TEACHER

AP Stats

Review Chapter 14

Name _____

1. Births are not evenly distributed across the days of the week. Fewer babies are born on Saturday and Sunday than on other days, probably because doctors find weekend births inconvenient. A random sample of 700 births from local records shows this distribution across the days of the week:

	EV 100	100	100	100	100	100	100
Day	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.
Births	84	110	124	104	94	112	72
							=700

$P_{EXP} = \frac{1}{7}$

(a) The null hypothesis is that all days are equally probable. What are the probabilities specified by this null hypothesis? proportions →

H_0: Prop. of Births are same for each day = 1/7

$P_M = 1/7 = P_T = P_W = P_{TH} = P_F = P_S = P_S$

H_a: at least one prop is diff. from 1/7

(b) What are the expected counts for each day in 700 births?

700 ÷ 7 = 100 For each

(c) Calculate the chi-square statistic for goodness of fit.

$\chi^2 = \chi^2GOF(L1, L2, df=6)$

$\chi^2 = 19.12$

$p = .004$

(d) What are the degrees of freedom for this statistic? $df = 6$

(e) Do 700 births give significant evidence that births are not equally probable on all days of the week?

The p-value associated with $\chi^2 = 19.12$ is .00397 which is significantly low. This is enough evidence to reject H_0 that the proportion of births are equally distributed and support that they are not equally distributed.

2. A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Here are the results:

Indep one sample

	Community Type		
	Rural	Suburban	Urban
Internet users	433	1072	536
Nonusers	463	627	388

3519

Is there a relationship between Internet use and community type? Give statistical evidence to support your findings.

H_0: There is no association Between Internet use and community type
(or
Internet use and community type are independ.)

H_a: There is an association Between Internet use and community type

Conditions

Random → given

10% → 3519 < 10%-population of internet users

Large Counts → Expected counts are all greater than 5 $\begin{bmatrix} 519.67 & 985.41 & 535.91 \\ 376.33 & 713.59 & 388.09 \end{bmatrix}$

Conditions are satisfied to do a χ^2 Test for independence

$\chi^2 = 52.535$

$p = 3.9 \times 10^{-12}$ with $df = 2$

[A] observed
[B] expected

the p-value is very small = approx zero therefore we can reject the Ho of no association. There is sufficient evidence to support that there is a relationship between internet use and community type.

Review 14 Multiple Choice

1. A chi-square goodness of fit test is used to test whether a 0 to9 spinner is "fair" (that is, the outcomes are all equally likely). The spinner is spun 100 times, and the results are recorded. Which member of the chi-square family of curves is used?

(a) $\chi^2(8)$ (b) $\chi^2(9)$ (c) $\chi^2(10)$ (d) $\chi^2(99)$ (e) None of the above

df

2. A study of accident records at a large engineering company in England reported the following number of injuries on each shift for 1 year: 4636

H_0: p's same df=2

H_a: p's not same

Shift:	Morning	Afternoon	Night
Number of injuries:	1372	1578	1686
Ha	1545.33	1545.33	1545.33

χ^2GOF L1= obs L2= Exp > Tests GOF

Is there sufficient evidence to say that the numbers of accidents on the three shifts are not the same? Test at the 0.05, 0.01, and 0.001 levels. $\chi^2=32.94$ $p=\approx 0$

(a) There is sufficient evidence at all three levels to say that the numbers of accidents on each shift are not the same.

(b) There is sufficient evidence at the 0.05 and 0.01 levels but not at the 0.001 level.

(c) There is sufficient evidence at the 0.05 level but not at the 0.01 or 0.001 levels.

(d) There is sufficient evidence at the 0.001 level but not at the 0.01 or 0.05 levels.

(e) There is insufficient evidence at any of these levels.

Questions 3 to 10 refer to the following situation.

In the paper "Color Association of Male and Female Fourth-Grade School Children" (*Journal of Psychology*, 1988, 383–388), reported on a study in which children were asked to indicate what emotion they associated with the color red. The response and the sex of the child are noted and summarized below. The first number in each cell is the count; the second number is the row percent.

	Anger	**Happy**	**Love**	**Pain**	**Total**
Female	27 26.47	19 18.63	39 38.24	17 16.67	**102**
Male	34 30.36	12 10.71	38 33.93	28 25.00	**112**
Total	**61**	**31**	77	**45**	**214**

~~102~~ $\frac{(112)(61)}{214}=31.93$

df=3

Statistic	DF	Value	Prob
Pearson Chi-Square	*	4.629	*****
Likelihood Ratio Chi-Square	*	4.661	*****
Mantel-Haenszel Chi-Square	1	0.307	*****

3. The null hypothesis is

Ho: Gender is indep. of red color feeling
Ha: Gender is not ind. of red color feeling

(a) emotional association with red is independent of gender.

(b) gender is dependent upon the emotional association with red.

(c) the probability of associating a specific emotion with red is related to gender.

(d) the number of children in each cell does not depend upon gender or upon emotion.

(e) the color red is independent of the emotion associated with it and with gender.

Wording indicates red & But its gender vs red feeling

4. Under a suitable null hypothesis, the expected frequency for the cell corresponding to Anger and Males is

%

(a) 15.9.

(b) 55.7.

(c) 30.4.

(d) 31.9. ← $\frac{(\text{row})(\text{colum})}{\text{tot}} = \frac{(112)(61)}{214}$

(e) 29.1.

5. The null hypothesis will be rejected at $\alpha = 0.05$ if the test statistic exceeds

Think backwards. what is χ^2 at $\alpha = .05$

(a) 3.84.

(b) 5.99.

(c) 7.81.

(d) 9.49.

(e) 14.07.

$df = (2-1)(4-1) = 3$

$df = 3$

6. The approximate *P*-value is

$\chi^2 cdf(4.629, \infty, 3) \approx 2.011$

(a) between 0.100 and 0.900.

(b) between 0.050 and 0.100.

(c) between 0.025 and 0.050.

(d) between 0.010 and 0.025.

(e) between 0.005 and 0.010.

7. Which of the following is NOT CORRECT?

(a) The children were classified by sex and emotion associated with red. Each child was counted in one and only one cell. correct

(b) The null hypothesis is that the type of emotion associated with red is independent of the sex of the child. correct

(c) The null hypothesis is that the proportion of emotions associated with red is the same for both sexes. ↳ can use prop./counts → gender=

(d) All expected cell counts should be greater than 5 in order that the distribution of the test statistic is an approximate chi-square distribution. T

(e) If we reject the null hypothesis, then we have proven that the two sexes associate red with emotions in different ways. red emot not indep of gender

not proof ⇒ support they are not independ. ∴ that's acp

8. Which of the following is NOT CORRECT?

(a) A lower percent of female students associate the emotion "anger" with the color red than do male students. T

(b) More students associate the color red with the emotion "love" than with the emotion "anger." T

(c) There is insufficient evidence of an association between gender and emotion associated with the color red. T

(d) We will be unable to compute a correlation for these data because the variables are both categorical.

(e) We compute row or column percents by dividing the cell count by the table total (214).

$\frac{(row \times col)}{tot}$ not $\frac{cell}{tot}$

9. A Type I error would be committed if

(a) we conclude that the sex of the child and the emotion associated with red are independent when in fact they are not independent.

(b) we conclude that the sex of the child and the emotion associated with red are not independent when in fact they are not independent.

(c) we conclude that the proportion of emotions associated with red differs between males and females when in fact they are the same.

(d) we conclude that the proportion of emotions associated with red is the same for males and females when in fact they are the same.

(e) we fail to find any association between the color red and emotions for either sex.

10. The test statistic and approximate *P*-value are

(a) 4.661 0.1983

(b) 4.661 0.3966

(c) 4.629 0.2011

(d) 4.629 0.4022

(e) 4.629 0.1006