AP Statistics
Chapter 14 Review

Name: Key
Hour: _____

1. Which of the following statements is not true?
(A) A chi-square test of independence that is statistically significant shows a cause and effect relationship.
(B) A segmented bar graph is useful in observing when two variables might be independent.
(C) As the number of categories increases, the χ^2 distribution approaches the normal distribution.
(D) The chi-square tests involve categorical variables.
(E) The chi-square goodness-of-fit test is an extension of the z-test for more than two categories.

2. The null hypothesis is rejected in a chi-square test of significance when
(A) the test conditions are satisfied.
(B) the P-value is larger than α, the level of significance.
(C) the P-value is larger than $1 - \alpha$.
(D) the χ^2 statistic is smaller than the critical value for the given α, the level of significance.
(E) the χ^2 statistic is larger than the critical value for the given α, the level of significance. Higher χ^2 stats give more evidence to reject H_0

3. An investigator was studying a territorial species of Central American termites, *Nasutitermes corniger*. Forty-nine termite pairs were randomly selected; both members of each of these pairs were from the same colony. Fifty-five additional termite pairs were randomly selected; the two members in each of these pairs were from different colonies. The pairs were placed in petri dishes and observed to see whether they exhibited aggressive behavior. The results are shown in the table below.

	Aggressive	Nonaggressive	Total
Same Colony	40 (33.5)	9 (15.5)	49
Different Colonies	31 (37.5)	24 (17.5)	55
Total	71	33	104

$p \approx .006$

A Chi-square test for homogeneity was conducted, resulting in $\chi^2 = 7.638$. The expected counts are shown in parentheses in the table. Which of the following sets of statements follows from these results?

(A) χ^2 is not significant at the 0.05 level.
(B) χ^2 is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.
(C) χ^2 is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.
(D) χ^2 is significant, $p < 0.01$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.
(E) χ^2 is significant, $p < 0.01$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.

4. Match the situations in parts a-c with the most appropriate chi-square test: goodness-of-fit, homogeneity, or independence.

a) You are told that the number of cracked M&M's depends on color. To check this claim, you randomly select 100 M&M's and sort by color and whether the M&M is cracked. Independence / Association

b) You are told that the distribution of M&M's colors is as follows: 15% red, 20% green, 30% yellow, 10% orange, and 25% brown. To check this claim, you randomly select a sample of M&M's and count the number of M&M's of each color. Goodness of fit

c) You are told that the number of cracked M&M's depends on color. To check this claim, you randomly select 100 M&M's of each color and count the number of cracked and uncracked M&M's of each color. Homogeneity

5.

In an experiment, two different species of flowers were crossbred. The resulting flowers from this crossbreeding experiment were classified, by color of flower and stigma, into one of four groups, as shown in the table below.

Flower Type Resulting from Crossbreeding	Number of Flowers Observed with These Colors
I: Magenta flower with green stigma	115
II: Magenta flower with red stigma	49
III: Red flower with green stigma	32
IV: Red flower with red stigma	21

A biologist expected that the ratio of 9:3:3:1 for the flower types I:II:III:IV, respectively, would result from this crossbreeding experiment. From the data above, a χ^2 value of approximately 8.04 was computed. Are the observed results inconsistent with the expected ratio at the 5 percent level of significance?

(A) Yes, because the computed χ^2 value is greater than the critical value. p-value = .045

(B) Yes, because the computed χ^2 value is less than the critical value.

(C) No, because the computed χ^2 value is less than the critical value.

(D) No, because the computed χ^2 value is greater than the critical value.

(E) It cannot be determined because some of the expected counts are not large enough to use the χ^2 test.

6. A high school principal decided to take a random sample of 200 students, 50 from each of grades 9 – 12, and record the frequency of absences for a randomly selected Friday during the school year. The number of Friday absences for the 50 students in each of grades 9 – 12 were 4, 4, 6, and 11, respectively. The principal wishes to test the null hypothesis that the absence rate is constant (the same) among the grade levels. Test the principal's claim.

① H_0: The absence rate among the grade levels is the same.
H_a: The absence rate among the grade levels is different.

② χ^2 test of Homogeneity

③ SRS is stated.

Expected counts:

	Absent	Present
9	6.25	43.75
10	6.25	43.75
11	6.25	43.75
12	6.25	43.75

All expected counts are ≥ 5

④ Observed Counts

9	4	46
10	4	46
11	6	44
12	11	39

$\chi^2 = \frac{(4-6.25)^2}{6.25} + \ldots \frac{(39-43.75)^2}{43.75} \approx 5.989$

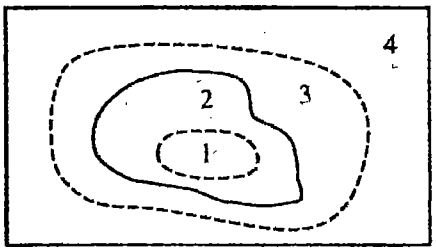
$df = (4-1)(2-1) = 3$ $p \approx .112$

⑤ Since $p = .112 > \alpha = .05$ we fail to reject H_0. There is evidence to show the absence rates among grade levels is the same.

7. A study was conducted to determine where moose are found in a region containing a large burned area. A map of the study area was partitioned into the following four habitat types.

(1) Inside the burned area, not near the edge of the burned area, (2) Inside the burned area, near the edge, (3) Outside the burned area, near the edge, and (4) Outside the burned area, not near the edge.

The figure at right shows these four habitat types.



Note: Figure not drawn to scale.

The proportion of total acreage in each of the habitat types was determined for the study area. Using an aerial survey, moose locations were observed and classified into one of the four habitat types. The results are given in the table.

Habitat Type	Proportion of Total Acreage	Number of Moose Observed
1	0.340	25 (39.78)
2	0.101	22 (11.817)
3	0.104	30 (12.168)
4	0.455	40 (53.235)
Total	1.000	117

(a) The researchers who are conducting the study expect the number of moose observed in a habitat type to be proportional to the amount of acreage of that type of habitat. Are the data consistent with this expectation? Conduct an appropriate statistical test to support your conclusion. Assume the conditions for inference are met.

① H_0: The number of moose in each habitat is proportional to the acreage of that habitat.

H_a: The number of moose are not proportional to acreage.

② χ^2 Goodness of fit test

③ We are told to assume conditions for inference are met

④ $\chi^2 = \frac{(25-39.78)^2}{39.78} + \ldots + \frac{(40-53.235)^2}{53.235} = 43.689$ $df = 4-1 = 3$ p-value ≈ 0

⑤ Since the p-value $= 0 < \alpha = .05$, we reject H_0. There is sufficient evidence that the number of moose in each habitat is not proportional to the acreage. The data is not consistent with the expectation.

(b) Relative to the proportion of total acreage, which habitat types did the moose seem to prefer? Explain.

The moose seem to prefer habitats 2 and 3 since the observed counts are at least twice the amount as is expected.

8. The Colorado Rocky Mountain Rescue Service wishes to study the behavior of lost hikers. If more were known about the direction in which lost hikers tend to walk, then more effective search strategies could be devised. 200 hikers selected at random from those applying for hiking permits are asked whether they would hike uphill, downhill, or remain in the same place if they became lost while hiking. Each hiker in the sample was also classified according to whether he or she was an experienced or novice hiker. The resulting data are summarized in the following table.

Expected = $\frac{\text{row} \times \text{column}}{\text{Total}}$

	Direction			
	Uphill	Downhill	Remain in same Place	
Novice	20 (18)	50 (48)	50 (54)	120
Experienced	10 (12)	30 (32)	40 (36)	80
	30	80	90	

Do these data provide convincing evidence of an association between the level of hiking expertise and the direction the hiker would head if lost? Give appropriate statistical evidence to support your conclusion.

① H_0: There is no association between level of hiking experience and direction.
H_a: There is an association

② χ^2 test for association

③ SRS is stated. Expected counts: $\begin{bmatrix} 18 & 48 & 54 \\ 12 & 32 & 36 \end{bmatrix}$ All expected counts are ≥ 5

④ $\chi^2 = \frac{(20-18)^2}{18} + \ldots + \frac{(40-36)^2}{36} = 1.505$ $df = (3-1)(2-1) = 2$

p-value $\approx .471$

⑤ Since $p = .471 > \alpha = .05$ we fail to reject H_0. There is not significant evidence to conclude there is an association between hiking experience and direction when lost.

9. 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.

	HIGHEST LEVEL OF EDUCATIONAL ACHIEVEMENT			
Primary Source for News	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) Conduct a statistical test to investigate whether there is an association between educational achievement and primary source for news for adults in the city.

① H_0: There is no association between the level of educational achievement and the primary source for news in the city.

H_a: There is an association.

② χ^2 test for association

③ SRS is stated. Expected counts $\begin{bmatrix} 65.416 & 254.06 & 122.52 \\ 49.58 & 192.56 & 92.862 \\ 111.89 & 434.55 & 209.56 \\ 101.68 & 394.89 & 190.44 \\ 41.44 & 160.94 & 77.616 \end{bmatrix}$ All expected counts ≥ 5

④ $\chi^2 = \frac{(49-65.416)^2}{65.416} + \ldots + \frac{(38-77.616)^2}{77.616} = 217.769$ $df = (5-1)(3-1) = 8$ p-value ≈ 0

⑤ Since the p-value $= 0 < \alpha = .05$, we reject H_0. There is sufficient evidence of an association between educational achievement and primary source for news in the city.

(b) Based on your conclusion in part (a), which type of error (Type I or Type II) might the you have made? Describe this error in the context of the question.

Rejecting H_0 may be a Type I error. This would mean we concluded there is an association between educational achievement and primary news source when there really isn't an association.