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

4. A survey conducted by a national research center asked a random sample of 920 teenagers in the United States how often they use a video streaming service. From the sample, 59% answered that they use a video streaming service every day.

(a) Construct and interpret a 95% confidence interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.

GO ON TO THE NEXT PAGE.

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

(b) Based on the confidence interval in part (a), do the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would respond that they use a video streaming service every day is not 0.5? Justify your answer.

GO ON TO THE NEXT PAGE.

Begin your response to **QUESTION 5** on this page.

5. Studies have shown that foods rich in compounds known as flavonoids help lower blood pressure. Researchers conducted a study to investigate whether there was a greater reduction in blood pressure for people who consumed dark chocolate, which contains flavonoids, than people who consumed white chocolate, which does not contain flavonoids. Twenty-five healthy adults agreed to participate in the study and add 3.5 ounces of chocolate to their daily diets. Of the 25 participants, 13 were randomly assigned to the dark chocolate group and the rest were assigned to the white chocolate group. All participants had their blood pressure recorded, in millimeters of mercury (mmHg), before adding chocolate to their daily diets and again 30 days after adding chocolate to their daily diets.

The reduction in blood pressure (before minus after) for each of the participants in the two groups is shown in the dotplots below.
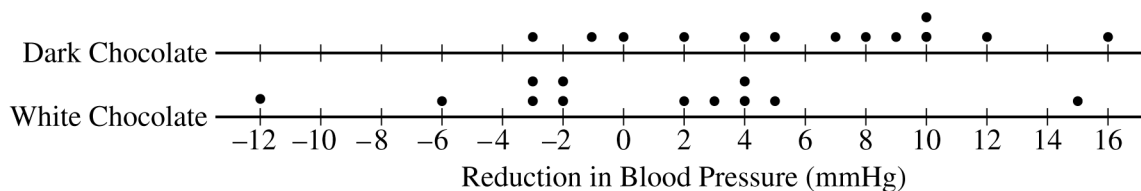
(a) Determine and compare the medians of the reduction in blood pressure for the two groups.

The researchers found the mean reduction in blood pressure for those who consumed dark chocolate is $\overline{x}_{dark} = 6.08$ mmHg and the mean reduction in blood pressure for those who consumed white chocolate is $\overline{x}_{white} = 0.42$ mmHg.

(b) One researcher indicated that because the difference in sample means of 5.66 mmHg is greater than 0 there is convincing statistical evidence to conclude that the population mean blood pressure reduction for those who consume dark chocolate is greater than for those who consume white chocolate. Why might the researcher's conclusion, based only on the difference in sample means of 5.66 mmHg, not necessarily be true?

GO ON TO THE NEXT PAGE.

Question 4: Focus on Inference   **4 points**

General Scoring Notes

- This two-part question is scored in four sections. Each section is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). Part (a) includes three sections that may appear in any order in the response. The first section includes identification of the appropriate confidence interval in part (a). The second section includes verifying the conditions for inference in part (a) and calculating the values of the endpoints of the confidence interval. The third section includes the interpretation of the confidence interval in part (a). The fourth section includes the response to part (b). The response is then categorized based on the scores assigned to each section and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each section of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

Model Solution	Scoring
(a) **Section 1** The appropriate procedure is a one-sample z-interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.	**Essentially correct (E)** if the response satisfies the following two components: 1. Identifies the appropriate procedure as a one-sample z-interval by name or formula or by the calculations of the correct confidence interval endpoint values 2. States that the parameter of interest is the population proportion **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- The response to component 2 concerning the statement of "population proportion" can be found in any of the three sections of part (a).
- Any notation used to represent sample proportion or population proportion should remain consistent throughout part (a).

Model Solution	Scoring
(a) **Section 2** This survey selected a random sample of 920 teenagers in the United States, which enables the interval to be generalized to the population of interest. This sample of 920 teenagers is less than 10% of the total number of teenagers in the United States, which is required as sampling was conducted without replacement from a finite population. In addition, there were more than 10 successes and 10 failures as $(920)(0.59) = 542.8$ (or 543) responded that they use a streaming service daily and $(920)(0.41) = 377.2$ (or 377) responded that they did not. Thus, the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal. Therefore, a 95% confidence interval for the population proportion is given by $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 0.59 \pm 1.96 \sqrt{\dfrac{(0.59)(0.41)}{920}}$, which is 0.59 ± 0.032, and the interval is $(0.558, 0.622)$.	**Essentially correct (E)** if the response satisfies the following four components: 1. States that a random sample was selected 2. Indicates 920 is less than ten percent of all teenagers in the United States 3. Verifies that there are at least 10 successes and failures by calculating the following values $n\hat{p} = (920)(0.59) \approx 542.8$ and $n(1-\hat{p}) = (920)(1-0.59) \approx 377.2$ 4. Reports the values for a correct interval consistent with the procedure stated in Section 1 **Partially correct (P)** if the response satisfies either component 3 or component 4 and at least one of the other three components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:
- Stating the large sample condition without verification is not sufficient for component 3.
- If the response includes an inappropriate check of conditions, such as $n > 30$, then component 3 is not satisfied.
- Supporting work, showing formulas or calculations, is not required for component 4.
- If the interval values are correct, the use of a one-sample z procedure for proportion can be used to satisfy component 1 of Section 1.
- A response that uses the value of $x = 543$ will result in an interval of $(0.5584, 0.6219)$, and a response that uses the value of $x = 542$ will result in an interval of $(0.5573, 0.6209)$. These interval endpoint values may be used to satisfy component 4.
- If the response includes supporting work for calculating the confidence interval that displays a correct formula with correct values inserted for $\hat{p}$, n, and z, then component 4 is satisfied even if values for the endpoints of the confidence interval are not displayed or calculated incorrectly.
- A response that computes an interval in percentages rather than proportions may satisfy component 4 if the response correctly indicates the use of percentages, $(55.8\%, 62.2\%)$.
- Minor errors or omissions when checking assumptions may be considered if holistic scoring is required.

Model Solution	Scoring
(b) **Section 4** The 95% confidence interval of $(0.558, 0.622)$ indicates that any value between 0.558 and 0.622 is a plausible value for the proportion of all teenagers in the United States who use video streaming every day. Because the value 0.5 is not contained in the interval, the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would use a streaming service every day is not 0.5.	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides a correct conclusion (there is convincing evidence or there is not convincing evidence that 0.50 is a plausible value), consistent with the interval calculated in part (a) 2. Provides correct justification based on whether the value of 0.5 is contained in the interval calculated in part (a) **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:
- Any use of a hypothesis test to answer Section 4 cannot be used to satisfy component 2. However, the statement of a conclusion that 0.50 is not a plausible value based on the results of the hypothesis test can be used to satisfy component 1.
- A response that correctly uses incorrectly calculated proportion interval endpoint values can satisfy components 1 and 2 if the response is consistent with interval values.
- A response that includes a correct interpretation of the confidence interval in Section 4 can receive credit for the interpretation in Section 3 if no interpretation was provided in Section 3.
- A response that presents a correct conclusion in the context of the population percentage may satisfy both component 1 and component 2.
- A response that indicates that there is evidence that the proportion of teenagers in the United States who would use a streaming service every day is greater than 0.5 based on the fact that the entire confidence interval is above 0.5 may satisfy both component 1 and component 2.
- A response that does not provide the correct conclusion and justification but does recognize that the value is not in the interval may be considered a positive if holistic scoring is required.

Scoring for Question 4
Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.

	Score
Complete Response	4
Substantial Response	3
Developing Response	2
Minimal Response	1

If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.