outliers in the distribution of length of stay. Justify your answer.

(c) Explain why method A might identify more data points as potential outliers than method B for a distribution that is strongly skewed to the right.

GO ON TO THE NEXT PAGE.

2. Researchers will conduct a year-long investigation of walking and cholesterol levels in adults. They will select a random sample of 100 adults from the target population to participate as subjects in the study.

(a) One aspect of the study is to record the number of miles each subject walks per day. The researchers are deciding whether to have subjects wear an activity tracker to record the data or to have subjects keep a daily journal of the miles they walk each day. Describe what bias could be introduced by keeping the daily journal instead of wearing the activity tracker.

GO ON TO THE NEXT PAGE.

During the course of the study, the subjects will have their cholesterol levels measured each month by a doctor. The researchers will perform a significance test at the end of the study to determine whether the average cholesterol level for subjects who walk fewer miles each day is greater than for those who walk more miles each day.

(b) Selecting a random sample creates a reasonable representative sample of the target population. Explain the benefit of using a representative sample from the population.

(c) Suppose the researchers conduct the test and find a statistically significant result. Would it be valid to claim that increased walking causes a decrease in average cholesterol levels for adults in the target population? Explain your reasoning.

GO ON TO THE NEXT PAGE.

3. To increase morale among employees, a company began a program in which one employee is randomly selected each week to receive a gift card. Each of the company's 200 employees is equally likely to be selected each week, and the same employee could be selected more than once. Each week's selection is independent from every other week.

(a) Consider the probability that a particular employee receives at least one gift card in a 52-week year.

(i) Define the random variable of interest and state how the random variable is distributed.

(ii) Determine the probability that a particular employee receives at least one gift card in a 52-week year. Show your work.

GO ON TO THE NEXT PAGE.

(b) Calculate and interpret the expected value for the number of gift cards a particular employee will receive in a 52-week year. Show your work.

(c) Suppose that Agatha, an employee at the company, never receives a gift card for an entire 52-week year. Based on her experience, does Agatha have a strong argument that the selection process was not truly random? Explain your answer.

GO ON TO THE NEXT PAGE.

5. A company that manufactures smartphones developed a new battery that has a longer life span than that of a traditional battery. From the date of purchase of a smartphone, the distribution of the life span of the new battery is approximately normal with mean 30 months and standard deviation 8 months. For the price of $50, the company offers a two-year warranty on the new battery for customers who purchase a smartphone. The warranty guarantees that the smartphone will be replaced at no cost to the customer if the battery no longer works within 24 months from the date of purchase.
   (a) In how many months from the date of purchase is it expected that 25 percent of the batteries will no longer work? Justify your answer.
   (b) Suppose one customer who purchases the warranty is selected at random. What is the probability that the customer selected will require a replacement within 24 months from the date of purchase because the battery no longer works?
   (c) The company has a gain of $50 for each customer who purchases a warranty but does not require a replacement. The company has a loss (negative gain) of $150 for each customer who purchases a warranty and does require a replacement. What is the expected value of the gain for the company for each warranty purchased?

3. A shopping mall has three automated teller machines (ATMs). Because the machines receive heavy use, they sometimes stop working and need to be repaired. Let the random variable X represent the number of ATMs that are working when the mall opens on a randomly selected day. The table shows the probability distribution of X.

Number of ATMs working when the mall opens	0	1	2	3
Probability	0.15	0.21	0.40	0.24

   (a) What is the probability that at least one ATM is working when the mall opens?
   (b) What is the expected value of the number of ATMs that are working when the mall opens?
   (c) What is the probability that all three ATMs are working when the mall opens, given that at least one ATM is working?
   (d) Given that at least one ATM is working when the mall opens, would the expected value of the number of ATMs that are working be less than, equal to, or greater than the expected value from part (b) ? Explain.