

AP[®] STATISTICS 2016 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question and (2) assess the effectiveness of treatments in a controlled experiment.

Solution

Part (a):

Step 1: States a correct pair of hypotheses.

H_0 : The proportion of children who would choose each snack is the same regardless of which type of ad is viewed.

H_a : The proportion of children who would choose each snack differs based on which type of ad is viewed.

Step 2: Identifies a correct test procedure (by name or formula) and checks appropriate conditions.

The appropriate procedure is a chi-square test of homogeneity.

The conditions for this test are satisfied because (1) the question states that the children were randomly assigned to groups, and (2) expected counts for the six cells of the table are all at least 5, as seen in the following table that lists expected counts beside observed counts.

Group	Choco-Zuties	Apple-Zuties	Total
A	21 (18.67)	4 (6.33)	25
B	13 (18.67)	12 (6.33)	25
C	22 (18.67)	3 (6.33)	25
Total	56	19	75

Step 3: Calculates the appropriate test statistic and p -value.

The test statistic is calculated as $\chi^2 = \sum \frac{(O - E)^2}{E}$, which is

$$\begin{aligned}\chi^2 &\approx \\ &0.292 + 0.860 + \\ &1.720 + 5.070 + \\ &0.595 + 1.754 \approx 10.291.\end{aligned}$$

The p -value is $P(\chi^2_{df2} \geq 10.291) \approx 0.006$.

Step 4: States a correct conclusion in the context of the study, using the result of the statistical test.

Because the p -value is very small (for instance, much smaller than $\alpha = 0.05$), we reject the null hypothesis at the 0.05 level (and at the 0.01 level). The data provide convincing statistical evidence that the proportions who would choose each snack differ based on which ad is viewed.

AP[®] STATISTICS

2016 SCORING GUIDELINES

Question 2 (continued)

Part (b):

When neither ad was viewed, $\frac{22}{25}$ or 88 percent of the children chose Choco-Zuties whereas only 12 percent chose Apple-Zuties.

When the Choco-Zuties ad was viewed, 84 percent of the children chose Choco-Zuties, which was very similar to the percentage that chose them without viewing any ad. So watching the Choco-Zuties ad did not affect the snack choice very much.

When the Apple-Zuties ad was viewed, only $\frac{13}{25}$ or 52 percent of the children chose Choco-Zuties, and 48 percent chose Apple-Zuties. Watching the Apple-Zuties ad seemed to increase the proportion of children choosing Apple-Zuties.

Scoring

This question is scored in four sections. Section 1 consists of steps 1 and 2 in part (a); section 2 consists of step 3 in part (a); section 3 consists of step 4 in part (a); and section 4 consists of part (b). Sections 1, 2, 3 and 4 are scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response includes the following three components:

1. Both hypotheses are stated correctly with at least one in context.
2. Identifies the correct test procedure (by name or by formula).
3. The technical conditions are checked (all expected counts are greater than or equal to 5).

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response correctly includes at most one of the three components.

Notes for component 1:

- It is acceptable if hypotheses are stated in terms of association/independence instead of proportions.
- Context must be included in at least one of the hypotheses, but is not required to be in both.
- If the hypotheses contain language that suggests that the response refers to the sample data, the component is not satisfied.

Note for component 2: It is acceptable if the test is identified as a test of independence instead of a test of homogeneity.

Notes for component 3:

- The random assignment condition was stated so need not be explicitly mentioned.
- Stating the expected count condition is met is not sufficient for component 3. The condition must be checked by reporting the expected counts and either:
 - noting that all are greater than or equal to 5;

OR

AP[®] STATISTICS

2016 SCORING GUIDELINES

Question 2 (continued)

- noting that all are greater than or equal to 1 and at most 20 percent are less than 5.
- Noting that the smallest expected count is 6.33 and that it is greater than or equal to 5 is sufficient to satisfy this component.
- Component 3 is not satisfied if the expected counts are reported as integers.
- This component is not satisfied if the response includes any of the following inappropriate conditions:
 - The response implies that a random sample was taken, e.g., “SRS – check.”
 - The response refers to independence of groups or independence of ads as a required condition.
 - The response indicates that a sample size greater than 30 ensures normality or the response implies normality as a condition.
- A response stating that children are independent can be ignored in the scoring of this component.

Section 2 is scored as follows:

Essentially correct (E) if the response includes the following three components:

1. Correct chi-square test-statistic
2. Correct degrees of freedom
3. Correct p -value

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response correctly includes at most one of the three components.

Notes:

- If the response makes an error in one calculation, future calculations are considered correct if they follow correctly from the initial miscalculation.
- A chi-square critical value approach is acceptable: The critical value for $\alpha = 0.05$ is $\chi^2 = 5.99$.

Section 3 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion about the alternative hypothesis in context, *AND* provides justification based on linkage between the p -value and conclusion.

Partially correct (P) if the response provides a correct conclusion in context, but without justification based on linkage to the p -value;

OR

if the response provides a correct conclusion, with linkage to the p -value, but not in context;

OR

if the response provides a correct decision stated in terms of the null hypothesis in context, with linkage to the p -value, but no conclusion is made about the alternative hypothesis.

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

AP[®] STATISTICS

2016 SCORING GUIDELINES

Question 2 (continued)

Notes:

- If the conclusion is consistent with the p -value from section 2, and also in context with justification based on linkage to the p -value, section 3 is scored as E.
- If no alpha level is given, the solution must be explicit about the linkage by giving a correct interpretation of the p -value or by explaining how the conclusion follows from the p -value such as saying: “Because the p -value is small, we reject the null hypothesis.” or “Because the p -value is large, we do not reject the null hypothesis.”
- If a conclusion contains language that suggests that the response refers to the sample data, the conclusion component is not correct, unless the same error occurred in the statement of hypotheses in section 1.

Section 4 is scored as follows:

Essentially correct (E) if the response, in context, concludes that ad type A had little effect and ad type B had an effect, both supported by the observed proportions (or counts) from the table.

Partially correct (P) if the response correctly concludes that ad type A had little effect, and ad type B had an effect but does not provide correct numerical justification;

OR

if the response compares all of the proportions (or counts) as required, but without correctly concluding ad effectiveness;

OR

if the response correctly describes the effect of one of the ad types A or B (in context, with correct numerical justification) but not the other;

OR

if the response compares all of the proportions (or counts) as required, but not in context.

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

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section counts as $\frac{1}{2}$ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, $2\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and communication.

2. Product advertisers studied the effects of television ads on children’s choices for two new snacks: The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
A	Choco-Zuties only	21	4
B	Apple-Zuties only	13	12
C	Neither	22	3

- (a) Do the data provide convincing statistical evidence that there is an association between type of ad and children’s choice of snack among all children similar to those who participated in the experiment?

H_0 : There is no association between type of ad and children’s choice of snack.

H_A : There is an association between type of ad and children’s choice of snack.

$\alpha = 0.05$

Conditions

- Reasonable to assume that the 75 children are representative of all children similar to those who participated in the experiment.

- All expected counts are greater than 5 (see table).

Expected counts

	Choco	Apple
A	18.667	6.333
B	18.667	6.333
C	18.667	6.333

Since all conditions are met, I will perform a chi-square test for independence.

$$\chi^2 = \frac{(21 - 18.667)^2}{18.667} + \dots + \frac{(3 - 6.333)^2}{6.333} = 10.2914$$

df = 2

$p(\chi^2 > 10.2914) = 0.00582$

2A2
of 2

If you need more room for your work in part (a), use the space below.

2A2
of 2

Since $p < \alpha$, we reject H_0 , and find sufficient evidence to support the hypothesis that there is an association between type of ad and children's choice of snack.

(b) Write a few sentences describing the effect of each ad on children's choice of snack.

The Choco-Zuties ad seemed to have little effect on children's choice of snack. The children who watched the Choco-Zuties ad chose the sugary snack 1 less time than those who watched neither ad, which is not very significant.

The Apple-Zuties ad had a much larger effect. Of the children who watched the ad for Apple-Zuties, 10 more chose Apple-Zuties than the children who watched neither ad.

2. Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

$xp = \frac{\text{row total} \times \text{column total}}{\text{table total}}$

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties	Total
A	Choco-Zuties only	21	4	25
B	Apple-Zuties only	13	12	25
C	Neither	22	3	25
Total		56	19	75

(a) Do the data provide convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

H_0 : There is no association between the type of ad and children's choice of snack.

H_a : There is an association between the type of ad and children's choice of snack.

$\alpha = 0.05$

I will conduct a Chi Square Test of Association

Conditions: Random? 75 children were randomly assigned to one of three groups

Independent? There are at least 10(75) = 750 children nationwide.

Large Counts? All expected counts are greater than 5. (see chart above)

$df = (row-1)(column-1) = (3-1)(2-1) = 2$

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} = \frac{(21 - 18.667)^2}{18.667} + \frac{(4 - 6.333)^2}{6.333} \dots = 10.29$$

Table C says the p -value ranges from 0.01 to 0.005.

Since our p -value is less than our α of 0.05, I reject the null hypothesis. There is convincing evidence that there is an association between the type of ad and children's choice of snack.

(b) Write a few sentences describing the effect of each ad on children's choice of snack.

In group A, kids were advertised Choco-Zuties and 84% of them chose Choco-Zuties because of the ad. In group B, kids were advertised Apple-Zuties and a significant increase was seen in those who chose Apple-Zuties compared to group A. Almost 50% in group B chose Apple-Zuties, thanks to the ad. When neither ad was shown, 88% chose Choco-Zuties which show children's natural inclination towards sweets.

2C1
of 2

2C1
of 2

2. Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in an experiment. One ad was for a new sugary snack called Choco-Zuties, and the other ad was for a new healthy snack called Apple-Zuties.

For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad nor the Apple-Zuties ad.

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices.

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
A	Choco-Zuties only	21 / 18.07	4 / 6.3
B	Apple-Zuties only	13 / 18.07	12 / 6.3
C	Neither	22 / 18.07	3 / 6.3
		56	19

(a) Do the data provide convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

H_0 : The null hypothesis is that
 H_a : the type of ad has no affect χ^2 test of homogeneity
 on what snacks the children choose. The alternative hypothesis is
 that the type of ad shown does have an effect on what snack the children choose. Let $\alpha = .05$
 Conditions: random-stated
 Independent-one sample
 child won't affect the other since they were separated.
 Counted Data greater than 5: yes
 $\chi^2 = 10.29$ Since P -value is less than α reject H_0 : There is enough evidence to suggest that the type of ad the children watched affected what snack they choose
 P -val = .006
 $df = 2$
 $\frac{56.25}{75}$ $\frac{19.75}{75}$

Unauthorized copying or reuse of any part of this page is illegal.

GO ON TO THE NEXT PAGE.

202
0F2

If you need more room for your work in part (a), use the space below.

202
0F2

(b) Write a few sentences describing the effect of each ad on children's choice of snack.

According to the test there is enough evidence to suggest that the type of ad the child watched has an effect on what snacks they choose. However we still have to take into account the confounding variables. It is possible that some children just really like chocolate for example.

Unauthorized copying or reuse of
any part of this page is illegal.

GO ON TO THE NEXT PAGE.

AP[®] STATISTICS

2016 SCORING COMMENTARY

Question 2

Overview

The primary goals of this question were to assess a student's ability to (1) identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question and (2) assess the effectiveness of treatments in a controlled experiment.

Sample: 2A

Score: 4

All three of the required components are present and correct for section 1. The response includes acceptable statements of hypotheses in terms of association between type of ad and children's choice of snack and indicates that a chi-square test for independence will be performed. A table of expected counts is provided, and a check that all are greater than 5 is made. The statement "Reasonable to assume that the 75 children are representative of all children similar to those who participated in the experiment" is considered extraneous. Section 1 was scored as essentially correct. All three of the required components are present and correct for section 2. The value of the chi-square statistic is correctly calculated, and the correct degrees of freedom and p -value are reported. (The p -value is identified by an acceptable probability statement.) Section 2 was scored as essentially correct. The statement $p < \alpha$ establishes linkage between the previously calculated p -value and the chosen level of α given in the response. A correct decision and a correct conclusion in context are provided. Section 3 was scored as essentially correct. In section 4 the response correctly states that the Choco-Zuties ad seemed to have little effect by noting that the difference between the number in the Choco-Zuties ad group and the number in the neither ad group is just one child. Establishing lack of ad effectiveness for group A because each group size is the same (25) is acceptable. By a similar argument, it is correctly stated that the Apple-Zuties ad had a much larger effect, noting that 10 more children in this group chose the Apple-Zuties snack than chose Apple-Zuties in the neither ad group. The value 10 here is incorrect; the correct value is 9. However, this is considered a minor arithmetic error that is overlooked. Thus the effectiveness of each type of ad is correctly addressed with numerical justification. Section 4 was scored as essentially correct. Because all four sections were scored as essentially correct, the response earned a score of 4.

Sample: 2B

Score: 3

All three of the required components are present and correct for section 1. The response includes acceptable statements of hypotheses in terms of association and indicates that a chi-square test for association will be conducted. (The test is also correctly identified by formula). Correct expected counts are added to the table in the problem and noted as all being greater than 5. Though not required, the response also indicates that random assignment occurred. An additional condition labeled "Independent?" was considered extraneous information. Section 1 was scored as essentially correct. All three of the required components are present and correct for section 2. The value of the chi-square test statistic and the degrees of freedom are both correctly calculated. Correct upper and lower bounds on the p -value are obtained from Table C of the exam. Section 2 was scored as essentially correct. A correct conclusion is provided in context with justification based on linkage between the p -value and the chosen α level given in the response. Section 3 was scored as essentially correct. In section 4 a comparison is made between group A (Choco-Zuties ad type) and group B (Apple-Zuties ad type), but this does not establish ad effectiveness or ineffectiveness. The response includes a correct assessment of ad effectiveness of the Apple-Zuties ad type with numerical justification by comparing the percentage of children who chose Choco-Zuties in group B to the corresponding percentage in group C. Section 4 was scored as partially correct. Because three sections were scored as essentially correct and one section was scored as partially correct, the response was initially scored as a $3\frac{1}{2}$.

AP[®] STATISTICS

2016 SCORING COMMENTARY

Question 2 (continued)

Looking holistically at the response, the absence of any assessment of the Choco-Zuties ad effectiveness in section 4 was viewed as a weakness. Therefore the response earned a score of 3.

Sample: 2C

Score: 2

All three of the required components are present and correct for section 1. The response includes correct null and alternative hypothesis statements for a test of homogeneity and correctly indicates that a chi-square test of homogeneity will be conducted. Note that traditional labels for hypotheses H_0 and H_a are included in the response but not used to identify the individual hypothesis statements. Correct expected counts (unlabeled) are added to the table in the problem, with unlabeled work shown on the page. The statement “Counted Data greater then [sic] 5: yes” is considered to be referencing the calculated expected counts given its proximity to the work shown. Section 1 was scored as essentially correct. All three of the required components (the value of the chi-square test statistic, the degrees of freedom, and the p -value) are present and correct for section 2, and this section was scored as essentially correct. In section 3 a statement comparing the p -value to the chosen α level given in the response establishes linkage. The correct decision to reject the null hypothesis is stated. A conclusion is provided in context. However, the choice of wording “the type of ad the children watched affected what snack they choose,” suggests that the response refers to the sample data; therefore, it does not satisfy the requirements for a correct conclusion. Because linkage, a correct decision, and context is included in the response, section 3 was scored as partially correct. In section 4 the response refers to the results of the hypothesis test performed in part (a) of the problem, but it does not compare children’s snack preference between ad types. Other comments in the response are considered extraneous to the question. Because ad effectiveness is not correctly addressed, section 4 was scored as incorrect. Because two sections were scored as essentially correct, one section was scored as partially correct, and one section was scored as incorrect, the response was initially scored as a 2½. Looking holistically at the paper, the lack of a correct response to part (b), together with weak communication concerning the technical conditions in section 2, was viewed as a weakness. Therefore the response earned a score of 2.