2021

AP CollegeBoard

AP® Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free Response Question 5

- ☑ **Scoring Guideline**
- ☑ **Student Samples**
- ☑ **Scoring Commentary**

© 2021 College Board. College Board, Advanced Placement, AP, AP Central, and the acorn logo are registered trademarks of College Board. Visit College Board on the web: collegeboard.org.

AP Central is the official online home for the AP Program: apcentral.collegeboard.org.

Question 5: Multi-Focus — 4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part 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 part 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)	No, the researcher's claim is not correct. Although the Baltimore survey has the least number of teens who consumed a soft drink in the past week, it also has the least number of teens surveyed among the three cities' samples. Comparing the numbers of teens who consumed a soft drink in the past week is meaningless without considering the sample sizes. The comparison should be based on proportions rather than counts. In fact, the proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportions for either of the other two cities, $\frac{1,232}{1,663} \approx 0.741$ for Detroit and $\frac{1,482}{2,280} = 0.65$ for San Diego.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates that the researcher's claim is not correct (or "may not be correct", if proportions are not reported) 2. Provides an explanation that is based on at least one of the following: • The proportions (or relative frequencies), not counts, should be compared because the sample sizes are not equal *OR* • The proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportion for at least one of the two other cities **Partially correct (P)** if the response satisfies component 1 AND states that the sample sizes are not equal *OR* if the response satisfies component 2 only *OR* if the response provides a correct proportion (or relative frequency) for Baltimore and at least one other city. **Incorrect (I)** if the response does not meet the criteria for E or P.

© 2021 College Board

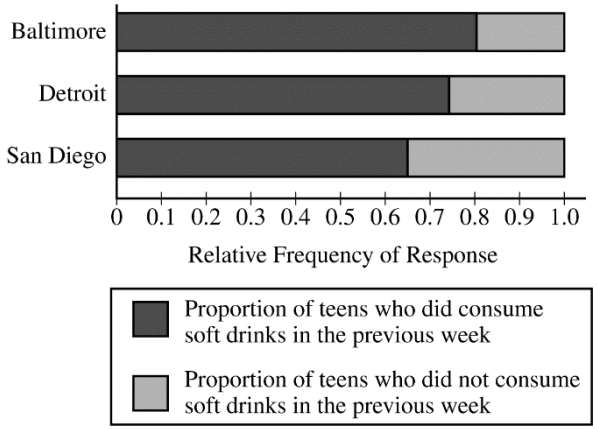
Additional Notes:

- A response that compares the proportion of Baltimore teens who consumed a soft drink in the past week to the combined proportion of the other two cities, $\frac{1,323 + 1,482}{1,663 + 2,280} = \frac{2,714}{3,943} \approx 0.688$ is not equivalent to comparing counts and may earn a P.
- If the "yes" count for each city is divided by the same number (e.g., the total number of respondents who consumed a soft drink in the past week, 3,441; or the total sample size, 4,847), then the response is equivalent to comparing counts and should be scored I.
- A response may satisfy component 2 by providing correct numerical values for the proportion of Baltimore teens who did *not* consume a soft drink in the past week, $\frac{177}{904} \approx 0.196,$ AND the proportion of teens who did *not* consume a soft drink in the past week for at least one of the two other cities, either $\frac{431}{1,663} \approx 0.259$ for Detroit and/or $\frac{798}{2,280} = 0.35$ for San Diego.
- If work is shown, calculation or transcription errors should be ignored in scoring.
- Statistical notation should be ignored in scoring.

© 2021 College Board

Model Solution	Scoring

(b) **(i)** A segmented bar graph of the relative frequencies based on the information in the table is shown below:



(ii) The proportion of teens who consumed a soft drink in the previous week are shown below:

- Baltimore: $\frac{727}{904} \approx 0.804$
- Detroit: $\frac{1,232}{1,663} \approx 0.741$
- San Diego: $\frac{1,482}{2,280} = 0.65$

San Diego has the smallest proportion of teens (0.65) who consumed a soft drink in the previous week.

Essentially correct (E) if the response satisfies the following four components:

1. Constructs a segmented bar graph in part (b-i), with the bars correctly segmented
2. Includes clear labeling of the proportions of teens who consumed a soft drink in the previous week and the proportions of teens who did not consume a soft drink in the previous week for the segmented bar graph provided in part (b-i)
3. Identifies San Diego as the city with the smallest proportion of teens who consumed a soft drink in the previous week in part (b-ii)
4. Reports the correct numerical value of the proportion of teens who consumed a soft drink in the previous week for the city identified in part (b-ii)

Partially correct (P) if the response satisfies only two or three of the four components.

Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that constructs a segmented bar graph with the lengths of the segments representing the relative frequencies of teens who consumed a soft drink in the previous week between 0.75 and 0.85 for Baltimore, between 0.7 and 0.8 for Detroit, and between 0.6 and 0.7 for San Diego satisfies component 1.
- A response that constructs a segmented bar graph with the lengths of the segments representing the relative frequencies of teens who did not consume a soft drink in the previous week between 0.15 and 0.25 for Baltimore, between 0.2 and 0.3 for Detroit, and between 0.3 and 0.4 for San Diego satisfies component 1.
- Incorrect proportions imported from part (a) may be used to satisfy component 1.
- Segmented bar graphs with more than two segments cannot satisfy either component 1 or component 2.
- Labels of “Yes” and “No” may satisfy component 2.
- A response to part (b-ii) that is consistent with an incorrect graph in (b-i) may satisfy components 3 and 4.

© 2021 College Board

Model Solution	Scoring
(c) (i) Since the data were collected from independent random samples from the three cities, a chi-square test for homogeneity should be conducted. **(ii)** The appropriate hypotheses are: H_0: There is no difference in the proportion of all teens who consumed a soft drink in the past week across the three cities. H_a: There is at least one difference in the proportion of all teens who consumed a soft drink in the past week across the three cities. *OR* H_0: The proportion of all teens who consumed a soft drink in the past week is the same across the three cities. H_a: The proportion of all teens who consumed a soft drink in the past week differs for at least two of the three cities.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies a chi-square test for homogeneity by name in part (c-i) 2. States the null hypothesis to imply homogeneous (or equal) proportions AND states the alternative hypothesis to imply that at least two proportions are not the same in part (c-ii) 3. Provides sufficient context for at least one of the hypotheses in part (c-ii) by including the parameters of interest (proportion of teens who consumed a soft drink) AND the populations (cities) **Partially correct (P)** if the response satisfies component 1 and only one of the other two components *OR* if the response identifies a "chi-square test" in part (c-i) by name or formula AND satisfies component 2. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that identifies two different tests is considered parallel solutions and the weaker solution is used when scoring component 1.
- Component 1 is not satisfied by the test statistic formula for a chi-square test unless the response includes "Homogeneity."
- Component 1 is not satisfied if the response presents a test statistic formula that is inconsistent with a chi-square test of homogeneity, even if the response identifies a chi-square test of homogeneity by name.
- A response that states the hypotheses in terms of distributions rather than proportions (e.g., H_0: There is no difference in distributions of teens who consumed or did not consume a soft drink in the past week across the three cities) satisfies component 2.
- A response that states either the null hypothesis or the alternative hypothesis by referring to sample proportions does not satisfy component 2.
- A response that uses symbols to describe the hypotheses must clearly identify the parameters in context (proportion of teens who consumed a soft drink) AND reference the populations (cities) in order to satisfy component 3.
- Any attempt to check test conditions should be ignored in scoring.
- Any discussion of the degrees of freedom for the test should be ignored in scoring.

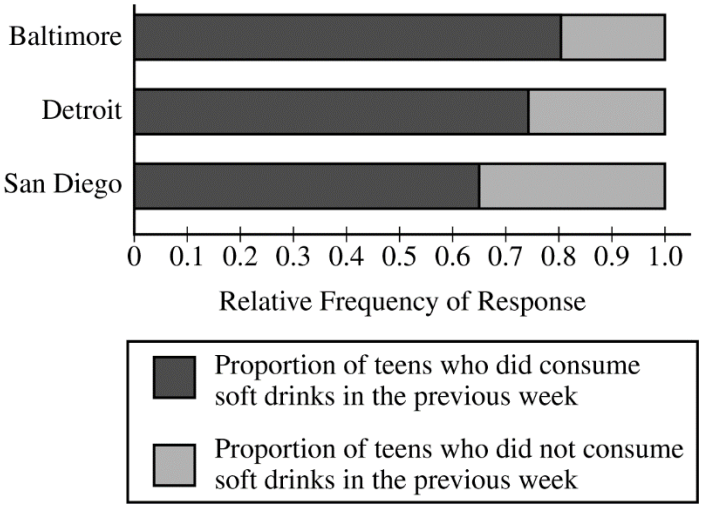
© 2021 College Board

Scoring for Question 5	Score
Complete Response Three parts essentially correct	**4**
Substantial Response Two parts essentially correct and one part partially correct	**3**
Developing Response Two parts essentially correct and no part partially correct *OR* One part essentially correct and one or two parts partially correct *OR* Three parts partially correct	**2**
Minimal Response One part essentially correct and no part partially correct *OR* No part essentially correct and two parts partially correct	**1**

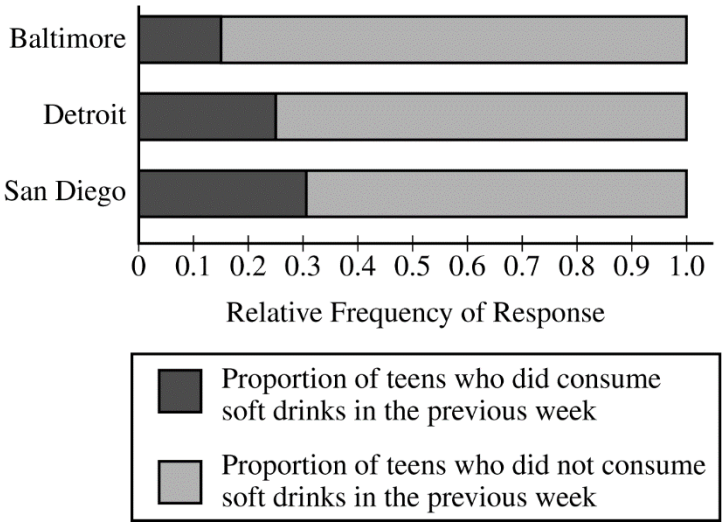
© 2021 College Board

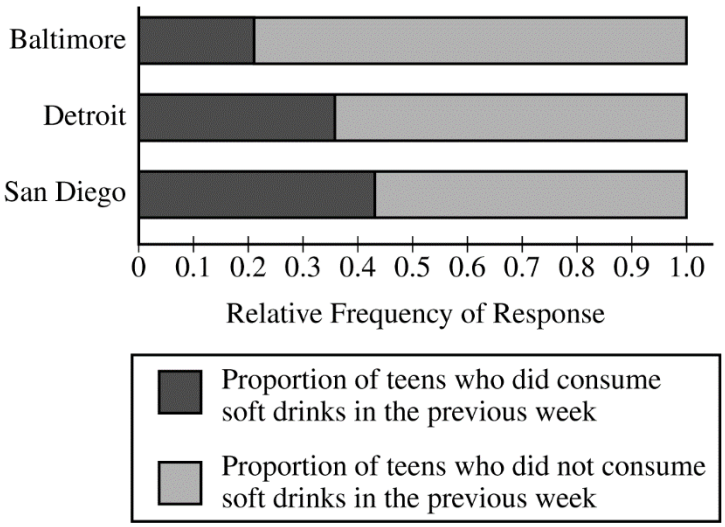
Common acceptable and unacceptable graphs for part (b-i)

Acceptable Graph



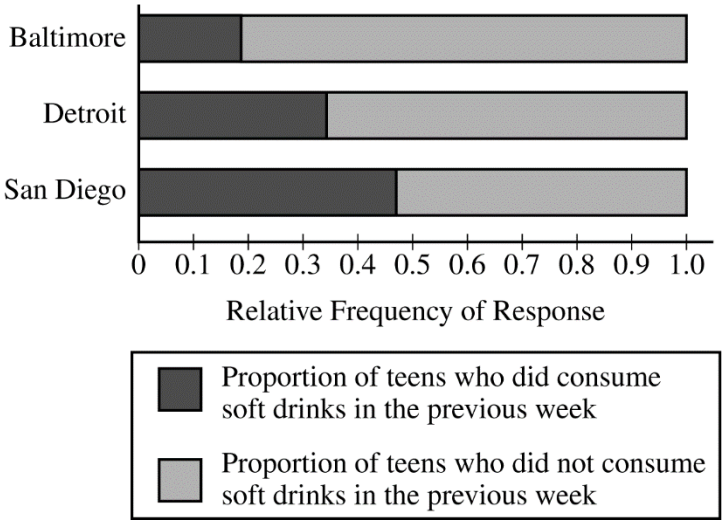
Common Unacceptable Graphs





© 2021 College Board

Common Unacceptable Graphs (continued)



© 2021 College Board

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

5. A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847

(a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.

The researcher's claim is incorrect. Just because the count of Baltimore teens who consumed a soft drink in the past week doesn't mean it is less likely to select a teen from Baltimore who answered yes than the other cities. In reality, the likelihood of a Baltimore teen answering yes is $\frac{727}{904} = 0.804$, which is higher than Detroit ($\frac{1232}{1663} = 0.741$) and San Diego ($\frac{1482}{2280} = 0.650$). There were differing numbers of teens surveyed in the three cities, so the numbers cannot be compared directly; they must be proportions first.

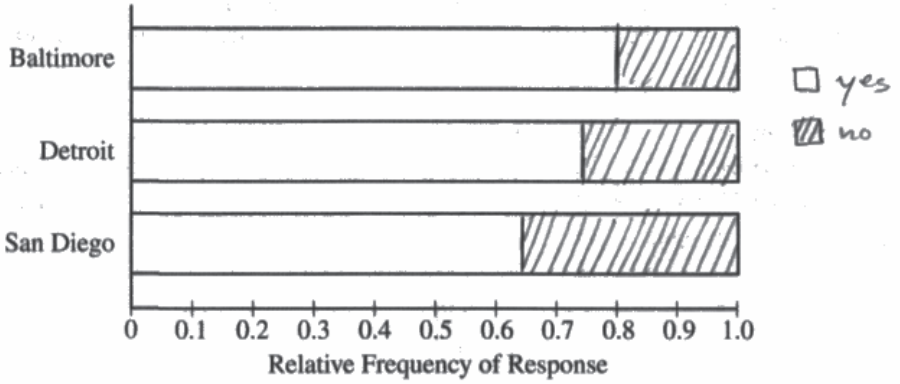
Unauthorized copying or reuse of this page is illegal.

GO ON TO THE NEXT PAGE.

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

(b) Consider the values in the table.

(i) Construct a segmented bar chart of relative frequencies based on the information in the table.



(ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.

San Diego $\rightarrow p = \frac{1482}{2280} = 0.65$

(c) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.

(i) Identify the appropriate inference procedure.

The appropriate inference procedure is a χ^2-test for homogeneity.

(ii) Identify the hypotheses of the test.

H_0: All three cities have the same proportion of teens who consumed a soft drink in the past week.

H_a: The proportion of teens who had a soft drink in the past week differs among the three cities.

Unauthorized copying or reuse of this page is illegal.

GO ON TO THE NEXT PAGE.

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

5. A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847

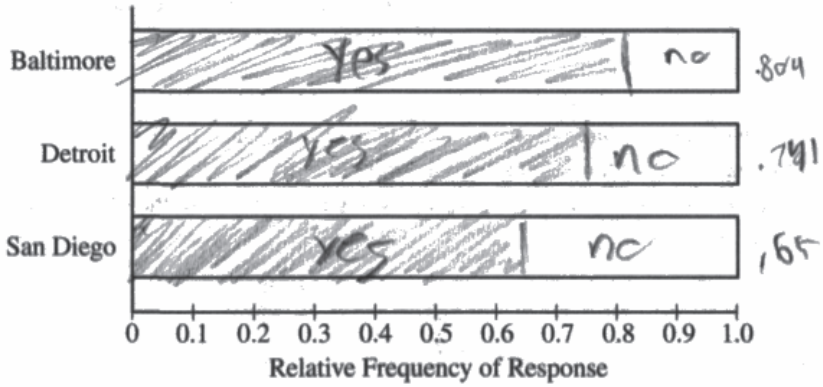
(a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.

The researcher's claim is incorrect, because the total number of teens in the survey was much less in Baltimore than it was in San Diego and Detroit, [illegible]

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

(b) Consider the values in the table.

(i) Construct a segmented bar chart of relative frequencies based on the information in the table.



(ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.

San diego had the smallest proportion of teens who consumed a soft drink in the previous week, with a proportion of .65 saying yes

(c) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.

(i) Identify the appropriate inference procedure.

χ^2 test

(ii) Identify the hypotheses of the test.

~~If we obtain the two~~

h_0: there is no difference in the proportions ~~of~~ of teens who drank soft drinks (at least once last week) in the 3 cities.

h_a: there is a difference in the proportions of teens who drank soft drinks in the last week in the 3 cities.

Unauthorized copying or reuse of this page is illegal.

GO ON TO THE NEXT PAGE.

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

5. A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847

(a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.

χ^2 test of homogeneity

Conditions
- Data of teens consumption of soft drinks are counts
- SRSs of teenagers soft drink consumption in Baltimore Detroit, and San Diego
- Expected counts all ≥ 5
conditions met, appropriate to approx. with χ^2 test of homogeneity

Hypothesis Statements
H_0: The consumption of soft drinks of teenagers is the same in Baltimore, Detroit, and San Diego
H_A: The consumption of soft drinks of teenagers in at least one city is different.

Expected counts

	B	D	S	
Y	641.77	1180.16	1618.63	3441
N	262.23	482.84	661.37	1406
	904	1663	2280	4847

$\chi^2 = \frac{(727-641.77)^2}{641.77} + \frac{(1232-1180.16)^2}{1180.16} + \frac{(798-661.37)^2}{661.37}$

$\chi^2 = 86.4996$

p-value = χ^2 CDF $86.4996, 10^{10}, 2$
$= 1.65582 \times 10^{-19}$

The researchers claim is not correct. Baltimore has a higher proportion of teenagers that consumed the soft drink, though its absolute number is lower.

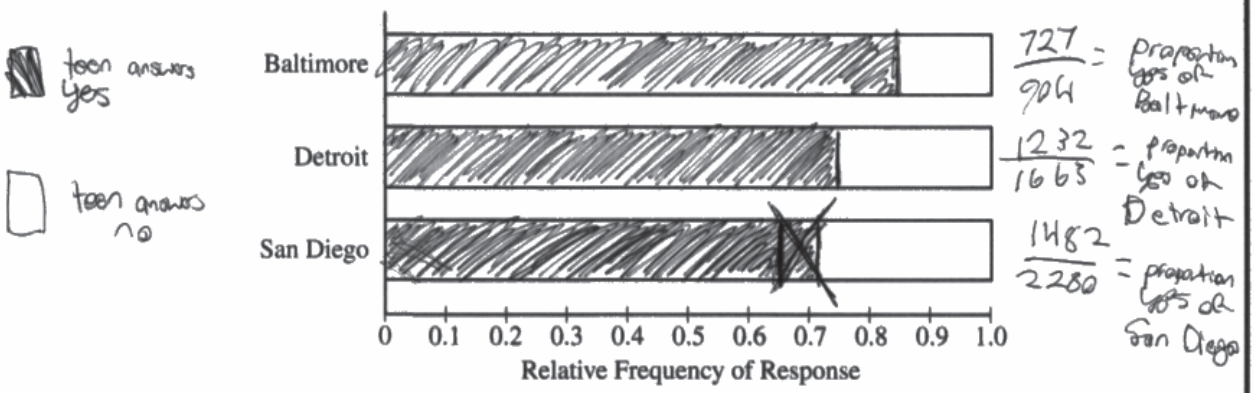
Unauthorized copying or reuse of this page is illegal.

GO ON TO THE NEXT PAGE.

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

(b) Consider the values in the table.

(i) Construct a segmented bar chart of relative frequencies based on the information in the table.



(ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.

San Diego had the smallest proportion of teens who consumed in the previous week, with a proportion of 0.65 that consumed a soft drink in the past week being equal to $\frac{1482}{2280}$.

(c) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.

(i) Identify the appropriate inference procedure.

χ^2 test of homogeneity

(ii) Identify the hypotheses of the test.

H_0: The consumption of soft drinks of teenagers is the same in Baltimore, Detroit, and San Diego

H_A: The consumption of soft drinks of teenagers in at least one city is different

Unauthorized copying or reuse of this page is illegal.

GO ON TO THE NEXT PAGE.

Question 5

Note: Student samples are quoted verbatim and may contain spelling and grammatical errors.

Overview

The primary goals of this question were to assess a student's ability to (1) recognize whether comparisons between samples should be based on proportions instead of counts when sample sizes are different; (2) identify appropriate proportions to compute from a table of counts; (3) construct and label a segmented bar chart; (4) use a segmented bar chart to make a comparison; (5) identify an appropriate inference procedure for investigating whether the distribution of a categorical random variable differs across populations; and (5) identify the null and alternative hypotheses for a chi-square test of homogeneity.

This question assesses skills in multiple skill categories, including skill category 1: Selecting Statistical Methods; skill category 2: Data Analysis; and skill category 4: Statistical Argumentation. Skills required for responding to this question include (1.E) Identify an appropriate inference method for significance tests, (1.F) Identify null and alternative hypotheses, (2.B) construct numerical or graphical representations of distributions, (2.D) compare distributions or relative positions of points within a distribution, and (4.B) Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from multiple units, including Unit 1: Exploring One-Variable Data, Unit 2: Exploring Two-Variable Data, and Unit 8: Inference for Categorical Data: Chi-Square of the course framework in the AP Statistics Course and Exam Description. Refer to topics 1.4, 2.2. 2.3, and 8.5, and learning objectives UNC-1.C UNC-1.P, UNC-1.R, VAR-8.I, and VAR-8.J.

Sample: 5A
Score: 4

The response earned the following: part (a) – E; part (b) – E; part (c) – E.

In part (a) the response indicates the claim is incorrect, satisfying component 1; proportions are shown and explicitly compared by the phrase "is higher than Detroit," satisfying component 2. Part (a) was scored essentially correct (E).

In part (b-i) the response correctly segments and labels the bars, satisfying components 1 and 2. In part (b-ii) the response correctly indicates San Diego and states the correct proportion, satisfying components 3 and 4. Part (b) was scored essentially correct (E).

In part (c-i) the response correctly identifies the chi-square test of homogeneity by name, satisfying component 1. In part (c-ii) the response correctly states the hypotheses for a chi-square test of homogeneity in words and provides sufficient context, satisfying components 2 and 3. Part (c) was scored essentially correct (E).

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Question 5 (continued)

Sample: 5B
Score: 2

The response earned the following: part (a) – P; part (b) – E; part (c) – P.

In part (a) the response indicates that the researcher's claim is incorrect, satisfying component 1. The response indicates that the sample sizes are not equal using the phrase "the total number of teens in the survey was much less in baltimore," but does not provide proportions; thus, component 2 is not satisfied. Because component 1 is satisfied, and the response indicates sample sizes are not equal, part (a) was scored partially correct (P).

In part (b-i) the response correctly segments and labels the segments, satisfying components 1 and 2. In part (b-ii) the response correctly indicates San Diego and states the correct proportion, satisfying components 3 and 4. Part (b) was scored essentially correct (E).

In part (c-i) the response identifies a "χ^2 test," which does not satisfy component 1 but does identify the test as chi-square; the response correctly states both hypotheses for a chi-square test of homogeneity, with context, satisfying components 2 and 3. Part (c) was scored partially correct (P).

Sample: 5C
Score: 1

The response earned the following: part (a) – I; part (b) – E; part (c) – I.

In part (a) the response indicates the claim is not correct, satisfying component 1; unequal sample sizes are not mentioned, and correct proportions are not shown. Thus component 2 is not satisfied. The chi-square test of homogeneity is ignored. Part (a) was scored incorrect (I).

In part (b-i) the response correctly segments and labels the segments, satisfying components 1 and 2. In part (b-ii) the response correctly indicates San Diego and states the correct proportion, satisfying components 3 and 4. Part (b) was scored essentially correct (E).

In part (c-i) the response correctly identifies the chi-square test of homogeneity by name, satisfying component 1. In part (c-ii) the response does not include the correct hypotheses for a chi-square test of homogeneity or the parameter of interest (the proportion of teens who consumed a soft drink); thus, components 2 and 3 are not satisfied. Part (c) was scored incorrect (I).

© 2021 College Board.
Visit College Board on the web: collegeboard.org.