Question 5

Intent of Question

The primary goals of this question were to assess students' ability to (1) calculate appropriate probabilities, including conditional probabilities, from a two-way table; (2) determine from a two-way table whether two events are independent; (3) identify an appropriate test procedure for assessing independence between two categorical variables.

Solution

Part (a):

Using the addition rule, the probability that the randomly selected adult is a college graduate or obtains news primarily from the internet is:

P(college graduate or internet) = P(college graduate) + P(internet) – P(college graduate and internet)

$$= \frac{693}{2500} + \frac{687}{2500} - \frac{245}{2500} = \frac{1135}{2500} = 0.454 \, .$$

Part (b):

Reading values from the table, the conditional probability that the selected adult receives news primarily from the internet given that he or she is a college graduate is: $\frac{245}{693} = 0.354$.

Part (c):

These events are not independent. One way to establish this is to note that the unconditional probability equals $P(\text{obtains news primarily from the internet}) = \frac{687}{2500} = 0.275$, but the conditional probability equals $P(\text{obtains news primarily from the internet}/\text{is a college graduate}) = 0.354$. Because these two probabilities are not equal, the events "is a college graduate" and "obtains news primarily from the internet" are not independent.

Part (d):

Chi-square test of association (or independence), with
$\text{degrees of freedom} = (\# \text{ of rows} - 1) \times (\# \text{ of columns} - 1) = (5 - 1) \times (3 - 1) = 8$.

Scoring

Parts (a), (b), (c) and (d) are each scored as essentially correct (E), partially correct (P) or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the probability is computed correctly and appropriate work is shown *OR* the probability calculation is set up correctly but a minor computational error is made.

Partially correct (P) if the probabilities of the two events are added without subtracting the probability of their intersection, resulting in $\frac{1380}{2500} = 0.552$.

OR

Independence is assumed in computing the probability of the intersection.

Incorrect (I) if the response does not meet the criteria for an E or P, or includes the correct decimal answer with no accompanying work or justification.

Note: An answer of $\frac{1135}{2500}$ in fraction form is sufficient to be scored as essentially correct (E).

Part (b) is scored as follows:

Essentially correct (E) if the conditional probability is correctly computed and appropriate work is included *OR* if the calculation is set up correctly but a minor computational error is made.

Partially correct (P) if the reverse conditional probability (being a college graduate given that he or she primarily obtains news from the internet) is computed, resulting in $\frac{245}{687} = 0.357$.

Incorrect (I) if the probability of the *intersection* of the two events is computed, resulting in $\frac{245}{2500} = 0.098$.

    OR

The *unconditional* probability of obtaining news primarily from the internet is computed, resulting in $\frac{687}{2500} = 0.275$.

    OR

The response otherwise fails to meet the requirements for an E or P.

Notes

- An answer of $\frac{245}{693}$ in fraction form is sufficient to be scored as essentially correct (E).
- A correct decimal answer with no work or justification is scored as incorrect (I).

Part (c) is scored as follows:

Essentially correct (E) if the response states that the events are not independent and gives a correct numerical justification based on the table.

Partially correct (P) if the response states that the events are not independent and gives a correct statistical justification, but numerical support is not included (for example, says that P(C/I) ≠ P(C) but never reports either probability) *OR* the response includes correct and relevant calculations related to independence of these events but reaches an incorrect conclusion that the events *are* independent.

Incorrect (I) if the response states that the events are not independent but the given justification is not based on a correct probability argument *OR* the response does not reveal an understanding of how to assess whether two events are independent by comparing appropriate probabilities.

Part (d) is scored as follows:

Essentially correct (E) if the chi-square test of association (or independence) is correctly identified and the correct degrees of freedom are given.
Note: It is not necessary to show work in calculating the degrees of freedom.

Partially correct (P) if the response includes the correct name (chi-square test of association or independence) but not the correct degrees of freedom.

Incorrect (I) if the response includes neither identification of the chi-square test of association or independence nor correct degrees of freedom.

Notes
- If the response includes only "chi-square test" without specifying "of association (or independence)," this part is scored as essentially correct (E) provided that the degrees of freedom are computed correctly but as incorrect (I) if the degrees of freedom are incorrect.
- If the response identifies the test as "chi-square test of goodness-of-fit" or "chi-square test of homogeneity of proportions," the response is scored as incorrect (I).
- If the response does not name a correct test and only gives correct degrees of freedom, the response is scored as incorrect (I).

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part 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. Also use the following guidelines:
- If part (a) was scored as partially correct (P), always score down.
- A holistic score of 1 may be given to a response with all four parts scored as incorrect (I), if parts (a) and (b) both provide correct decimal answers but received no credit because supporting work was not included.

5. An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

| Primary Source for News | HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT | | | |
	Not High School Graduate	High School Graduate But Not College Graduate	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

(a) If an adult is to be selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

The probability that the selected adult is a college graduate or obtains news primarily from the Internet is

$$\frac{693 + 687 - 245}{2500} = \frac{1135}{2500} = 0.454.$$

(b) If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

~~Let A denote the~~

The probability that the selected adult obtains news primarily from the Internet in the condition that an adult who is a college graduate is to be selected is

$$\frac{\frac{245}{2500}}{\frac{693}{2500}} = \frac{245}{693} \approx 0.354$$

(c) When selecting an adult at random from the sample of 2,500 adults, are the events "is a college graduate" and "obtains news primarily from the internet" independent? Justify your answer.

The probability of the event "is a college student" is $\frac{693}{2500}$, while that for "obtains news primarily from the internet" is $\frac{687}{2500}$.

Since $\frac{245}{2500} \neq \frac{693}{2500} \cdot \frac{687}{2500}$. They are not independent.

(d) The company wants to conduct a statistical test to investigate whether there is an association between educational achievement and primary source for news for adults in the city. What is the name of the statistical test that should be used?

The Chi-square test for independence should be used.

What are the appropriate degrees of freedom for this test?

The appropriate degrees of freedom for this test is $(3-1) \times (5-1) = 8$.

(3 is the number of columns, while 5 is the number of rows.)

GO ON TO THE NEXT PAGE.

5. An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

| Primary Source for News | HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT | | | |
	Not High School Graduate	High School Graduate But Not College Graduate	College Graduate	Total
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

(a) If an adult is to be selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

$$\frac{693 + (687 - 245)}{2500} = .454$$

Probability of selecting adult who is college graduate or obtains news primarily from internet is 45.4%

(b) If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

$$\frac{245}{693} = .353535$$

Probability of college graduate, who gets primarily news source from internet is 35.354

(c) When selecting an adult at random from the sample of 2,500 adults, are the events "is a college graduate" and "obtains news primarily from the internet" independent? Justify your answer.

No they are not independent.
This is because (693)(687) ≠ (245)(2500)

(d) The company wants to conduct a statistical test to investigate whether there is an association between educational achievement and primary source for news for adults in the city. What is the name of the statistical test that should be used?

You would use a χ^2 test.

What are the appropriate degrees of freedom for this test?

Appropriate degrees of freedom would be 100 as it is the highest amount in table C.

GO ON TO THE NEXT PAGE.

5. An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

Primary Source for News	HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT			Total
	Not High School Graduate	High School Graduate But Not College Graduate	College Graduate	
Newspapers	49	205	188	442
Local television	90	170	75	335
Cable television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1,437	693	2,500

(a) If an adult is to be selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
$$= \left(\frac{693}{2500} + \frac{687}{2500} \right) - \frac{245}{2500}$$
$$= (.2772 + .2748) - .098$$
$$= .454$$

→ There is a 45.4% chance that the selected adult is a college graduate or obtains news primarily from the internet

(b) If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

$$\frac{\text{Obtains new from internet mostly}}{\text{college grad.}} = \frac{245}{693} = 35.35\%.$$

There is a 35.35% chance that an adult who is a college graduate is to be selected at random from this sample the selected adult obtains news primarily from the internet.

(c) When selecting an adult at random from the sample of 2,500 adults, are the events "is a college graduate" and "obtains news primarily from the internet" independent? Justify your answer.

No they are not independant because college graduates may have higher amount of access to internet than adults of other educational levels.

(d) The company wants to conduct a statistical test to investigate whether there is an association between educational achievement and primary source for news for adults in the city. What is the name of the statistical test that should be used?

What are the appropriate degrees of freedom for this test?

2499 degrees of freedom would be appropriate.

GO ON TO THE NEXT PAGE.

-13-

Question 5

Sample: 5A
Score: 4

The probability calculation in part (a) is correct, and supporting work is shown. In part (b) the correct conditional probability is calculated, and supporting work is given. Parts (a) and (b) were scored as essentially correct. In part (c) the student incorrectly labels the event "is a college graduate" as "is a college student." This was considered a minor error. The response shows that the probability of the intersection of the two events "is a college graduate" and "obtains news primarily from the internet" is not equal to the product of the two individual event probabilities and correctly concludes that the two events are not independent. Part (c) was scored as essentially correct. In part (d) the response correctly identifies the appropriate test as the "[c]hi-square test for independence" and includes a correct calculation of degrees of freedom. Part (d) was also scored as essentially correct. The entire answer, based on all four parts, was judged a complete response and earned a score of 4.

Sample: 5B
Score: 3

The probability calculations in parts (a) and (b) are correct, and supporting work is given. These two parts were scored as essentially correct. In part (c) the response correctly concludes that the two events "is a college graduate" and "obtains news primarily from the internet" are not independent and supports this conclusion with an argument that is equivalent to showing that the product of the individual event probabilities is not equal to the probability of the intersection of the two events. Part (c) was scored as essentially correct. In part (d) the chi-square test is identified, but the response does not specify that the test is for association or independence. This omission, accompanied by incorrect degrees of freedom, resulted in a score of incorrect for part (d). The entire answer, based on all four parts, was judged a substantial response and earned a score of 3.

Sample: 5C
Score: 2

The probability calculation in part (a) is correct, and supporting work is shown. In part (b) the correct conditional probability is calculated, and supporting work is given. Parts (a) and (b) were scored as essentially correct. Although the response indicates that the two events "is a college graduate" and "obtains news primarily from the internet" are not independent, no statistical justification is provided, so part (c) was scored as incorrect. In part (d) no test procedure is identified, and the degrees of freedom given are incorrect. Part (d) was scored as incorrect. The entire answer, based on all four parts, was judged a developing response and earned a score of 2.