4. The Behavioral Risk Factor Surveillance System is an ongoing health survey system that tracks health conditions and risk behaviors in the United States. In one of their studies, a random sample of 8,866 adults answered the question "Do you consume five or more servings of fruits and vegetables per day?" The data are summarized by response and by age-group in the frequency table below.

Age-Group (years)	Yes	No	Total
18–34	231	741	972
35–54	669	2,242	2,911
55 or older	1,291	3,692	4,983
Total	2,191	6,675	8,866

Do the data provide convincing statistical evidence that there is an association between age-group and whether or not a person consumes five or more servings of fruits and vegetables per day for adults in the United States?

5. Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation.

The researchers wanted to perform a hypothesis test of

$$H_0 : p_m - p_c = 0$$
$$H_a : p_m - p_c < 0,$$

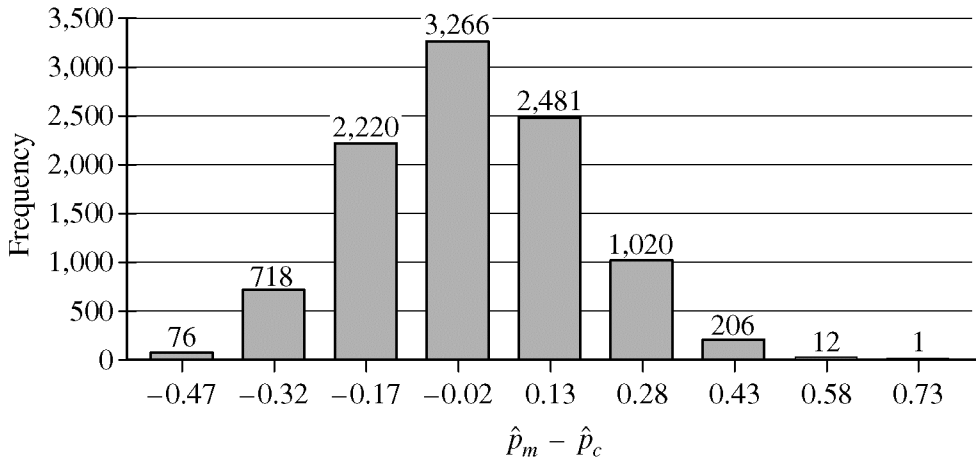
where p_m is the proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation and p_c is the proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation.

(a) If the study were to provide significant evidence against H_0 in favor of H_a, would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

(b) Let $\hat{p}_m$ represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let $\hat{p}_c$ represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to use a normal approximation for the sampling distribution of $\hat{p}_m - \hat{p}_c$?

Although a normal approximation cannot be used, it is possible to simulate the distribution of $\hat{p}_m - \hat{p}_c$. Under the assumption that the null hypothesis is true, 10,000 values of $\hat{p}_m - \hat{p}_c$ were simulated. The histogram below shows the results of the simulation.



(c) Based on the results of the simulation, what can be concluded about the relationship between blood pressure and meditation among men in the retirement community?

AP® STATISTICS
2013 SCORING GUIDELINES

Question 4

Intent of Question

The primary goal of this question was to assess students' ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and *p*-value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

Solution

Step 1: States a correct pair of hypotheses.

The null hypothesis is that fruit and vegetable consumption is independent of (that is, it is not associated with) age group for the population of adults in the United States.

The alternative hypothesis is that fruit and vegetable consumption is not independent of (that is, it is associated with) age group for the population of adults in the United States.

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

The appropriate test is a chi-square test of independence.

The conditions for this test were satisfied because:
1. The question states that the sample was randomly selected.
2. The expected counts for all six cells of the table were all at least 5, as seen in the following table that lists expected counts in parentheses beside the observed counts:

	Five or more servings of fruit and vegetables	Four or fewer servings of fruit and vegetables	Total
18–34 years	231 (240.2)	741 (731.8)	972
35–54 years	669 (719.4)	2242 (2191.6)	2911
55+ years	1291 (1231.4)	3692 (3751.6)	4983
Total	2191	6675	8866

Step 3: Correct mechanics, including the value of the test statistic and *p*-value (or rejection region).

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

$$\chi^2 = 0.353 + 0.116 + 3.528 + 1.158 + 2.883 + 0.946 = 8.983.$$

The *p*-value is $P(\chi^2 \geq 8.983) = 0.011$, based on $(3-1)\times(2-1) = 2$ degrees of freedom.

Question 4 (continued)

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 would reject the null hypothesis at the 0.05 level and conclude that the sample data provide strong evidence that there is an association between age group and consumption of fruits and vegetables for adults in the United States. In particular, older (55+ years of age) people were more likely to eat five or more servings of fruits and vegetables, and middle-aged people (35–54 years of age) were less likely to eat five or more servings of fruits and vegetables.

Scoring

Each of steps 1, 2, 3, and 4 were scored as essentially correct (E), partially correct (P), or incorrect (I).

Step 1 is scored as follows:

Essentially correct (E) if the response correctly states both hypotheses with at least one in context.

Partially correct (P) if the response correctly states both hypotheses but not in context,
OR
the hypotheses were reversed with at least one stated in context.

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

Notes:

- If the hypotheses contain language that suggests that the response refers to the sample data, step 1 is scored as incorrect (I).

- If the null and alternative hypotheses were correctly stated in terms of population proportions, step 1 is scored essentially correct (E):
  $H_0 : p_1 = p_2 = p_3$, where p_i is the population proportion of adults at the indicated age group, 1, 2, or 3, who eat at least 5 servings of fruits and vegetables.

  H_a : At least one of the population proportions, p_1, p_2, p_3, differs from the other two
  OR
  H_a : The population proportions for the three age groups were not the same

Step 2 is scored as follows:

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

1. Identifies a chi-square test of independence by name or by formula for the chi-square test statistic.
2. States *AND* verifies the random sampling condition.
3. States *AND* verifies the technical condition that all expected counts were greater than 5.

Question 4 (continued)

Partially correct (P) if the response correctly includes two of the three components listed above.

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

Notes:

- If the response identifies the test procedure as a chi-square test of homogeneity of proportions, step 2 does not receive credit for component 1.
- If the response identifies the correct test procedure but gives an incorrect formula for the test statistic, then this is considered a contradiction and does not meet the criteria for component 1.
- Stating the condition that the expected counts must be greater than 5 is not in itself sufficient for satisfying component 3; the condition must be checked by reporting expected counts, or minimally reporting the value of the smallest expected count and indicating that it is at least 5.
- If the response includes an incorrect technical condition, such as "$n \geq 30$" or "normality," then this will be considered a parallel solution and credit will not be granted for component 3.
- If the response states and verifies the condition that 80 percent of all expected counts must be ≥ 5 and all expected counts must be ≥ 1, then the response can receive credit for component 3.

Step 3 is scored as follows:

Essentially correct (E) if the response correctly calculates the following two components:

1. Test statistic
2. *p*-value or critical value

Partially correct (P) if the response correctly calculates one of the two components listed above.

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

Notes:

- When a response has an error in one calculation, future calculations were considered correct if they follow correctly from the initial miscalculation.
- The correct critical value is 5.99 for a significance level of 0.05.

Step 4 is scored as follows:

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

1. A correct conclusion
2. Justification of the conclusion based on linkage between the *p*-value and conclusion (or linkage between test statistic and critical value)
3. The conclusion is stated in context

Question 4 (continued)

Partially correct (P) if the response includes a correct conclusion and one of the other two components listed above.

Incorrect (I) if the response does not provide a correct conclusion or is missing both linkage and context.

Notes:

- If the conclusion is consistent with an incorrect *p*-value from step 3, stated in context, and has justification based on linkage to the *p*-value, then step 4 is scored as E.
- To meet the criteria for component 1, the conclusion must be consistent with the hypotheses.
- If the response does not have linkage between the *p*-value and conclusion, then in order to meet the criteria for component 1, the conclusion must be correct based on a 0.05 level of significance.
- If no significance level, α, is given, then in order to meet the criteria for the linkage component, the solution must explain 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."
- A *p*-value interpretation, in terms of the probability of obtaining a sample result as or more extreme as the observed result if the null hypothesis were true, does not in itself meet the criteria for the linkage component. A response with an incorrect interpretation of the *p*-value lowers the score one level (that is, from E to P, or from P to I)
- A response that states the alternative hypothesis has been proven lowers the score one level (that is, from E to P, or from P to I)
- A response including incorrect statistical language lowers the score one level (that is, from E to P, or from P to I)

Each essentially correct (E) step counts as 1 point, and a partially correct (P) step in counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2½ points), use a holistic approach to determine whether to score up or down, depending on the strength of the response and communication.