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 Sorting one-sentence characteristics into piles:

	1	2	3	4	5	Total
Observed	8	10	20	8	4	50
Expected	5	10	20	10	5	50
Exp. %	10%	20%	40%	20%	10%	100%

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

$$= \frac{(8-5)^2}{5} + \frac{(10-10)^2}{10} + \frac{(20-20)^2}{20} + \frac{(8-10)^2}{10} + \frac{(4-5)^2}{5}$$

$$= 2.4 \qquad [\chi^2_{.05(4)} = 9.49]$$

Do not reject H_0 that your friend's child's sorting behavior is in line with your theory.

[1] The answers to these questions may differ substantially, depending on the number of decimal places that are carried for the calculations. (e. g. for Exercise 6.18 answers can vary between 37.14 and 37.339.)

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

	Black	White	Total
Observed	83	169	252
Expected	126	126	252

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

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

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

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.

6.7 Combining the two racial choice experiments:

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

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

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

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

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

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.

6.9 **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.

6.11 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.

6.13 Gender and voting behavior

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

6.15 The relationship of assistance-seeking behavior to number of bystanders:

		Sought Assistance		
		Yes	**No**	**Total**
Number of Bystanders	0	11 (7.75)	2 (5.25)	13
	1	16 (15.5)	10 (10.5)	26
	4	4 (7.75)	9 (5.25)	13
		31	21	52 = *N*

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

$$= \frac{(11-7.75)^2}{7.75} + \frac{(2-5.25)^2}{5.25} + \dots + \frac{(9-5.25)^2}{5.25}$$

$$= 7.908 \qquad [\chi^2_{.05(2)} = 5.99]$$

Reject H_0. The number of bystanders influences whether or not subjects seek help.

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

	Reducers	Maintainers	Gainers	Total
White	352 (336.7)	152 (151.9)	31 (46.4)	535
Black	47 (62.3)	28 (28.1)	24 (8.6)	99
	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.

6.19 Analyzing Exercise 6.12 (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 .

6.21 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

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(20-12.5)^2}{12.5} + \frac{(5-12.5)^2}{12.5}$$

$= 4.5 + 4.5 = 9.0$ on 1 df. 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.

6.23 **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.

6.25 For data in Exercise 6.24a:

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.

6.27 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.

6.29 Dabbs and Morris (1990) study of testosterone.

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

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

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

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

6.31 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*

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

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

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

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.

6.33 Good touch/Bad touch

a.

		Abused Yes	Abused 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.

6.35 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.

6.37 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.

6.39 I am trying to get students to think about the issues of measurement and about what we can, and cannot, tell from data. In Exercise 6.39, if scale points mean different things to different sexes, it is possible that the relationship could be distorted by the closed-end nature of the scales.

6.41 Mantel-Haenszel statistic on race and the death penalty by seriousness of the crime

	Death Penalty	
Seriousness	$\mathbf{O_{11k}}$	$\mathbf{E_{11k}}$
1	2	0.7623
2	2	1.3077
3	6	4.3333
4	9	7.3333
5	9	7.3125
6	17	17

$$M^2 = \frac{(\left|\Sigma O_{11k} - \Sigma E_{11k}\right| - \frac{1}{2})^2}{\sum (n_{1+k} n_{2+k} n_{+1k} n_{+2k} / n_{++k}^2 (n_{++k} - 1))}$$

$$= \frac{(\left|45 - 38.049\right| - \frac{1}{2})^2}{62*182*3*241/(244^2*243+...+(17*4*21*0)/(21^2*20)}$$

$$= \frac{(6.951 - .5)^2}{\left(0.564 + 0.699 + 1.382 + 1.007 + 0.640 + 0\right)} = \frac{6.451^2}{4.291} = 9.698$$

This is a chi-square on 1 *df* and is significant. Death sentence and race are related even after we condition on the seriousness of the crime.

$$OR = \frac{\Sigma\left(f_{11k} f_{22k} / n_{.k}\right)}{\Sigma\left(f_{21k} f_{12k} / n_{.k}\right)}$$

$$= \frac{\left(2*181/244 + 2*21/39 + ... + 17*0/21\right)}{(60*1/244 + 15*1/39 + ... + 0)4/21)}$$

$$= \frac{8.498}{1.5471} = 5.493$$

Controlling for the seriousness of a crime, a nonwhite defendant is 5.5 times as likely to receive the death penalty.

6.43 Seatbelt data

Whereas only 9% of the occupants of cars were not belted at the time of the accident, 22% of those who were injured were unbelted and 74% of those who were killed were unbelted.

The chi-square statistics for these two statements are 1738.00 and 363.2, both of which are clearly significant. A disproportionate number of those killed or injured were not wearing seat belts relative to the seatbelt use of occupants in general.