Chapter 6 - Categorical Data and Chi-Square

6.1 Popularity of psychology professors:

	Anderson	Klatsky	Kamm	Total
Observed	32	25	10	67
Expected	22.3	22.3	22.3	67

$$\chi^2 = \Sigma\frac{(O-E)^2}{E}$$

$$= \frac{(32-22.3)^2}{22.3} + \frac{(25-22.3)^2}{22.3} + \frac{(10-22.3)^2}{22.3}$$

= 11.33[1]

Reject H_0 and conclude that students do not enroll at random.

6.3 Racial choice in dolls (Clark & Clark, 1939):

	Black	White	Total
Observed	83	169	252
Expected	126	126	252

$$\chi^2 = \frac{(O-E)^2}{E}$$

$$= \frac{(83-126)^2}{126}\frac{(169-126)^2}{126}$$

$$= 29.35 \quad \left[\chi^2_{.05(1)} = 3.84\right]$$

Reject H_0 and conclude that the children did not chose dolls at random (at least with respect to color). It is interesting to note that this particular study played an important role in Brown v. Board of Education (1954). In that case the U.S. Supreme Court ruled that the principle of "separate but equal", which had been the rule supporting segregation in the public schools, was no longer acceptable. Studies such as those of the Clarks had illustrated the negative effects of segregation on self-esteem and other variables.

[1] The answers to these questions may differ substantially, depending on the number of decimal places that are carried for the calculations.

6.7 Combining the two racial choice experiments:

Study	Black	White	Total
1939	83 (106.42)	169 (145.58)	252
1970	61 (37.58)	28 (51.42)	89
	144	197	341 = N

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(83-106.42)^2}{106.42} + \frac{(169-145.58)^2}{145.58} + \frac{(61-37.58)^2}{37.58} + \frac{(28-51.42)^2}{51.42}$$

$$= 5.154 + 3.768 + 14.595 + 10.667$$

$$= 34.184 \quad \left[\chi^2_{.05(1)} = 3.84\right]$$

Reject the H_0 and conclude that the distribution of choices between Black and White dolls was different in the two studies. Choice is *not* independent of Study. We are no longer asking whether one color of doll is preferred over the other color, but whether the *pattern* of preference is constant across studies. In analysis of variance terms we are dealing with an interaction.

6.7 **a.** Take a group of subjects at random and sort them by gender and life style (categorized three ways).

b. Deliberately take an equal number of males and females and ask them to specify a preference among 3 types of life style.

c. Deliberately take 10 males and 10 females and have them divide themselves into two teams of 10 players each.

6.9 Doubling the cell sizes:

a. $\chi^2 = 10.306$

b. This demonstrates that the obtained value of χ^2 is exactly doubled, while the critical value remains the same. Thus the sample size plays a very important role, with larger samples being more likely to produce significant results—as is also true of other tests.

6.11 Gender and voting behavior

	Vote		
	Yes	**No**	**Total**
Women	35	9	44
	(28.83)	(15.17)	
Men	60	41	101
	(66.17)	(34.83)	
Total	95	50	145

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(35-28.83)^2}{28.83} + \frac{(9-15.17)^2}{15.17} + \frac{(60-66.17)^2}{66.17} + \frac{(41-34.83)^2}{34.83}$$

$$= 5.50 \qquad [\chi^2_{.05(1)} = 3.84]$$

Reject H_0 and conclude that women voted differently from men. The odds of women supporting civil unions much greater than the odds of men supporting civil—the odds ratio is (35/9)/(60/41) = 3.89/1.46 = 2.66. The odds that women support civil unions were 2.66 times the odds that men did. That is a substantial difference, and likely reflects fundamental differences in attitude.

6.13 **a.** Weight preference in adolescent girls:

	Reducers	Maintainers	Gainers	Total
White	352	152	31	535
	(336.7)	(151.9)	(46.4)	
Black	47	28	24	99
	(62.3)	(28.1)	(8.6)	
	399	180	55	634 = N

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

$$= \frac{(352-336.7)^2}{336.7} + \frac{(152-151.9)^2}{151.9} + \ldots + \frac{(24-8.6)^2}{8.6}$$

$$= 37.141 \qquad [\chi^2_{.05(2)} = 5.99]$$

Adolescents girls' preferred weight varies with race.

b. The number of girls desiring to lose weight was far in excess of the number of girls who were overweight.

6.15 Analyzing Exercise 6.10 (Regular or Remedial English and frequency of ADD diagnosis) using the likelihood-ratio approach:

	1st	2nd	4th	2 & 4	5th	2 & 5	4 & 5	2,4,&5	Total
Rem.	22	2	1	3	2	4	3	4	41
Reg.	187	17	11	9	16	7	8	6	261
	209	19	12	12	18	11	11	10	302

$$\chi^2 = 2\left(\Sigma O_{ij} \ln\left[\frac{O_{ij}}{E_{ij}}\right]\right)$$
$$= 2\times[22\times\ln(22/28.374) + 2\times\ln(2/2.579) + ... + 6\times\ln(6/8.642)]$$
$$= 2\times[22(-.25443) + 2(-0.25444) + ... + 6(-0.36492)]$$
$$= 12.753 \text{ on } 7\ df$$

Do not reject H_0 .

6.17 Monday Night Football opinions, before and after watching:

	Pro to Con	Con to Pro	Total
Observed Frequencies	20	5	25
Expected Frequencies	12.5	12.5	25

$$\chi^2 = \Sigma\frac{(O-E)^2}{E} = \frac{(20-12.5)^2}{12.5} + \frac{(5-12.5)^2}{12.5}$$
$$= 4.5 + 4.5 = 9.0 \text{ on } 1\ df. \text{ Reject } H_0$$

b. If watching Monday Night Football really changes people's opinions (in a negative direction), then of those people who change, more should change from positive to negative than vice versa, which is what happened.

c. The analysis does not take into account all of those people who did not change. It only reflects direction of change if a person changes.

6.19 **b.** Row percents take entries as a percentage of row totals, while column percents take entries as percentage of column totals.

c. These are the probabilities (to 4 decimal places) of a $\chi^2 \geq \chi^2_{\text{obt}}$

d. The correlation between the two variables is approximately .25.

6.21 For data in Exercise 6.20a:

a. $\phi_c = \sqrt{26.90/22,071} = 0.0349$

b. Odds Fatal | Placebo = 18/10,845 = .00166.
Odds Fatal | Aspirin = 5/10,933 = .000453.
Odds Ratio = .00166/.000453 = 3.66
The odds that you will die from a myocardial infarction are 3.66 times higher if you do not take aspirin than if you do.

6.23 For Table 6.4 the odds ratio for a death sentence as a function of race is (33/251)/(33/508) = 2.017. A person is about twice as likely to be sentenced to death if they are nonwhite than if they are white.

6.25 Dabbs and Morris (1990) study of testosterone.

		Testosterone High	Testosterone Normal	Total
Delinquency	No	345 (395.723)	3614 (3563.277)	3959
	Yes	101 (50.277)	402 (452.723)	503
		446	4016	4462 = N

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(345-395.723)^2}{395.723} + \frac{(3614-3563.277)^2}{3563.277} + \frac{(101-50.277)^2}{50.277} + \frac{(402-452.723)^2}{452.723}$$

$$= 64.08 \quad \left[\chi^2_{.05(1)} = 3.84\right] \quad \text{Reject } H_0$$

6.27 Childhood delinquency in the Dabbs and Morris (1990) study.

a.

		Testosterone High	Testosterone Normal	Total
Delinquency	No	366 (391.824)	3554 (3528.176)	3920
	Yes	80 (54.176)	462 (487.824)	542
		446	4016	4462 = N

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(366-391.824)^2}{391.824} + \frac{(3554-3528.176)^2}{3528.176} + \frac{(80-54.176)^2}{54.176} + \frac{(462-487.824)^2}{487.824}$$

$$= 15.57 \quad \left[\chi^2_{.05(1)} = 3.84\right] \quad \text{Reject } H_0$$

b. There is a significant relationship between high levels of testosterone in adult men and a history of delinquent behavior during childhood.

c. This result shows that we can tie the two variables (delinquency and testosterone) together historically.

6.29 Good touch/Bad touch

a.

		Abused		
		Yes	No	Total
Received Program	Yes	43 (56.85)	457 (443.15)	500
	No	50 (36.15)	268 (281.85)	318
		93	725	818 = N

$$\chi^2 = \Sigma\frac{(O-E)^2}{E}$$

$$= \frac{(43-56.85)^2}{56.85} + \frac{(457-443.15)^2}{443.15} + \ldots + \frac{(268-281.85)^2}{281.85}$$

$$= 9.79 \qquad \chi^2_{.05(1)} = 3.84 \quad \text{Reject } H_0$$

b. Odds ratio

OR = (43/457)/(50/268) = 0.094/0.186 = .505. Those who receive the program have about half the odds of subsequently suffering abuse.

6.31 Gender of parents and children.

a.

		Lost Parent Gender		
		Male	Female	Total
Child	Male	18	34	52
	Female	27	61	88
		45	95	140 = N

$\chi^2 = .232$
($p = .630$)

b. There is no relationship between the gender of the lost parent and the gender of the child.

c. We would be unable to separate effects due to parent's gender from effects due to the child's gender. They would be completely confounded.

6.33 We could ask a series of similar questions, evenly split between “right” and “wrong” answers. We could then sort the replies into positive and negative categories and ask whether faculty were more likely than students to give negative responses.

6.35 I alluded to this when I referred to the meaning of kappa in the previous question. Kappa would be noticeably reduced if the scales used by husbands and wives were different, but the relationship could still be high.

6.37 Fidalgo’s study of bullying in the work force.

a. Collapsing over job categories

	Not Bullied	**Bullied**	**Total**
Male	461 (449.54)	68 (79.46)	529
Female	337 (342.46)	72 (60.54)	403
Total	792	140	932

$$\chi^2 = \Sigma\left(\frac{(O-E)^2}{E}\right)$$

$$= \frac{(461-449.54)^2}{449.54} + \frac{(68-79.46)^2}{79.46} + \frac{(337-342.46)^2}{342.46} + \frac{(72-60.54)^2}{60.54}$$

$$= 0.292 + 1.653 + 0.087 + 2.169 = 4.20$$

This chi-square is significant on 1 *df*

b. The odds ratio is

$$OR = \frac{68/461}{72/337} = \frac{.1478}{.2136} = .70$$

The odds that a male will be bullied are about 70% those of a female being bullied.

c. & d. Breaking the data down by job category

Using SPSS

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	2.602	1	.107
Mantel-Haenszel	2.285	1	.131

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			1.361
ln(Estimate)			.308
Std. Error of ln(Estimate)			.193
Asymp. Sig. (2-sided)			.111
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.931
		Upper Bound	1.988
	ln(Common Odds Ratio)	Lower Bound	-.071
		Upper Bound	.687

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

When we condition on job category there is no relationship between bullying and gender and the odds ratio drops to 1.36

e. For Males

Chi-Square Tests[b]

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.609[a]	4	.158
Likelihood Ratio	7.273	4	.122
Linear-by-Linear Association	5.591	1	.018
N of Valid Cases	529		

a. 1 cells (10.0%) have expected count less than 5. The minimum expected count is 3.98.

b. Gender = Male

For Females

Chi-Square Tests[b]

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.510[a]	4	.973
Likelihood Ratio	.550	4	.968
Linear-by-Linear Association	.246	1	.620
N of Valid Cases	403		

a. 1 cells (10.0%) have expected count less than 5. The minimum expected count is 1.61.

b. Gender = Female

For males bullying declines as job categories increase, but this is not the case for women.

6.39 Appleton, French, & Vanderpump (1996) study:

There is a tendency for more younger people to smoke than older people. Because younger people generally have a longer life expectancy than older people, that would make the smokers appear as if they had a lower risk of death. What looks like a smoking effect is an age effect.

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Dead (1.00 / 2.00)	1.460	1.141	1.868
For cohort Smoker = No	1.173	1.062	1.296
For cohort Smoker = Yes	.804	.693	.932
N of Valid Cases	1314		

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	9.121	1	.003
Mantel-Haenszel	8.745	1	.003

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.