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.

6.2 We cannot tell in Exercise 6.1 if students chose different sessions because of the instructor or because of the times at which the sections are taught—Instructor and Time are confounded. We would *at least* have to offer the sections at the same time.

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

[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.)

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.

6.4 Racial choice in dolls revisited (Hraba & Grant, 1970):

	Black	White	Total
Observed	61	28	89
Expected	44.5	44.5	89

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$
$$= \frac{(61-44.5)^2}{44.5} + \frac{(28-44.5)^2}{44.5}$$
$$= 12.36 \qquad [\chi^2_{.05(1)} = 3.84]$$

Again we reject H_0, but this time the departure is in the opposite direction.

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*

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(83-106.42)^2}{106.42} + \frac{(169-145.58)^2}{145.58} + \frac{(61-37.58)^2}{37.58} + \frac{(28-51.42)^2}{51.42}$$
$$= 5.154 + 3.768 + 14.595 + 10.667$$
$$= 34.184 \quad \left[\chi^2_{.05(1)} = 3.84\right]$$

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.

6.6 Smoking and pregnancy:

	1 cycle	2 Cycles	3+ Cycles	Total
Smokers	29 (38.74)	16 (22.70)	55 (40.27)	100
Non-smokers	198 (188.26)	107 (110.30)	181 (195.73)	486
Total	227	133	236	586

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

$$= \frac{(29-38.74)^2}{38.74} + \frac{(16-22.70)^2}{22.70} + \ldots + \frac{(181-195.73)^2}{195.73}$$

$$= 11.54 \qquad [\chi^2_{.05(2)} = 5.99]$$

Reject H_0 and conclude that smoking is related to ease of getting pregnant.

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.

6.8 Prediction of High School English level from ADD classification in elementary school:

	Remed. Eng.	Reg. Eng.	Total
Normal	22 (28.374)	187 (180.626)	209
ADD	19 (12.626)	74 (80.374)	93
	41	261	302 = N

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

$$= \frac{(22-28.374)^2}{28.374} + \frac{(187-180.626)^2}{180.626} + \frac{(19-12.626)^2}{12.626} + \frac{(74-80.374)^2}{80.374}$$

$$= 5.38 \qquad [\chi^2_{.05(1)} = 3.84]$$

Reject H_0 and conclude that achievement level during high school varies as a function of performance during elementary school.

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.

6.10 Frequency of ADD diagnosis and High School English level:

a. Chi-square analysis:

	Never	**2nd**	**4th**	**2 & 4**	**5th**	**2 & 5**	**4 & 5**	**2,4,&5**	**Total**
Rem.	22	2	1	3	2	4	3	4	41
	(28.374)	(2.579)	(1.629)	(1.629)	(2.444)	(1.493)	(1.493)	(1.358)	
Reg.	187	17	11	9	16	7	8	6	261
	(180.626)	(16.421)	(10.371)	(10.371)	(15.556)	(9.507)	(9.507)	(8.642)	
	209	19	12	12	18	11	11	10	302 = *N*

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

$$= \frac{(22-28.374)^2}{28.374} + \frac{(2-2.579)^2}{2.579} + \ldots + \frac{(6-8.642)^2}{8.642}$$

$$= 19.094 \qquad [\chi^2_{.05(7)} = 14.07]$$

b. Reject H_0 .

c. Since nearly half of the cell frequencies are less than 5, I would feel very uncomfortable. One approach would be to combine adjacent columns.

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.

6.12 Inescapable shock and implanted tumor rejection:

	Inescapable Shock	Escapable Shock	No Shock	Total
Rejection	8 (14.52)	19 (14.52)	18 (15.97)	45
No Rejection	22 (15.48)	11 (15.48)	15 (17.03)	48
	30	30	33	93 = *N*

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

$$= \frac{(8-14.52)^2}{14.52} + \frac{(19-14.52)^2}{14.52} + ... + \frac{(15-17.03)^2}{17.03}$$

$$= 8.85 \; [\chi^2_{.05(2)} = 5.99]$$

Reject H_0. The ability to reject a tumor is affected by the shock condition.

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.

6.14 Analyzing Exercise 6.8 (Regular or Remedial English and ADD) using the likelihood-ratio approach:

	Remed. Eng.	Reg. Eng.	Total
Normal	22	187	209
ADD	19	74	93
	41	261	302 = *N*

$$\chi^2 = 2\left(\Sigma O_{ij}\ln\left[\frac{O_{ij}}{E_{ij}}\right]\right)$$

$$= 2\times[22\times\ln(22/28.374)+187\times\ln(187/180.626)+19\times\ln(19/12.626)+74\times\ln(74/80.374)]$$

$$= 2\times[22(-.25443)+187(.03468)+19(.40868)+74(-.08262)]$$

$$= 2\times[2.53874] = 5.077$$

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

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

Do not reject H_0 .

6.16 If we were to calculate a one-way chi-square test on row 2 alone, we would be asking if the students are evenly distributed among the eight categories. What we really tested in Exercise 6.12 is whether that distribution, *however it appears*, is the same for those who later took remedial English as it is for those who later took non-remedial English.

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.

6.18 Pugh’s study of decisions in rape cases.

Fault	**Guilty**	**Not Guilty**	**Total**
Little	153	24	177
	(127.56)	(49.44)	
Much	105	76	181
	(130.44)	(50.56)	
Total	258	100	358

$$\chi^2 = \Sigma\frac{(O-E)^2}{E}$$
$$= \frac{(153-127.56)^2}{127.56} + \frac{(24-49.44)^2}{49.44} + \frac{(105-130.44)^2}{130.44} + \frac{(76-50.56)^2}{50.56}$$
$$= 35.93 \qquad \chi^2_{.05} = 3.84$$

Judgments of guilt and innocence are related to the amount of fault attributed to the victim.

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.

6.20 Death rates from myocardial infarction:

	Fatal Attack	Non-Fatal Attack	No Attack	
Placebo	18 (11.498)	171 (134.982)	10,845 (10,887.52)	11,034
Aspirin	5 (11.502)	99 (135.018)	10,933 (10,890.48)	11,037
	23	270	21,778	22,071 = N

a.

$$\chi^2 = \Sigma\frac{(O-E)^2}{E}$$
$$= \frac{(18-11.498)^2}{11.498} + \frac{(171-134.982)^2}{134.982} + \dots + \frac{(10,933-10,890.48)^2}{10,890.48}$$
$$= 26.90$$

$$\chi^2 = 2\left(\Sigma O_{ij}\ln\left[\frac{O_{ij}}{E_{ij}}\right]\right)$$
$$= 2\times[18\times\ln(18/11.498) + 171\times\ln(171/134.982) + \dots + 10,933\times\ln(10,933/10,890.48)]$$
$$= 2\times[8.0675 + 40.4453 - 42.4369 - 4.1654 - 30.7185 + 42.6029]$$
$$= 27.59 \text{ on } 2\ df.\ \text{Reject } H_0$$

b. Using only the data from those with heart attacks

	Fatal Attack	Non-Fatal Attack	
Placebo	18 (14.836)	171 (174.163)	189
Aspirin	5 (8.164)	99 (95.836)	104
	23	270	293 = N

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

$$= \frac{(18-14.836)^2}{14.836} + \frac{(171-174.163)^2}{174.163} + \ldots + \frac{(99-95.836)^2}{95.836}$$

$$= 2.06$$

$$\chi^2 = 2\left(\Sigma O_{ij} \ln\left[\frac{O_{ij}}{E_{ij}}\right]\right)$$

$$= 2 \times [18 \times \ln(18/14.836) + 171 \times \ln(171/174.163) + \ldots + 99 \times \ln(99/95.836)]$$

$$= 2 \times [3.4797 - 3.1341 - 2.4515 + 3.2157]$$

$= 2.22$ on 1 df. Do not reject H_0

c. Combining the myocardial infarction groups:

	Attack	No Attack	
Placebo	189 (146.480)	10,845 (10,887.52)	11,034
Aspirin	104 (146.520)	10,933 (10,890.48)	11,037
	293	21,778	22,071 = N

$$\chi^2 = 2\left(\Sigma O_{ij} \ln\left[\frac{O_{ij}}{E_{ij}}\right]\right)$$

$$= 2 \times [189 \times \ln(189/146.48) + 10,845 \times \ln(10,845/10,887.52) + \ldots$$

$$+10,933 \times \ln(10,933/10,890.48)]$$

$$= 2 \times [48.1682 - 42.4368 - 35.6482 + 42.6029]$$

$= 25.3720$ on 1 df. Reject H_0

d. Combining **b.** and **c.**:

For Pearson chi-square, the sum = 2.06 + 25.01 = 27.07. The χ^2 for the full table was 26.90.

For likelihood-ratio chi-square, the sum = 2.22 + 25.37 = 27.59 = likelihood-ratio chi-square for the full table.

We can see that likelihood-ratios neatly partition a larger table.

WHEW! That's a lot of calculating and typing.

e. Aspirin significantly reduces the likelihood of a heart attack. The risk ratio of heart attack versus no heart attack is 1.81, meaning that the placebo group is 1.8 times more likely than the aspirin group to have a heart attack.

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.

6.22 Odds ratio for Exercise 6.10:

Odds of being in remedial English class if ADDSC score was normal = 22/187 = .1176.
Odds of being in remedial English class if ADDSC score was high = 19/74 = .2568.
Odds Ratio = .2568/.1176 = 2.18. The odds of taking remedial English are twice as high if you had a high ADDSC score than if you had a low one.

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.

6.24 Tests on data in Exercise 6.11.

Fisher's Exact test has a *p* value of .0226, while the chi-square test has a *p* value of .01899. We would come to the same conclusion with either test. (If we use the correction for continuity on chi-square (a poor idea) the probability would be .0311.)

6.25 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

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

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

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

6.26 Odds ratio for Dabbs and Morris (1990) data.

Odds of adult delinquency for high testosterone group = 101/345 = .2928
Odds of adult delinquency for normal testosterone group = 402/3614 = .1112
Odds ratio = .2928/.1112 = 2.63. The odds of engaging in behaviors of adult delinquency are 2.63 times higher if you are a member of the high testosterone group.

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

a.

		Testosterone		
		High	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.

6.28 Percentage agreement and Cohen's Kappa:

a.

		Rater A		
		Presence	Absence	Total
Extreme Verbal Abuse	No	12 (4.55)	2	14
	Yes	1	25 (17.55)	26
		13	27	40 = *N*

Percentage agreement = (12 + 25)/40 = .925 = 92.5% agreement

b. Cohen's Kappa

$$\kappa = \frac{\Sigma f_o - \Sigma f_e}{N - \Sigma f_e} = \frac{37 - 22.10}{40 - 22.10} = .83$$

c. Kappa is less than the percentage of agreement because the bias in favor of the behavior being absent means that if the judges each chose the rating of Absent a high percentage of the time, they would automatically agree often.

d. Bias the data even more toward ratings of Absent.

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*

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

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

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

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.

6.30 Gender vs. College in Mireault's (1990) data.

b.

	College					
	1	2	3	4	5	Total
Male	68	0	18	35	4	125
Female	95	21	6	37	16	175
	163	21	24	72	20	300 = *N*

$\chi^2 = 31.263$
(p = .000)

c. The distribution of students across the different colleges in the University varies as a function of gender.

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.

6.32 **a.** I would agree with the researcher. The probability of a Type I error is held at α, regardless of the sample size.

b. The reviewer is forgetting that the greater variability in the means of small samples is compensated for in the sampling distribution of the test statistic.

c. I would calculate the *number* of people in each category who sided with, *and against*, the researcher.

d. The level of accuracy varies by group. $\chi^2_{.05}(1) = 11.95$. Actually the students numerically outperform the other groups.

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.

6.34 Hout, Duncan, & Sobel (1987) study

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.955[a]	9	.049
Likelihood Ratio	15.486	9	.078
Linear-by-Linear Association	10.014	1	.002
N of Valid Cases	91		

[a] 7 cells (43.8%) have expected count less than 5. The minimum expected count is 2.51.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Phi	.432			.049
	Cramer's V	.249			.049
Interval by Interval	Pearson's R	.334	.098	3.338	.001[c]
Ordinal by Ordinal	Spearman Correlation	.314	.100	3.123	.002[c]
Measure of Agreement	Kappa	.129	.069	2.114	.035
N of Valid Cases		91			

[a] Not assuming the null hypothesis.

[b] Using the asymptotic standard error assuming the null hypothesis.

[c] Based on normal approximation.

c. Cramér's V is a general measure of the correlation between husband and wife's scores. Although it is significant (barely), it is not very high.

d. Odds ratios don't make much sense here because we don't have a basic control condition against which to compare others.

e. Kappa represents a measure of agreement, but if females were shifted slightly up the scale the agreement would change simply because they had a different reference point.

f. Combining categories

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.565[a]	1	.003		
Continuity Correction[b]	7.361	1	.007		
Likelihood Ratio	8.657	1	.003		
Fisher's Exact Test				.005	.003
Linear-by-Linear Association	8.471	1	.004		
N of Valid Cases	91				

[a] 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.14.

[b] Computed only for a 2x2 table

Notice that the result has a much lower probability value. Combining in this way makes sense if the categories are ordered, but would not make much sense if they are not ordered.

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.

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

Seriousness	**Death Penalty**	
	O_{11k}	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

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

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

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

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.

$$OR = \frac{\Sigma(f_{11k} f_{22k} / n_{.k})}{\Sigma(f_{21k} f_{12k} / n_{.k})}$$
$$= \frac{(2*181/244 + 2*21/39 + ... + 17*0/21)}{(60*1/244 + 15*1/39 + ... + 0)4/21)}$$
$$= \frac{8.498}{1.5471} = 5.493$$

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

6.37 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.

6.38 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.

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

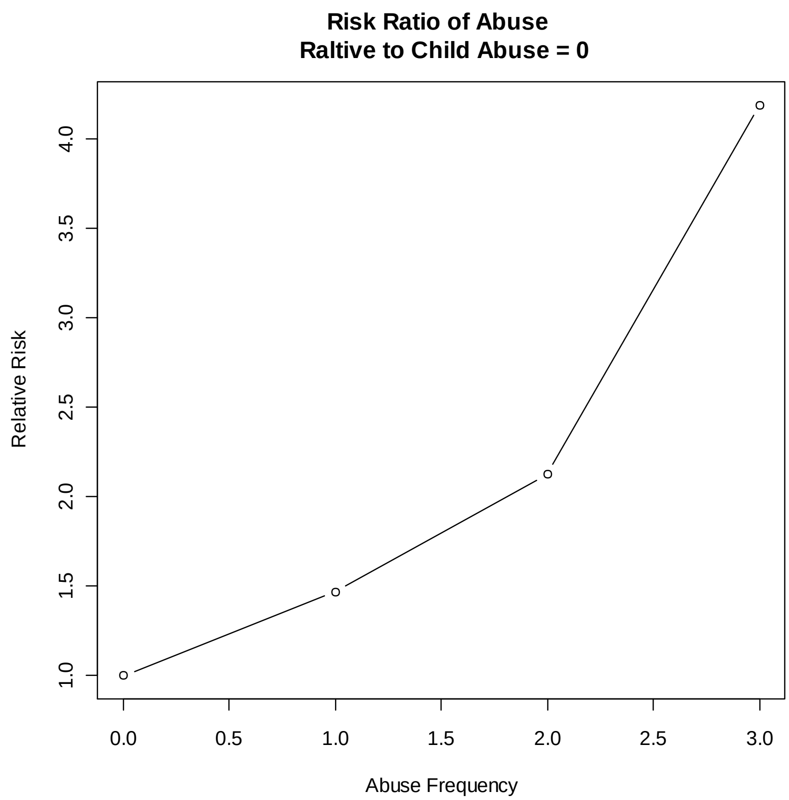
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Dead (1.00 / 2.00)	1.460	1.141	1.868
For cohort Smoker = No	1.173	1.062	1.296
For cohort Smoker = Yes	.804	.693	.932
N of Valid Cases	1314		

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.

6.40 Relative risk in Table 6.12



Chapter 7 - Hypothesis Tests Applied to Means

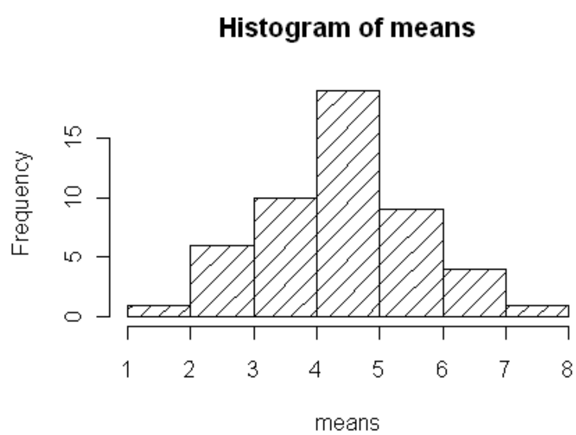
7.1 Distribution of 100 random numbers:

DV

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	7	7.0	7.0	7.0
	1.00	9	9.0	9.0	16.0
	2.00	14	14.0	14.0	30.0
	3.00	9	9.0	9.0	39.0
	4.00	16	16.0	16.0	55.0
	5.00	4	4.0	4.0	59.0
	6.00	10	10.0	10.0	69.0
	7.00	14	14.0	14.0	83.0
	8.00	13	13.0	13.0	96.0
	9.00	4	4.0	4.0	100.0
	Total	100	100.0	100.0	

mean(dv) = 4.46
st. dev(dv) = 2.687
var(dv) = 7.22

7.2 Sampling distribution of means of 50 samples (N = 5) from the distribution of random numbers in Exercise 7.1:



Mean	Frequency
1 - 1.9	1
2 - 2.9	6
3 - 3.9	7
4 - 4.9	20
5 - 5.9	10
6 - 6.9	5
7-7.9	1

mean of means = 4.448
st. dev. of means = 1.198
variance of means = 1.44

7.3 Does the Central Limit Theorem work?

The mean and standard deviation of the sample are 4.46 and 2.69. The mean and standard deviation are very close to the other parameters of the population from which the sample was drawn (4.5 and 2.7, respectively.) The mean of the distribution of means is 4.45, which is close to the population mean, and the standard deviation is 1.20.

Population Parameters	Predictions from Central Limit Theorem	Empirical Sampling distribution
$\mu = 4.5$	$\bar{X} = 4.5$	$\bar{X} = 4.45$
$\sigma^2 = 7.22$	$s^2 = \frac{\sigma^2}{n} = \frac{7.22}{5} = 1.44$	$s^2 = 1.44$

The mean of the sampling distribution is approximately correct compared to that predicted by the Central Limit theorem. The variance of the sampling distribution is almost exactly what we would have predicted..

7.4 The distribution would have been smoother, and the mean and standard error would have been closer to what the Central Limit Theorem would have predicted, but the fundamental properties would stay the same.

7.5 The standard error would have been smaller, because it would be estimated by $\sqrt{\frac{7.29}{15}}$ instead of $\sqrt{\frac{7.29}{5}}$.

7.6 Kruger and Dunning study

$$\bar{X} = 67.9$$

$$s = 12.8$$

$$t = \frac{\bar{X} - 50}{s/\sqrt{n}} = \frac{67.9 - 50}{12.8/\sqrt{11}} = \frac{17.9}{3.89} = 4.64$$

p = .0009 (two-tailed)

These students, who really scored in the lowest quartile estimated that their performance was significantly above average.

7.7 I used a two-tailed test in the last problem, but a one-tailed test could be justified on the grounds that we had no interest is showing that these students thought that they were below average, but only in showing that they thought that they were above average.

7.8 Performance of best performing students

$$\bar{X} = 70$$
$$s = 14.92$$
$$t = \frac{\bar{X} - 86}{s/\sqrt{n}} = \frac{70 - 86}{14.92/\sqrt{11}} = \frac{-16}{4.498} = 3.557$$

This t has a two-tailed probability of .005, which means that this group significantly underestimated their performance. Notice that the estimate from the best scoring group was almost exactly the same as the estimate from the worst performing group.

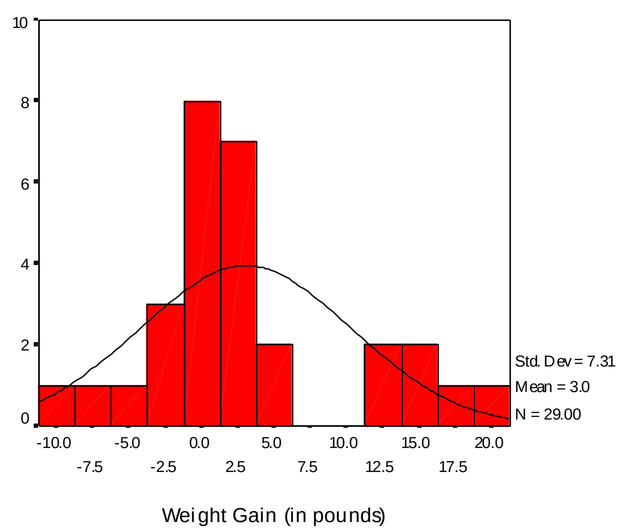
7.9 While the group that was near the bottom certainly had less room to underestimate their performance than to overestimate it, the fact that they overestimated by so much is significant. (If they were in the bottom quartile the best that they could have scored was at the 25th percentile, yet their mean estimate was at the 68th percentile.)

7.10 95% confidence limits on data in Exercise 7.8

$$\begin{aligned} CI_{.95} &= \bar{X} \pm t_{.025,10} s_{\bar{X}} \\ &= 70 \pm (2.228)(14.92/\sqrt{11}) = 70 \pm 2.228 * 4.498 \\ &= 70 \pm 10.02 \end{aligned}$$
$$59.98 \le \mu \le 80.02$$

7.11 Everitt's data on weight gain:

The Mean gain = 3.01, standard deviation = 7.31. $t = 2.22$. With 28 *df* the critical value = 2.048, so we will reject the null hypothesis and conclude that the girls gained at better than chance levels. The effect size is 3.01/7.31 = 0.41.



7.12 Confidence Limits on data for Anorexia:

$$
\begin{aligned}
CI_{.95} &= \bar{X} \pm t_{.025} s_{\bar{X}} \\
&= 3.01 \pm (2.048)\left(7.31/\sqrt{29}\right) = 3.01 \pm (2.048)(1.357) \\
&= 3.01 \pm 2.779
\end{aligned}
$$

$$0.231 \le \mu \le 5.789$$

7.13 **a.** Performance when not reading passage

$$
\begin{aligned}
t &= \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \\
&= \frac{46.6 - 20.0}{\frac{6.8}{\sqrt{28}}} = \frac{26.6}{1.285} = 20.70
\end{aligned}
$$

b. This does not mean that the SAT is not a valid measure, but it does show that people who do well at guessing at answers also do well on the SAT. This is not very surprising.

7.14 Testing the experimental hypothesis that children tend to give socially-approved responses:

a. I would compare the mean of this group to the mean of a population of children tested under normal conditions.

b. The null hypothesis would be that these children come from a population with a mean of 3.87 (the mean of children in general). The research hypothesis would be that these children give socially-approved responses at a different rate from normal children because of the stress they are under.

c.

$$
\begin{aligned}
t &= \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} \\
&= \frac{4.39 - 3.87}{\frac{2.61}{\sqrt{36}}} = \frac{0.52}{0.435} = 1.20
\end{aligned}
$$

With 35 *df* the critical value of t at $\alpha = .05$, two-tailed, is 2.03. We retain H_0 and conclude that we have no reason to think that these stressed children give socially-approved answers at a higher than normal rate.

7.15 Confidence limits on μ for Exercise 7.14:

$$CI_{.95} = \bar{X} \pm t_{.05}\frac{s}{\sqrt{n}}$$
$$= 4.39 \pm 2.03\frac{2.61}{\sqrt{36}} = 4.39 \pm 0.883$$
$$= 3.507 \le \mu \le 5.273$$

An interval formed as this one was has a probability of .95 of encompassing the mean of the population. Since this interval includes the hypothesized population mean of 3.87, it is consistent with the results in Exercise 7.14.

7.16 Beta-endorphin levels:

Gain Scores

10.00	7.50	5.50	6.00	9.50	-2.50	13.00	3.00	-.10	.20	20.30	4.00
8.00	25.00	7.20	35.00	-3.50	-1.90	.10					

Mean = 7.70 St. dev. = 9.945

$$t = \frac{\bar{D} - \mu}{s_{\bar{D}}} = \frac{7.70 - 0.00}{\frac{9.945}{\sqrt{19}}} = \frac{7.70}{2.282} = 3.37$$

Reject H_0 and conclude that beta-endorphin levels were higher just before surgery.

7.17 Confidence limits on beta-endorphin changes:

$$CI_{.95} = \bar{D} \pm t_{.05}\frac{s_D}{\sqrt{n}}$$
$$= 7.70 \pm 2.101\frac{9.945}{\sqrt{19}} = 7.70 \pm 4.794$$
$$= 2.906 \le \mu \le 12.494$$

7.18 Effect size for Exercise 17.16 :

Neither group is a control group, so we can't use that st. dev. as a standardizing constant. It doesn't make a lot of sense to use the standard deviation of the differences. I would be inclined to use the square root of the average of the two variances.

$$s_{pooled} = \sqrt{\frac{s^2_{beta12} + s^2_{beta10}}{2}} = 9.38$$

$$d = \frac{\bar{X}_{12} - \bar{X}_{10}}{s_{pooled}} = \frac{8.35 - 16.05}{9.38} = \frac{-7.7}{9.38} = -0.82$$

If you wanted to use the standard deviation of the differences, *d* would be 0.77.

7.19 Paired *t* test on marital satisfaction:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{D}}{s_{\bar{D}}} = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}$$

$$= \frac{2.725 - 2.791}{\frac{1.30}{\sqrt{91}}} = \frac{-.066}{.136} = -.485$$

We cannot reject the null hypothesis that males and females are equally satisfied. A paired-*t* is appropriate because it would not seem reasonable to assume that the sexual satisfaction of a husband is independent of that of his wife.

7.20 The answer in Exercise 7.19 asks whether males and females are equally satisfied. It does not speak directly to the question of whether there is a relationship between the satisfaction of husbands and wives.

7.21 Correlation between husbands and wives:

$$r = \frac{\text{cov}_{XY}}{s_X s_Y} = \frac{0.420}{\sqrt{(1.357)(1.167)}} = \frac{0.420}{1.584} = \frac{.420}{1.259} = .334$$

The correlation between the scores of husbands and wives was .334, which is significant, and which confirms the assumption that the scores would be related.

7.22 Confidence limits on data in Exercise 7.19:

$$CI_{.95} = \bar{D} \pm t_{.025(90)} s_{\bar{D}}$$

$$= -.066 \pm (1.98)(0.136) = -.066 \pm .269$$

$$-0.335 \le \mu \le 0.203$$

The probability is .95 that an interval constructed as we have constructed this one will include the true mean difference between satisfaction scores of husbands and wives. Since the interval includes 0.00, it is consistent with our *t* test on the difference.

7.23 The important question is what would the sampling distribution of the mean (or differences between means) look like, and with 91 pairs of scores that sampling distribution would be substantially continuous with a normal distribution of means.

7.24 If we wanted to study the effectiveness of two methods of treating breast cancer (radical versus limited mastectomy) we couldn't use the same subjects, since the effects of each treatment would obviously carry over to the other.

7.25 Sullivan and Bybee study:

$$\bar{X}_{\text{int}} = 5.03 \qquad s_{\text{int}} = 1.01 \qquad n_{\text{int}} = 135$$
$$\bar{X}_{ctrl} = 4.61 \qquad s_{\text{int}} = 1.13 \qquad n_{\text{int}} = 130$$

$$t = \frac{\bar{X}_{\text{int}} - \bar{X}_{ctrl}}{\sqrt{\frac{s^2_{\text{int}}}{n_{\text{int}}} + \frac{s^2_{ctrl}}{n_{ctrl}}}} = \frac{5.03 - 4.61}{\sqrt{\frac{1.01^2}{135} + \frac{1.13^2}{130}}}$$
$$= \frac{5.03 - 4.61}{\sqrt{\frac{1.02}{135} + \frac{1.277}{130}}} = \frac{0.42}{\sqrt{0.027}} = \frac{0.42}{0.165} = 2.545$$
$$p(t > abs(2.545)) = .011$$

The quality of life was significantly better for the intervention group.

7.26 Confidence interval for difference of group means in Exercise 7.25

$$\bar{X}_{diff} = 0.42$$
$$se_{diff} = 0.165$$
$$t_{.025,263} = 1.969$$
$$CI_{.95} = \bar{X}_{diff} \pm 1.969 * 0.165 = 0.42 \pm 0.325$$
$$CI_{.95} = 0.095 \le \mu_{diff} \le 0.745$$

Effect Size:

$$s_p = \sqrt{\frac{(n_1 - 1)s^2_{\text{int}} + (n_2 - 1)s^2_{ctrl}}{n_1 + n_2 - 2}} = \sqrt{\frac{134 * 1.02 + 129 * 1.13}{134 + 129}} = \sqrt{\frac{282.45}{263}} = \sqrt{1.07} = 1.04$$
$$d = \frac{\bar{X}_{\text{int}} - \bar{X}_{ctrl}}{s_p} = \frac{5.03 - 4.61}{1.04} = \frac{0.42}{1.04} = 0.40$$

7.27 Paired t-test on before and after intervention quality of life

$$\bar{X}_{before} = 4.47 \quad \bar{X}_{after} = 5.03 \quad s_{diff} = 1.30 \quad n = 135$$

$$t = \frac{\bar{D} - 0}{\frac{s_{diff}}{\sqrt{n}}} = \frac{5.03 - 4.47}{\frac{1.30}{\sqrt{135}}} = \frac{0.56}{.006} = 93.33$$

$$p < .000$$

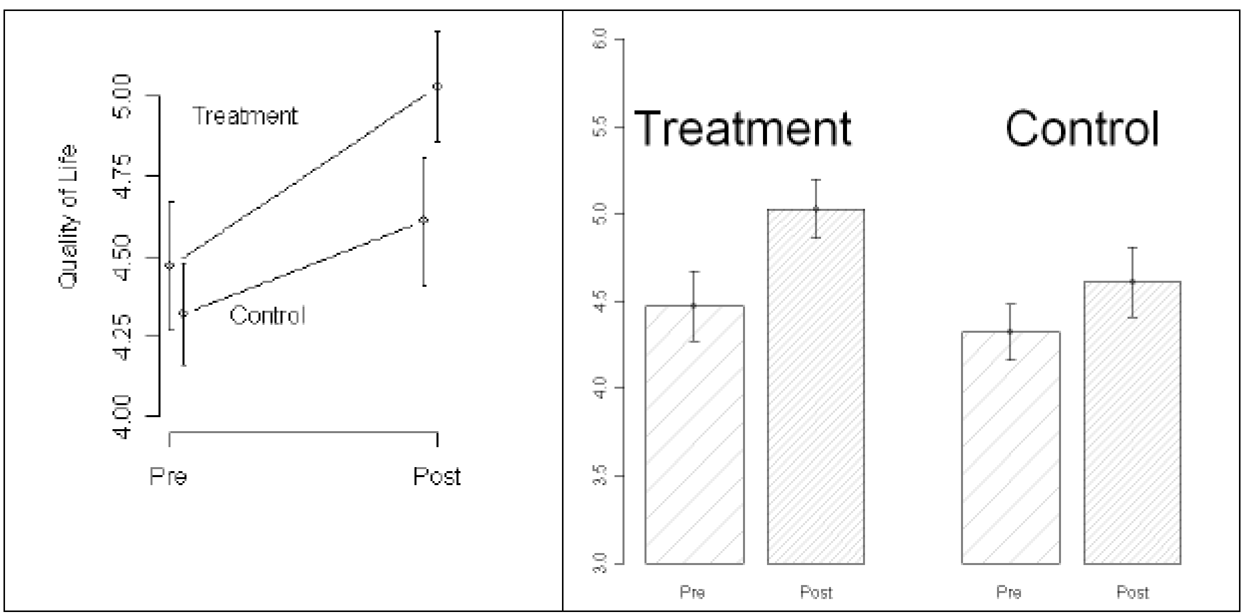
Confidence limits on weight gain in Cognitive Behavior Therapy group:

$$CI_{.95} = \bar{D} \pm t_{.025(28)} s_{\bar{D}}$$
$$= 3.02 \pm (2.05)(1.357) = 3.02 \pm 2.78$$
$$0.24 \le \mu \le 5.80$$

The probability is .95 that this procedure has resulted in limits that bracket the mean weight gain in the population.

7.28 Pre-Post scores for both groups

This can be done as line graphs or as bar plots—I have done it both ways. The error bars are calculated as $\bar{X} \pm t_{.025,df}\, s/\sqrt{n}$, where the means and standard deviations are given in the problem and n = 135 or 130.



Although both groups increased their ratings of quality of life, the treatment group increased more.

7.29 Katz et al (1990) study

a. Null hypothesis—there is not a significant difference in test scores between those who have read the passage and those who have not.

b. Alternative hypothesis—there is a significant difference between the two conditions.

c.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{16(10.6^2) + 27(6.8^2)}{17 + 28 - 2} = \frac{3046.24}{43} = 70.843$$

$$t = \frac{69.6 - 46.6}{\sqrt{\frac{70.843}{17} + \frac{70.843}{28}}} = \frac{23.0}{\sqrt{70.843\left(\frac{1}{17} + \frac{1}{28}\right)}} = \frac{23.0}{\sqrt{6.697}} = 8.89$$

t = 8.89 on 43 *df* if we pool the variances. This difference is significant.

d. We can conclude that students do better on this test if they read the passage on which they are going to answer questions.

7.30 Depression in new mothers:

The simplest approach would be to obtain an unselected sample of mothers who are in their first trimester of pregnancy and obtain a depression measure on each of them. Some time after they give birth we would obtain another depression score from the same mothers and compare the two means. (The length of the post-birth interval would be crucial.) An alternative approach would be to unsystematically collect a sample of new mothers and a sample of non-mothers of the same age and environmental characteristics and obtain depression measures from each sample. There would probably be greater variability in the second approach, but you would have the advantage of matching on environmental characteristics. Doing this would help to rule out alternative explanations for any change in depression.

7.31

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{25(63.82) + 28(53.41)}{26 + 29 - 2} = \frac{3090.98}{53} = 58.32$$

$$t = \frac{-0.45 - 3.01}{\sqrt{\frac{58.32}{26} + \frac{58.32}{29}}} = \frac{-3.46}{\sqrt{58.32\left(\frac{1}{26} + \frac{1}{29}\right)}} = \frac{-3.46}{\sqrt{4.254}} = \frac{--3.46}{2.062} = -1.68$$

A t on two independent groups = -1.68 on 53 *df*, which is not significant. Cognitive behavior therapy did not lead to significantly greater weight gain than the Control condition. (Variances were homogeneous.)

7.32 Confidence interval of difference in weight gain:

$$CI_{.95} = \left(\bar{X}_1 - \bar{X}_2\right) \pm t_{.025(53)} s_{\bar{X}_1 - \bar{X}_2}$$

$$= -3.46 \pm (2.006)\left(\sqrt{4.254}\right) = -3.46 \pm 0.677$$

$$-7.597 \le \mu_1 - \mu_2 \le 0.677$$

7.33 If those means had actually come from independent samples, we could not remove differences due to couples, and the resulting t would have been somewhat smaller.

7.34 Analysis of Exercise 7.19 treating samples as independent.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{91\left(1.16^2\right) + 91\left(1.08^2\right)}{91 + 91 - 2} = \frac{228.592}{180} = 1.27$$

$$t = \frac{2.791 - 2.725}{\sqrt{\frac{1.27}{91} + \frac{1.27}{91}}} = \frac{0.066}{\sqrt{0.028}} = 0.39$$

7.35 The difference between the two answers in not greater than it is because the correlation between husbands and wives was actually quite low.

7.36 Random assignment assures that any differences between the groups will be attributable to the different ways in which the groups were treated, not to other differences that might exist if we used nonrandom assignment. Often people do not want to participate if they are just going to serve in a control group, and therefore the people who are in that group will not be a random selection from those available for the study.

7.37 **a.** I would assume that the experimental hypothesis is the hypothesis that mothers of schizophrenic children provide TAT descriptions that show less positive parent-child relationships.

b.

Normal	Mean = 3.55	$s = 1.887$	$n = 20$
Schizophrenic	Mean = 2.10	$s = 1.553$	$n = 20$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{3.55 - 2.10}{\sqrt{\frac{1.887^2}{20} + \frac{1.553^2}{20}}}$$

$$= \frac{1.45}{\sqrt{0.299}} = \frac{1.45}{0.546} = 2.66$$

[$t_{.05}(38) = \pm 2.02$] Reject the null hypothesis

This t is significant on 38 *df*, and I would conclude that the mean number of pictures portraying positive parent-child relationships is lower in the schizophrenic group than in the normal group.

7.38 In Exercise 7.37 it could well have been that there was much less variability in the schizophrenic group than in the normal group because the number of TATs showing positive parent-child relationships could have an a floor effect at 0.0. The fact that this did not happen does not mean that it is important to check. The fact that sample sizes were equal makes this less of a problem if it did happen.

7.39 There is no way to tell cause and effect relationships in Exercise 7.37. It could be that people who experience poor parent-child interaction are at risk for schizophrenia. But it could also be that schizophrenic children disrupt the family and poor relationships come as a result.

7.40 Experimenter bias effect:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \qquad \text{where} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{9(15.44) + 7(17.41)}{9 + 8 - 2} = 16.362$$

$$t = \frac{18.778 - 17.625}{\sqrt{\frac{16.362}{9} + \frac{16.362}{8}}} = \frac{1.153}{\sqrt{3.863}} = \frac{1.153}{1.966} = 0.586$$

[$t_{.05}(15) = \pm 2.13$]

Do not reject the null hypothesis. There is no evidence of an experimenter bias effect in these data.

7.41 95% confidence limits:

$$CI_{.05} = \left(\bar{X}_1 - \bar{X}_2\right) \pm t_{.025}\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$$

$$= (18.778 - 17.625) \pm (2.131)\sqrt{\frac{16.362}{9} + \frac{16.362}{8}} = 1.153 \pm 4.189$$

$$-3.036 \le (\mu_1 - \mu_2) \le 5.342$$

7.42 Problem solving versus time-filling instructions:

(We do not need to pool variances because we have equal sample sizes.)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{5.4 - 8.4}{\sqrt{\frac{4.3}{5} + \frac{3.8}{5}}} = \frac{-3.00}{\sqrt{1.62}} = \frac{-3.00}{1.273} = -2.36$$

[$t_{.025(8)} = \pm 2.306$]

Reject the null hypothesis.

7.43 Repeating Exercise 7.42 with time as the dependent variable:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

$$t = \frac{2.102 - 1.246}{\sqrt{\dfrac{0.714}{5} + \dfrac{0.091}{5}}} = \frac{0.856}{\sqrt{0.161}} = \frac{0.856}{0.401} = 2.134$$

The variances are very different, but even if we did not adjust the degrees of freedom, we would still fail to reject the null hypothesis.

7.44 Perfectly legitimate and reasonable transformations of the data can produce quite different results. It is important to consider seriously the nature of the dependent variable before beginning an experiment.

7.45 If you take the absolute differences between the observations and their group means and run a t test comparing the two groups on the absolute differences, you obtain $t = 0.625$. Squaring this you have $F = 0.391$, which makes it clear that Levene's test in SPSS is operating on the absolute differences. (The t for squared differences would equal 0.213, which would give an F of 0.045.)

7.46 Data on young adults who had lost a parent:

(We can assume homogeneity of variance in each case.)

Independent Samples Test

Equal variances assumed

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DEPRESST	.298	314	.766	.318	1.066	-1.780	2.415
ANXT	.624	314	.533	.674	1.080	-1.451	2.798
GSIT	.270	314	.788	.275	1.021	-1.733	2.284

b. The tests are not independent because they involve the same participants.

7.47 Differences between males and females on anxiety and depression:

(We cannot assume homogeneity of regression here.)

Independent Samples Test

Equal variances not assumed

	t-test for Equality of Means					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DEPRESST	3.256	248.346	.001	3.426	1.052	1.353	5.499
ANXT	1.670	246.260	.096	1.805	1.081	-.324	3.933

7.48 Pairwise comparisons among groups:

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
GSIT	Assume equal variances	1 vs 2	.275	1.051	.262	372	.794
		1 vs 3	1.881	1.443	1.304	372	.193
		2 vs 3	1.606	1.386	1.159	372	.247
	Does not assume equal variances	1 vs 2	.275	1.038	.265	269.575	.791
		1 vs 3	1.881	1.604	1.173	101.167	.244
		2 vs 3	1.606	1.516	1.059	83.935	.292

7.49 Effect size for data in Exercise 7.25:

$$d = \frac{\bar{X}_{After} - \bar{X}_{Before}}{s_{Before}} = \frac{3.02}{4.85} = 0.62$$

I chose to use the standard deviation of the before therapy scores because it provides a reasonable base against which to standardize the mean difference. The confidence intervals on the difference, which is another way to examine the size of an effect, were given in the answer to Exercise 7.27.

7.50 Effect size for data in Exercise 7.31:

$$d = \frac{\bar{X}_1 \bar{X}_2}{s_p} = \frac{-0.45 - 3.01}{\sqrt{63.82}} = \frac{-3.46}{7.99} = -0.43$$

The two means are approximately ½ a standard deviation apart. (I used the standard deviation of the control group in calculating *d*.

7.51 **a.** The scale of measurement is important because if we rescaled the categories as 1, 2, 4, and 6, for example, we would have quite different answers.

b. The first exercise asks if there is a relationship between the satisfaction of husbands and wives. The second simply asks if males (husbands) are more satisfied, on average, than females (wives).

c. You could adapt the suggestion made in the text about combining the *t* on independent groups and the *t* on matched groups.

d. I'm really not very comfortable with the *t* test because I am not pleased with the scale of measurement. An alternative would be a ranked test, but the number of ties is huge, and that probably worries me even more.

7.52 Everitt (in Hand, 1994) compared the weight gain in a group receiving cognitive behavior therapy and a Control group receiving no therapy. The Control group lost 0.45 pounds over the interval, while the cognitive behavior therapy group gained 3.01 pounds. This difference was statistically not significant (t (53) = -1.676, $p < .05$). Using the standard deviation of the control group to calculate *d*, the effect size measure for this difference produced d =- 0.43, indicating that the groups differed by less than one half of a standard deviation. (Because the effect was not significant, though it would be significant with a one-tailed test, which Jones and Tukey would probably suggest, it is difficult to know what to make of this value of *d*.)

Chapter 8 - Power

8.1 Peer pressure study:

a.

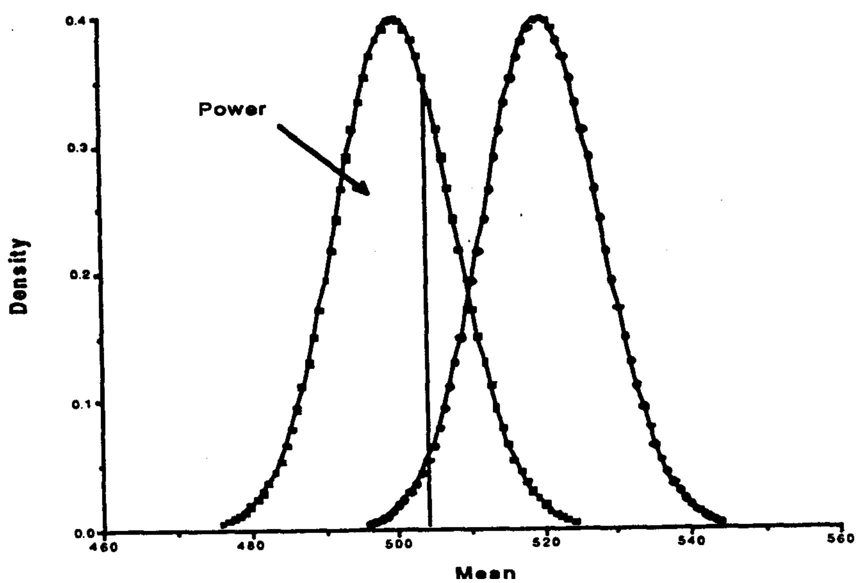
$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{520 - 500}{80} = .25$$

b. $f(n)$ for 1-sample t-test = $\sqrt{n}$

$$\delta = d\sqrt{n} = .25\sqrt{100} = 2.5$$

c. Power = .71

8.2 Sampling distributions of the mean for situation in Exercise 8.1:



8.3 Changing power in Exercise 8.1:

a. For power = .70, δ = 2.475

$$\delta = d\sqrt{n}$$
$$2.475 = .25\sqrt{n}$$
$$n = 98.01 \approx 99 \text{ (Round up, because students come in whole lots)}$$

b. For power = .80, δ = 2.8

$$\delta = d\sqrt{n}$$
$$2.8 = .25\sqrt{n}$$
$$n = 125.44 \approx 126 \text{ (Round up)}$$

c. For power = .90, δ = 3.25

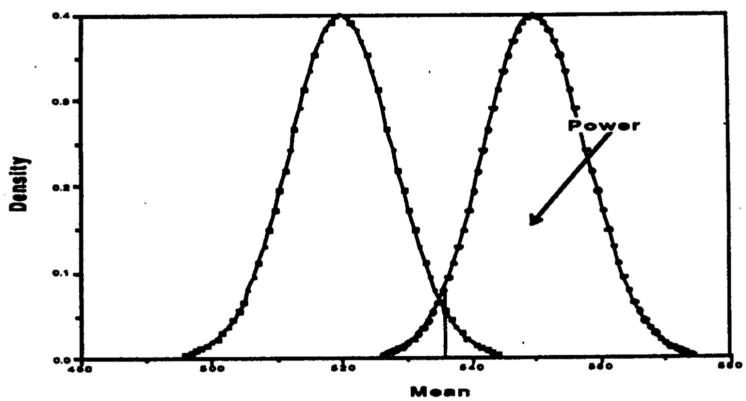
$$\delta = d\sqrt{n}$$
$$3.25 = .25\sqrt{n}$$
$$n = 169$$

8.4 Alternative peer pressure study:

$$d = \frac{30}{80} = .375$$
$$\delta = .375\sqrt{100}$$
$$= 3.75$$

power = .965

8.5 Sampling distributions of the mean for the situation in Exercise 8.4:



8.6 Combining Exercises 8.1 and 8.4:

a. The experimenter expects that one mean will be 550 and the other mean will be 500. She assumes a population standard deviation of 80. Therefore **d** = (550 - 500)/80 = .625.

b.

$$\delta = d\sqrt{\frac{n}{2}}$$

$$= .625\sqrt{\frac{50}{2}} = 3.125$$

c. Power = .88

8.7 Avoidance behavior in rabbits using 1-sample t test:

a.

$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{5.8 - 4.8}{2} = \frac{1}{2} = .50$$

For power = .50, δ = 1.95

$$\delta = d\sqrt{n}$$

$$1.95 = .5\sqrt{n}$$

$$n = 15.21 \approx 16$$

b. For power = .80, $\delta = 2.8$

$$\delta = d\sqrt{n}$$
$$2.8 = .5\sqrt{n}$$
$$n = 31.36 \approx 32$$

8.8 Avoidance behavior in rabbits using 2-sample *t* test:

a. For 2-sample *t* test f(*n*) = $\sqrt{n/2}$

For power = .60, $\delta = 2.2$

$$\delta = d\sqrt{n/2}$$
$$2.2 = .5\sqrt{n/2}$$
$$n = 38.72 \approx 39 \text{ in each group, or 78 overall}$$

b. For power = .90, $\delta = 3.25$

$$\delta = d\sqrt{n/2}$$
$$3.25 = .5\sqrt{n/2}$$
$$n = 84.5 \approx 85 \text{ in each group, or 170 overall}$$

8.9 Avoidance behavior in rabbits with unequal *N*s:

$$d = .5$$
$$n = \bar{n}_h = \frac{2n_1n_2}{n_1 + n_2}$$
$$= \frac{2(20)(15)}{20+15} = 17.14$$
$$\delta = d\sqrt{\frac{n}{2}} = 5\sqrt{\frac{17.14}{2}} = 1.46$$

power = .31

8.10 Cognitive development of LBW and normal babies at 1 year:

$$d = \frac{\mu_2 - \mu_1}{\sigma} = \frac{30-25}{8} = 0.625$$

$$\delta = \mathrm{d}\sqrt{n/2} = .625\sqrt{20/2} = 1.98$$

power= .51

8.11 *t* test on data for Exercise 8.10

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$= \frac{25 - 30}{\sqrt{\frac{64}{20} + \frac{64}{20}}}$$

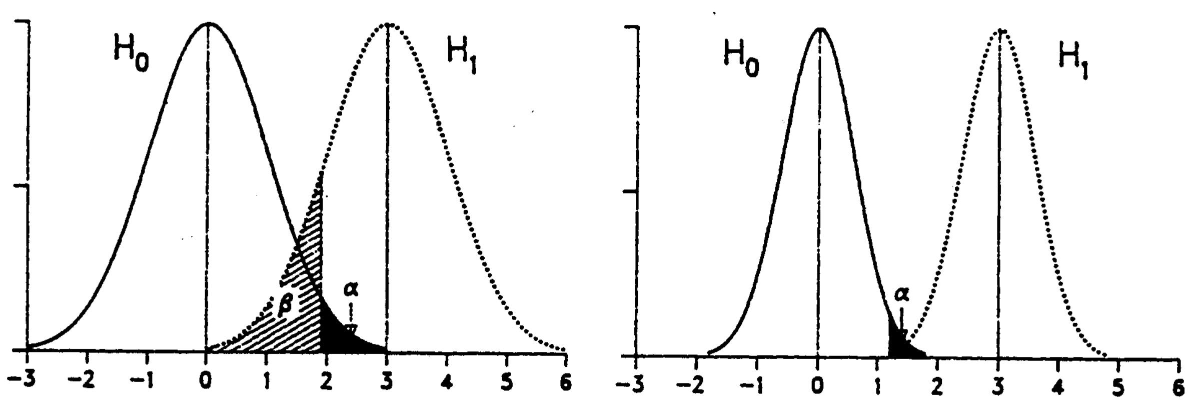
$$= -1.98$$

[$t_{.025}(38) = \pm 2.025$] Do not reject the null hypothesis

c. t is numerically equal to δ although t is calculated from statistics and δ is calculated from parameters. In other words, δ = the t that you would get if the data exactly match what you think are the values of the parameters.

8.12 The first one. A significant t with a smaller n is the more impressive, and since a significant difference was found with an experiment having relatively little power, the first experimenter is presumably dealing with a fairly large effect.

8.13 Diagram to defend answer to Exercise 8.12:



With larger sample sizes the sampling distribution of the mean has a smaller standard error, which means that there is less overlap of the distributions. This results in greater power, and therefore the larger n's significant result was less impressive.

8.14 Power increases as sample sizes become more nearly equal:

	Exp. 1	Exp. 2	Exp. 3	Calculations
$n_1 =$	25	20	15	$\bar{n}_{h1} = \dfrac{3}{\frac{1}{25}+\frac{1}{5}} = 8.33$
$n_2 =$	5	10	15	$\bar{n}_{h2} = \dfrac{2}{\frac{1}{20}+\frac{1}{10}} = 13.33$
$\bar{n}_h =$	8.33	13.33	15.00	
				Assume **d** = .50
$\delta =$	1.02	1.29	1.37	$\delta_1 = .5\sqrt{\dfrac{8.33}{2}} = 1.02$
Power =	0.18	0.25	0.28	$\delta_1 = .5\sqrt{\dfrac{13.33}{2}} = 1.29$
				$\delta_1 = .5\sqrt{\dfrac{15}{2}} = 1.37$

8.15 Social awareness of ex-delinquents--which subject pool would be better to use?

$\bar{X}_{\text{normal}} = 38$ $\quad n = 50$

$\bar{X}_{\text{H.S. Grads}} = 35$ $\quad n = 100$

$\bar{X}_{\text{dropout}} = 30$ $\quad n = 25$

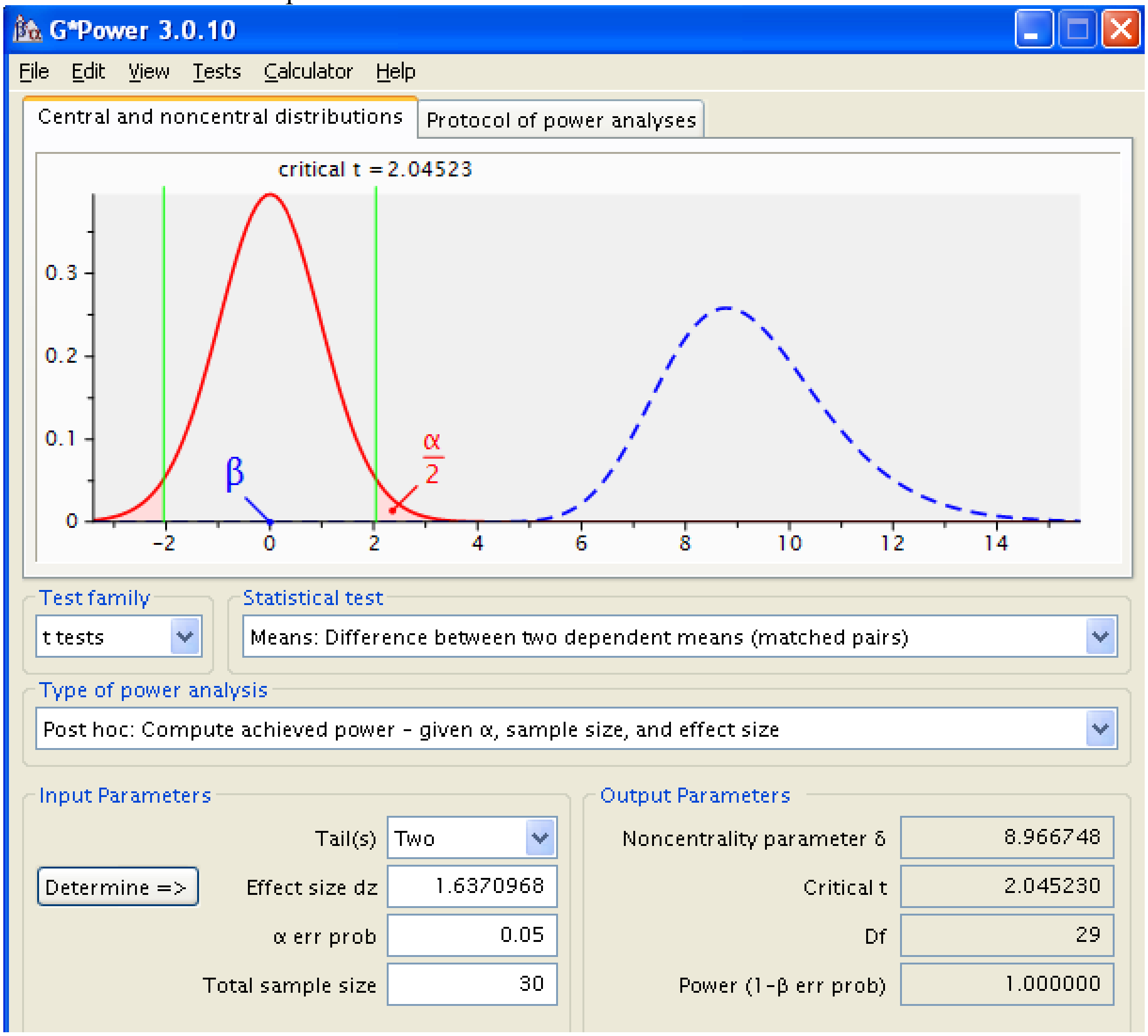
$$\mathrm{d} = \frac{38-35}{\sigma} \qquad \mathrm{d} = \frac{38-30}{\sigma}$$

$$\bar{n}_h = \frac{2(50)(100)}{150} = 66.67 \qquad \bar{n}_h = \frac{2(50)(25)}{75} = 33.33$$

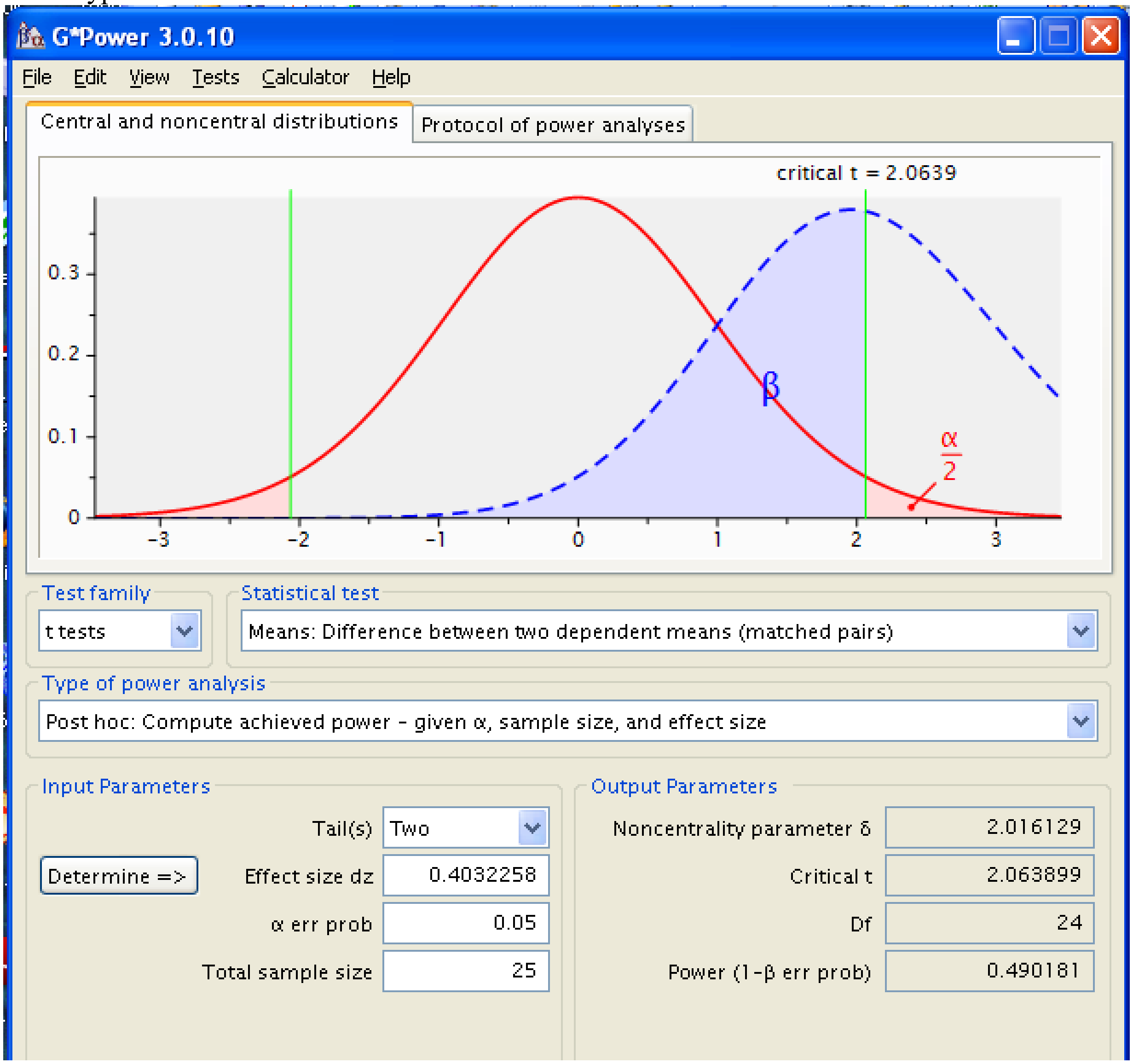
$$\delta = \frac{3}{\sigma}\sqrt{\frac{66.67}{2}} = \frac{17.32}{\sigma} \qquad \delta = \frac{8}{\sigma}\sqrt{\frac{33.33}{2}} = \frac{32.66}{\sigma}$$

Assuming equal standard deviations, the H.S. dropout group of 25 would result in a higher value of δ and therefore higher power. (You can let σ be any value you choose, as long as it is the same for both calculations. Then calculate δ for each situation.)

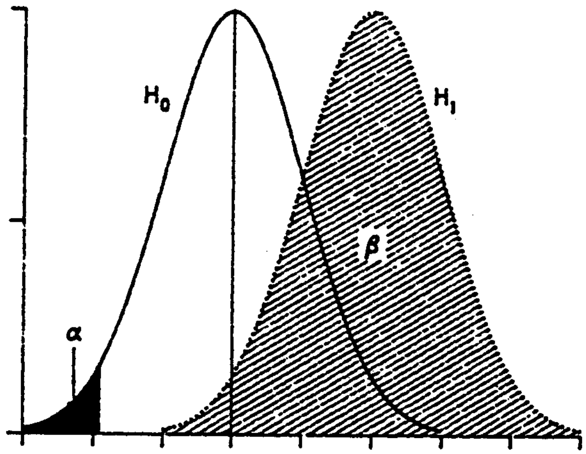
8.16 Power for example in Section 8.5



8.17 Stereotyped threat in women

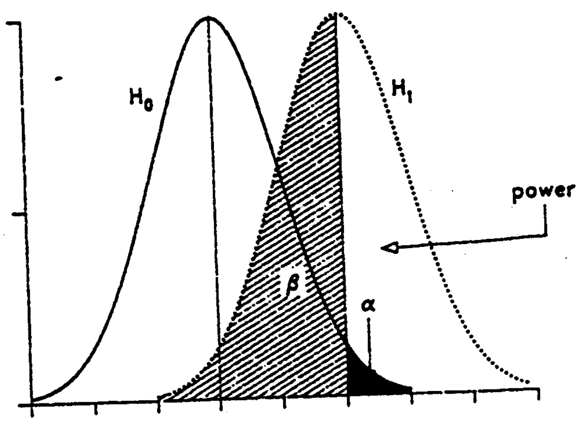


Here the power is about one half of what it was in the study using men, reflecting the fact that our group of men had a stronger identification with their skills in math.



8.18 Can power ever be less than α?
Not unless we choose the wrong tail for our one-tailed test. In that case power could be approximately zero.

8.19 When can power = β?



The mean under H_1 should fall at the critical value under H_0. The question implies a one-tailed test. Thus the mean is 1.645 standard errors above μ_0, which is 100.

$$\begin{aligned}\mu &= 100 + 1.64\sigma_X \\ &= 100 + 1.645\left(15/\sqrt{25}\right) \\ &= 104.935\end{aligned}$$

When $\mu = 104.935$, power would equal β.

8.20 I don't see that Prentice and Miller (1992) are really talking about experiments with small power. They are talking about relatively small experimental manipulations, but those manipulations are sufficient to generate enough of a group difference for the effect to be apparent.

Here I am trying to get students to think about what we mean by power and what we mean by small effects. I would also like them to come to realize that we don't have to find a huge difference between two means for the result to be meaningful.

8.21 Aronson's study:

a. The study would confound differences in lab that have nothing to do with the independent variable with the effect of that variable. You would not be able to draw sound conclusions unless you could persuade yourself that the labs were similar in all other relevant ways.

b. I would randomize the conditions across all of the students in the two labs combined.

c. The stereotypes do not apply to women, so I don't have any particular hypothesis about what would happen.

8.22 **a**. The control condition has to come first or else you will "tip off" the students as to the purpose of the study. It would be impossible to give the threat condition first and then expect that students would respond neutrally to the control condition.

b. I probably can't get around the problem directly, so I would have two sets of problems and randomize the order of presentation over weeks. (I could still have the control condition first, but simply randomize which questions the students receive.)

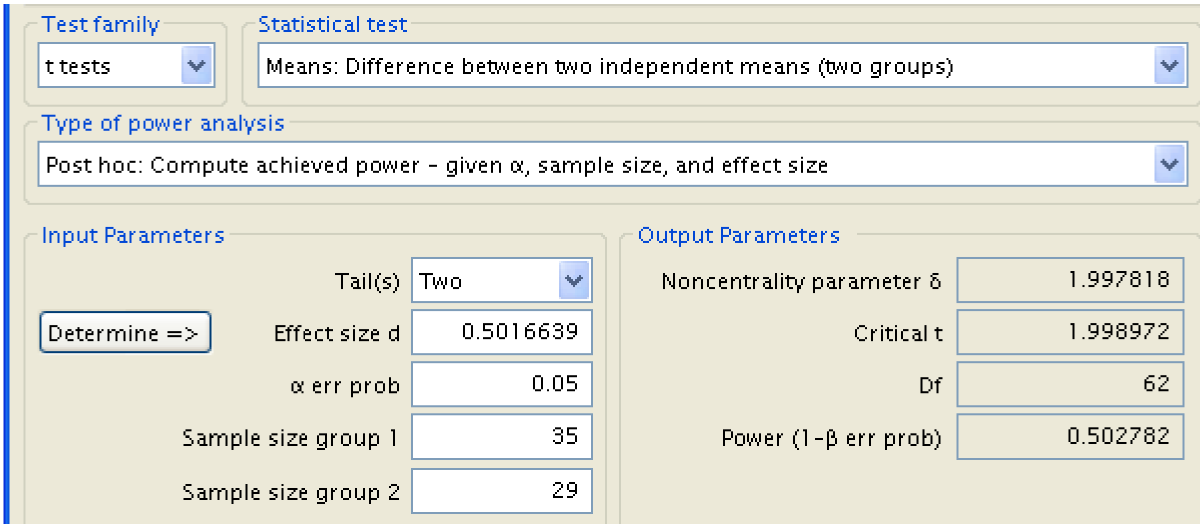
8.23 Both of these questions point to the need to design studies carefully so that the results are clear and interpretable.

8.24 Going back to the study by Adams et al. (1996) of homophobia, discussed in Section 7.5, assume that the homophobic group had a mean of 22.53 instead of 24, but that all other statistics were the same. Then

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(22.53 - 16.50)}{\sqrt{\frac{144.48}{35} + \frac{144.48}{29}}} = \frac{6.03}{\sqrt{9.11}} = 2.00$$

The critical value for $t_{.95,\,62}$ is 1.999, so this difference would barely be significant using a two-tailed test at $\alpha = .05$.

Now using G*Power we find:

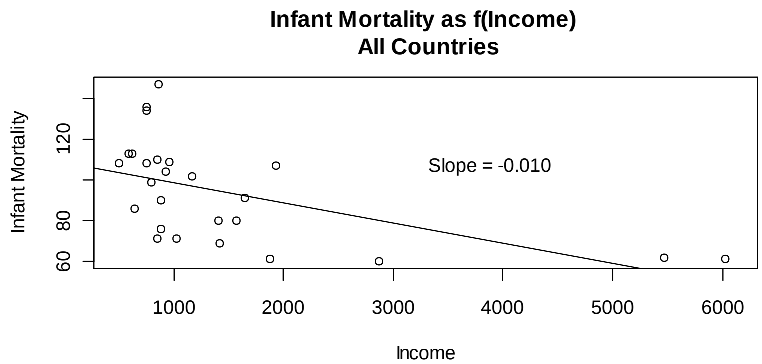


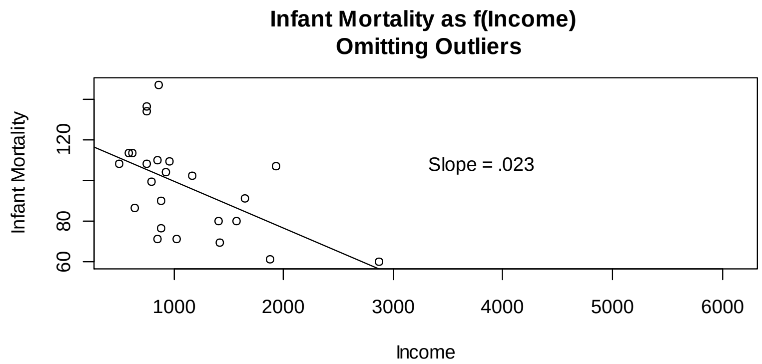
which shows that the power is .50. In other words if a test is just barely significant, you have a 50-50 chance of finding it significant in a follow-up study if you have estimated the parameters correctly.

Chapter 9 - Correlation and Regression

9.1 Infant Mortality in Sub-Saharan Africa

a. & b.





c. Those two points would almost certainly draw the line toward them, which will flatten the slope. If we remove those countries we have the second graph with a steeper slope.

9.2 Intercorrelation matrix

	InfMort	Income	PctMomLE20	PctMomGT40	LT2YrsApart	PctUsingContr	PctNeedFamPlan
InfMort	1.00	-0.56	0.22	-0.05	0.12	-0.44	-0.26
Income	-0.56	1.00	0.06	0.04	-0.14	0.33	-0.06
PctMomLE20	0.22	0.06	1.00	-0.58	0.01	0.30	-0.46
PctMomGT40	-0.05	0.04	-0.58	1.00	-0.19	-0.14	0.23
LT2YrsApart	0.12	-0.14	0.01	-0.19	1.00	-0.32	0.01
PctUsingContr	-0.44	0.33	0.30	-0.14	-0.32	1.00	-0.17
PctNeedFamPlan	-0.26	-0.06	-0.46	0.23	0.01	-0.17	1.00

9.3 Significance of correlations

The minimum sample size in this example is 25, and we will use that. We would need t = 2.069 for a two-tailed test on $N - 2 = 23$ *df.* A little (well, maybe a lot) of algebra will show that a correlation of .396 will produce that t value.

9.4 The strongest predictor of infant mortality is by far the family income, followed by the percentage of mothers using family planning.

9.5 If we put these two predictors together using methods covered in Chapter 15, the multiple correlation will be .58, which is only a small amount higher than Income alone.

9.6 As mentioned in Exercise 9.5, the increase top the correlation is minor. This is most likely due to the fact that there is a correlation between contraception and income, so that the two variables are not adding independent pieces of information.

9.7 I suspect that a major reason why this variable does not play a more important role is the fact that it has very little variance. The range is 3% - 7%. One cause of this may be the very high death rate among women in sub-saharan Africa. There are many fewer women giving birth at ages above 40. To quote from a United Nations report (http://www.un.org/ecosocdev/geninfo/women/women96.htm):

- Women are becoming increasingly affected by HIV. Today about 42 per cent of estimated cases are women, and the number of infected women is expected to reach 15 million by the year 2000.

- An estimated 20 million unsafe abortions are performed worldwide every year, resulting in the deaths of 70,000 women.

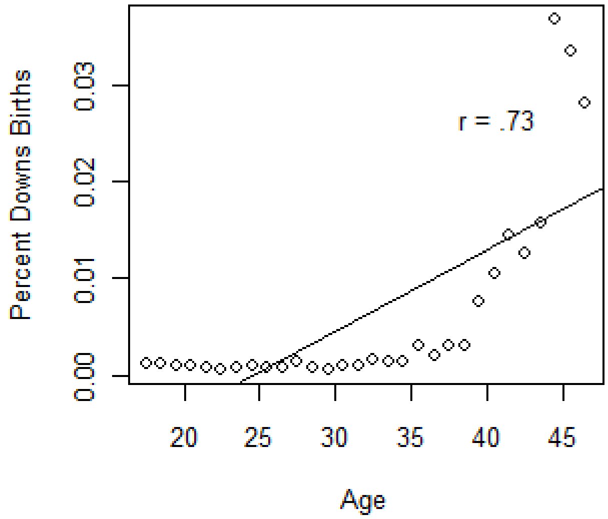
- Approximately 585,000 women die every year, over 1,600 every day, from causes related to pregnancy and childbirth. In sub-Saharan Africa, 1 in 13 women will die from pregnancy or childbirth related causes, compared to 1 in 3,300 women in the United States.

- Globally, 43 per cent of all women and 51 per cent of pregnant women suffer from iron-deficiency anemia.

9.8 Low income is associated with a lot of other variables that would contribute to infant mortality, and it is likely that it is not a cause by itself. It certainly is *associated* with infant mortality.

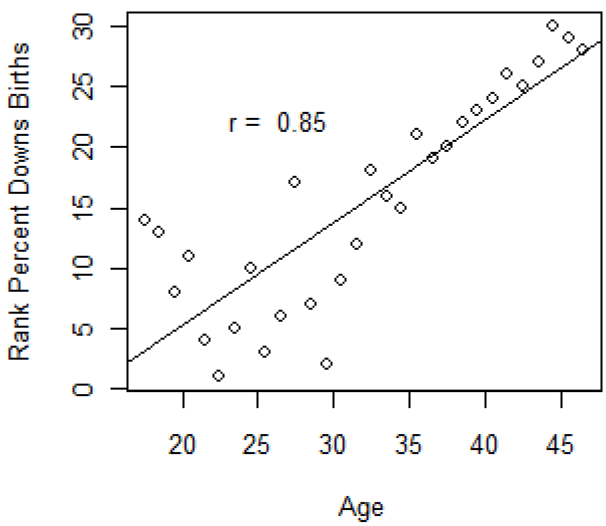
9.9 Psychologists are very much interested in studying variables related to behavior and in finding ways to change behavior. I would guess that they would have a good deal to say about educating women in ways that would decrease infant mortality.

9.10 Scatterplot:



9.11 The relationship is decidedly curvilinear, and Pearson's *r* is a statistic on linear relationships.

9.12 Using ranks of percent Downs births



This is technically not a Spearman correlation because Age is not ranked. However the age categories are equally spaced between 17.5 and 46.5, which will have the same effect as the ranks because it is a perfect linear transformation of ranks.

9.13 Power for n = 25, ρ = .20

$$d = \rho_1 = .20$$
$$\delta = \rho_1\sqrt{N-1} = .20\sqrt{24} = 0.98$$
$$\text{power} \approx .17$$

9.14 Sample sizes needed for power = .80

$$d = \rho_1 = .20$$
$$\delta = \rho_1\sqrt{N-1}$$
$$2.8^2 = \rho_1^2(N-1) = .04(N-1)$$
$$N-1 = 7.84/.04 = 196$$
$$N = 197$$

9.15 Number of symptoms predicted for a stress score of 8 using the data in Table 9.2 :

Regression equation: $Y = 0.0086(X) + 4.30$

If Stress score (X) = 8: $Y = 0.0086(8) + 4.30$

Predicted ln(symptoms) score is : $Y = 4.37$

9.16 Number of symptoms predicted for a mean stress score using the data in Table 9.2.

Regression equation: $Y = 0.0086(X) + 4.30$

If Stress score (X) = 21.467: Y = 0.0086(21.467) + 4.30 = 4.48

Predicted Number of symptoms: Y = 90.701, which is $\bar{Y}$

9.17 Confidence interval on Y :

I will calculate them for X incrementing between 0 and 60 in steps of 10

$$CI(Y) = Y \pm t_{\alpha/2}\left(s'_{Y.X}\right)$$

$$s'_{Y.X} = s_{Y.X}\sqrt{1+\frac{1}{N}+\frac{\left(X_i-\bar{X}\right)^2}{(N-1)s_X^2}} = 0.1726\sqrt{1+\frac{1}{107}+\frac{\left(X_i-\bar{X}\right)^2}{106(156.05)}}$$

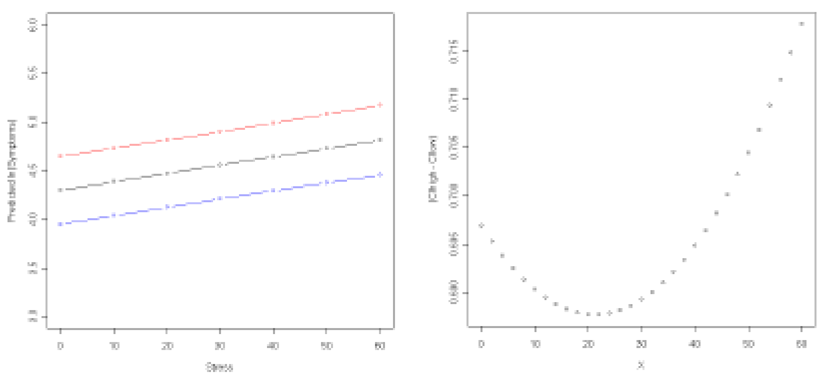
$$Y = 0.00856X + 4.30$$
$$t_{\alpha/2} = 1.983$$

For X from 0 to 60 in steps of 10, $s'_{Y.X} =$
0.1757 0.1741 0.1734 0.1738 0.1752 0.1776 0.1810

$$CI(Y) = \hat{Y} \pm (t_{\alpha/2})(s'_{Y.X})$$

For several different values of X, calculate Y and $s'_{Y.X}$ and plot the results.

X =	0	10	20	30	40	50	60
Y =	4.300	4.386	4.471	4.557	4.642	4.728	4.814



The curvature is hard to see, but it is there, as can be seen in the graphic on the right, which plots the width of the interval as a function of X. (It's fun to play with R).

9.18 When data are standardized, the slope equals r. Therefore the slope will be less than one for all but the most trivial case, and predicted deviations from the mean will be less than actual parental deviations.

9.19 Galton's data

a.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	23.942	2.811		8.517	.000
	midparent	.646	.041	.459	15.711	.000

a. Dependent Variable: child

b. Predicted height = 0.646*(Midparent) + 23.942

c. Child Means

Descriptives

child

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1	392	67.12	2.247	.113	66.90	67.35
2	219	68.02	2.240	.151	67.72	68.32
3	183	68.71	2.465	.182	68.35	69.06
4	134	70.18	2.269	.196	69.79	70.57
Total	928	68.09	2.518	.083	67.93	68.25

Parent means

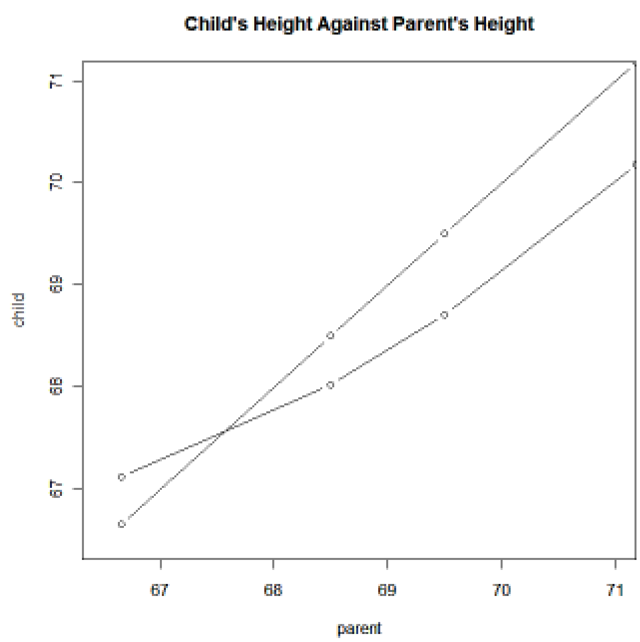
Descriptives

midparent

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
1	392	66.66	1.068	.054	66.56	66.77
2	219	68.50	.000	.000	68.50	68.50
3	183	69.50	.000	.000	69.50	69.50
4	134	71.18	.786	.068	71.04	71.31
Total	928	68.31	1.787	.059	68.19	68.42

d. Parents in the highest quartile have a mean of 71.18, while their children have a mean of 70.18. Those parents in the lowest quartile have a mean of 66.66, while their children have a mean of 67.14. This is what we would expect to happen.

e.



9.20 Power for study of relationship between the amount of money school districts spend on education, and the performance of students on a standardized test such as the SAT:

$$\delta = \rho_1\sqrt{N-1}$$
$$= .40\sqrt{30-1} = 2.154$$

Power = 0.58

9.21 Number of subjects needed in Exercise 9.20 for power = .80:

For power = .80, $\delta = 2.80$

$$\delta = \rho_1\sqrt{N-1}$$
$$2.80 = .40\sqrt{N-1}$$
$$\sqrt{N-1} = 2.80 / .40 = 7$$
$$N = 50$$

9.22 Guber's data on educational expenditures

The data would *appear* to suggest that as expenditures increase, school performance decreases. We will later see that this is very misleading.

9.23 Katz et al. correlations with SAT scores.

a. $r_1 = .68 \quad r_1' = .829$

$r_2 = .51 \quad r_2' = .563$

$$z = \frac{r_1' - r_2'}{\sqrt{\frac{1}{N_1-3}+\frac{1}{N_2-3}}} = \frac{.829 - .563}{\sqrt{\frac{1}{14}+\frac{1}{25}}}$$
$$= 0.797$$

The correlations are not significantly different from each other.

b. We do not have reason to argue that the relationship between performance and prior test scores is affected by whether or not the student read the passage.

9.24 Difference in correlation between Katz' two groups

$r_1 = .88 \quad r'_1 = 1.376$

$r_2 = .72 \quad r'_2 = .908$

$$z = \frac{r_1' - r_2'}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} = \frac{1.376 - .908}{\sqrt{\frac{1}{49} + \frac{1}{71}}}$$
$$= 2.52$$

The difference is significant.

9.25 It is difficult to tell whether the significant difference between the results of the two previous problems is to be attributable to the larger sample sizes or the higher (and thus more different) values of r'. It is likely to be the former.

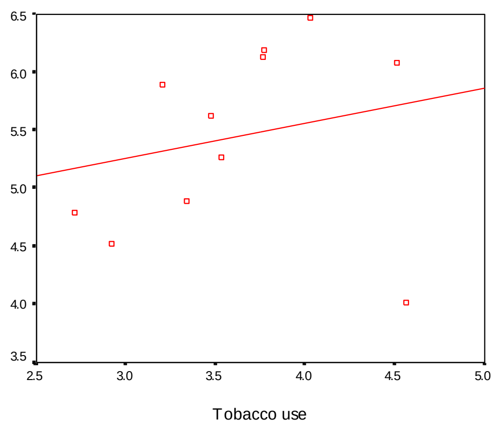
9.26 No one answer would be relevant here.

9.27 Moore and McCabe example of alcohol and tobacco use:

Correlations

		ALCOHOL	TOBACCO
ALCOHOL	Pearson Correlation	1.000	.224
	Sig. (2-tailed)	.	.509
	N	11	11
TOBACCO	Pearson Correlation	.224	1.000
	Sig. (2-tailed)	.509	.
	N	11	11

b. The data suggest that people from Northern Ireland actually drink relatively little.



c. With Northern Ireland excluded from the data the correlation is .784, which is significant at $p = .007$.

9.28 Relationship between GSIT and GPA in Mireault.dat:

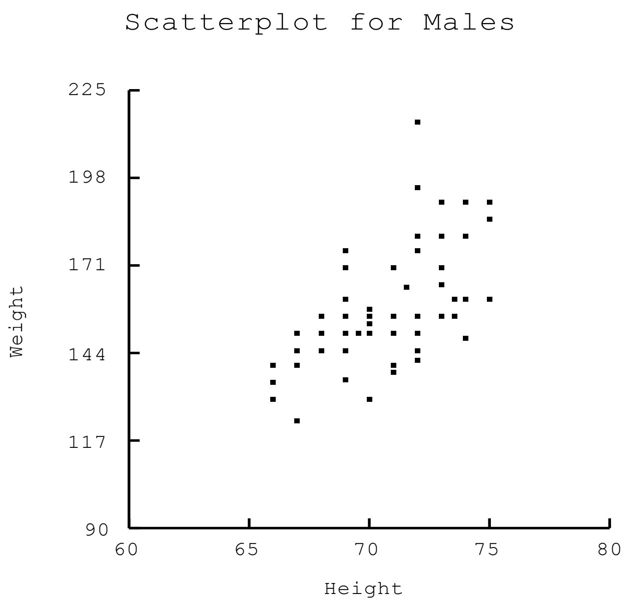
$r = .086$ $F(1,361) = 2.66$; Not significant

9.29 **a.** The correlations range between .40 and .80.

b. The subscales are not measuring independent aspects of psychological well-being.

9.30 Computer problem

9.31 Relationship between height and weight for males:



The regression solution that follows was produced by SPSS and gives all relevant results.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.604[a]	.364	.353	14.9917

a. Predictors: (Constant), HEIGHT

b. Gender = Male

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7087.800	1	7087.800	31.536	.000[a]
	Residual	12361.253	55	224.750		
	Total	19449.053	56			

a. Predictors: (Constant), HEIGHT

b. Dependent Variable: WEIGHT

c. Gender = Male

Coefficients[a,b]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-149.934	54.917		-2.730	.008
	HEIGHT	4.356	.776	.604	5.616	.000

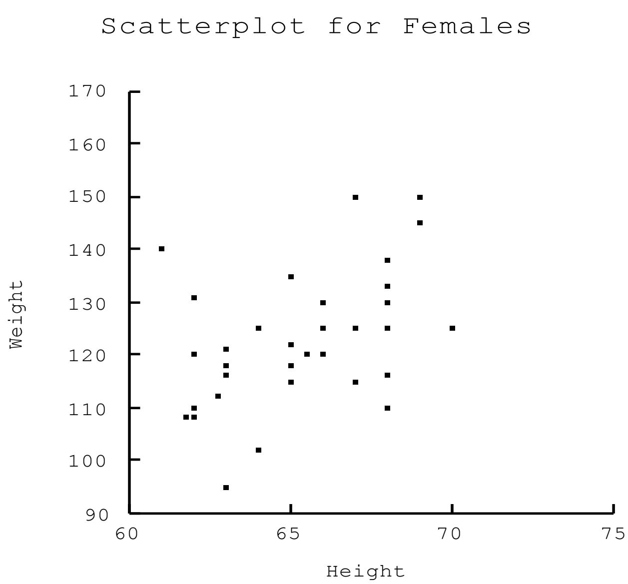
a. Dependent Variable: WEIGHT

b. Gender = Male

With a slope of 4.36, the data predict that two males who differ by one inch will also differ by approximately 4 1/3 pounds. The intercept has no meaning because people are not 0 inches tall, but the fact that it is so largely negative suggests that there is some curvilinearity in this relationship for low values of Height.

Tests on the correlation and the slope are equivalent tests when we have one predictor, and these tests tell us that both are significant. Weight increases reliably with increases in height.

9.32 Relationship between height and weight for females:



The regression solution that follows was produced by SPSS and gives all relevant results.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.494[a]	.244	.221	11.7997

a. Predictors: (Constant), HEIGHT

b. Gender = Female

ANOVA[b,c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1484.921	1	1484.921	10.665	.003[a]
	Residual	4594.679	33	139.233		
	Total	6079.600	34			

a. Predictors: (Constant), HEIGHT

b. Dependent Variable: WEIGHT

c. Gender = Female

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-44.859	51.684		-.868	.392
	HEIGHT	2.579	.790	.494	3.266	.003

a. Dependent Variable: WEIGHT

b. Gender = Female

9.33 As a 5’8” male, my predicted weight is $\hat{Y}$ = 4.356(Height) - 149.934 = 4.356*68 - 149.934 = 146.27 pounds.

a. I weigh 146 pounds. (Well, I did two years ago.) Therefore the residual in the prediction is $Y-\hat{Y}$ = 146 - 146.27 = -0.27.

b. If the students on which this equation is based under- or over-estimated their own height or weight, the prediction for my weight will be based on invalid data and will be systematically in error.

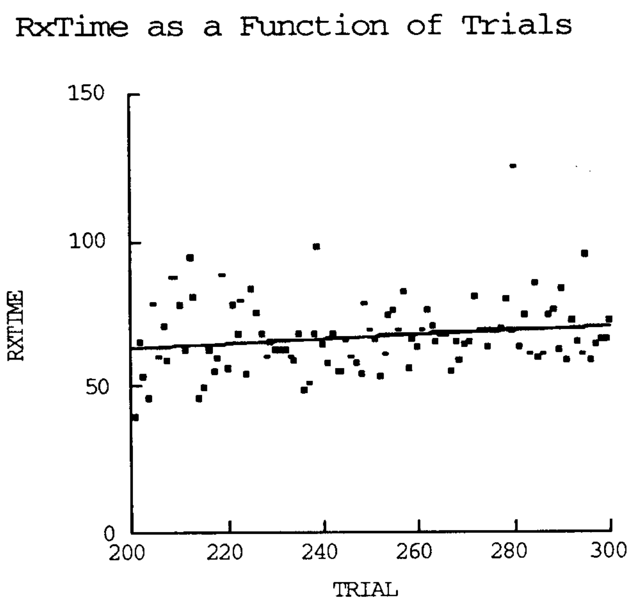
9.34 The largest residual for males is 51.311 points. This person was 6 feet tall and weighed 215 pounds. His predicted weight was only 163.7 pounds.

9.35 The male would be predicted to weigh 137.562 pounds, while the female would be predicted to weigh 125.354 pounds. The predicted difference between them would be 12.712 pounds.

9.36 Males are denser. By this I mean that a male weighs more per inch than does a female.

9.37 Independence of trials in reaction time study.

The data were plotted by "trial", where a larger trial number represents an observation later in the sequence.



Although the regression line has a slight positive slope, the slope is not significantly different from zero. This is shown below.

```
DEP VAR:   TRIAL      N:      100  MULTIPLE R: 0.181  SQUARED MULTIPLE R: 0.033
ADJUSTED SQUARED MULTIPLE R: 0.023    STANDARD ERROR OF ESTIMATE:     28.67506

VARIABLE      COEFFICIENT    STD ERROR     STD COEF TOLERANCE    T    P(2 TAIL)

CONSTANT        221.84259     15.94843      0.00000    .        .14E+02  .10E-14
RXTIME            0.42862      0.23465      0.18146  1.00000  1.82665  0.07080

                       ANALYSIS OF VARIANCE
SOURCE       SUM-OF-SQUARES   DF  MEAN-SQUARE     F-RATIO       P

REGRESSION      2743.58452     1   2743.58452     3.33664     0.07080
RESIDUAL       80581.41548    98    822.25934
```

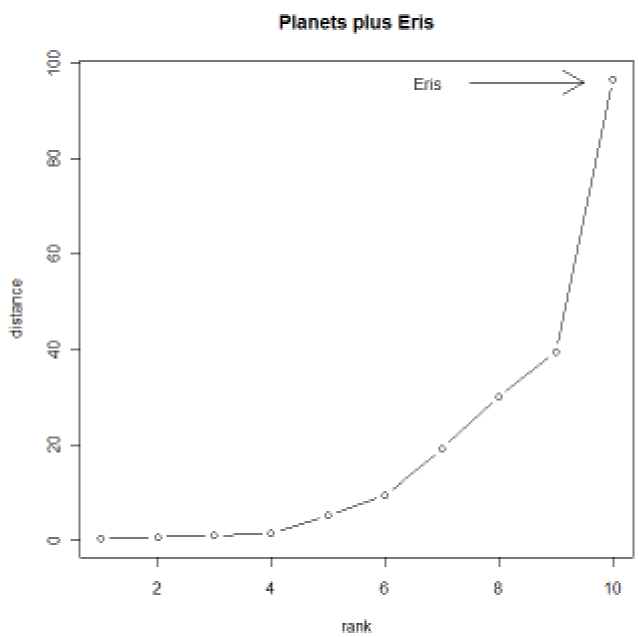
There is not a systematic linear or cyclical trend over time, and we would probably be safe in assuming that the observations can be treated as if they were independent. Any slight dependency would not alter our results to a meaningful degree.

9.38 Air quality measures.

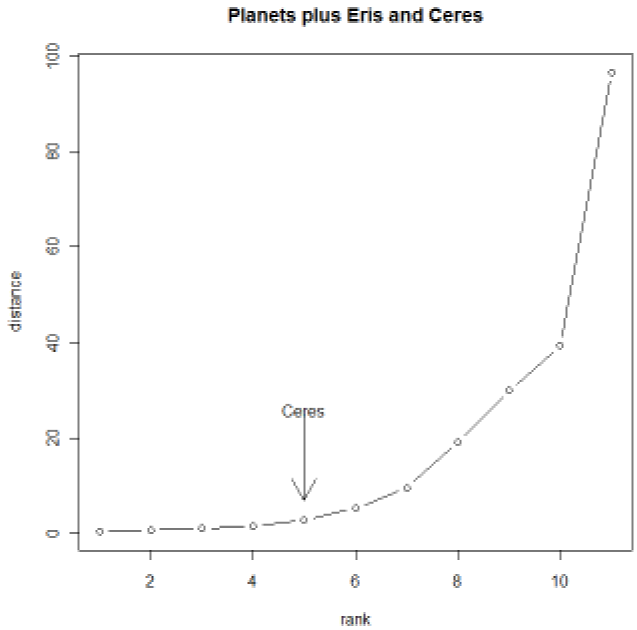
In these data (found as Ex9-38.dat) I wanted students to see that there are many ways of looking at a relationship between variables. Comparing the means would tell us only that one instrument read higher than the other, it wouldn't get at whether they are measuring the same thing. The data are somewhat curvilinear, and we need to take that into account. I put together a fairly extensive set of lecture notes on this example, and they can be found at http://www.uvm.edu/~dhowell/gradstat/psych340/Lectures/Class3.html. The notes develop a large number of simple ideas out of this one example.

9.39 What about Eris?

Eris doesn't fit the plot as well as I would have liked. It is a bit too far away.



9.40 What about Ceres? Here we have a good fit—in fact an even better fit



.

9.41 Comparing correlations in males and females.

$$z = \frac{r_1' - r_2'}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

$$= \frac{.648 - .343}{\sqrt{\frac{1}{284} + \frac{1}{222}}} = \frac{.305}{\sqrt{0.0085}} = \frac{.305}{.092}$$

$$= 3.30$$

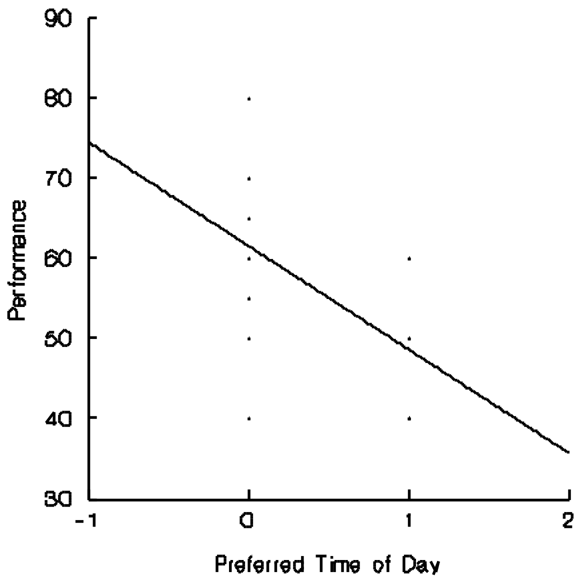
The difference between the two correlations is significant.

9.42 This is an Internet search question with no fixed answer.

Chapter 10 - Alternative Correlational Techniques

10.1 Performance ratings in the morning related to perceived peak time to day:

a. Plot of data with regression line:



b.

$$s_X = 0.489$$

$$s_Y = 11.743$$

$$\text{cov}_{XY} = -3.105$$

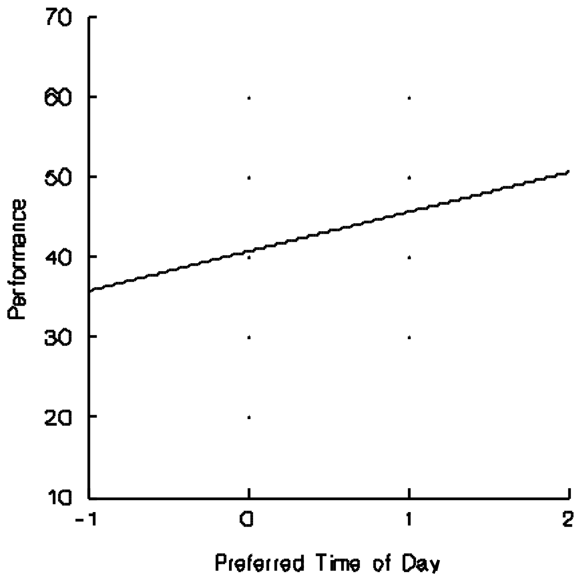
$$r_{pb} = \frac{\text{cov}_{XY}}{s_X s_Y} = \frac{-3.105}{(0.489)(11.743)} = -.540$$

$$t = \frac{r\sqrt{(N-2)}}{\sqrt{1-r^2}} = \frac{(-.540)\sqrt{18}}{\sqrt{.708}} = \frac{-2.291}{.842} = -2.723 \quad [p < .01]$$

c. Performance in the morning is significantly related to people's perceptions of their peak periods.

10.2 Performance ratings in the evening related to perceived peak time of day:

a. Plot of data with regression line:



b.

$$s_X = 0.489$$
$$s_Y = 10.699$$
$$\text{cov}_{XY} = 1.184$$
$$r_{pb} = \frac{\text{cov}_{XY}}{s_X s_Y} = \frac{1.184}{(0.489)(10.699)} = .226$$
$$t = \frac{r\sqrt{(N-2)}}{\sqrt{1-r^2}} = \frac{(.226)\sqrt{18}}{\sqrt{.949}} = \frac{.959}{.974} = .985 \quad \text{[not significant]}$$

c. Performance in the evening is not significantly related to perceived peak periods.

10.3 It looks as if morning people vary their performance across time, but that evening people are uniformly poor.

10.4 We believe that the underlying distribution is bimodal, and not continuous.

10.5 Running a t test on the data in Exercise 10.1:

$\bar{X}_1 = 61.538$	$s_1^2 = 114.103$	$n_1 = 13$
$\bar{X}_2 = 48.571$	$s_2^2 = 80.952$	$n_2 = 7$

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(13-1)114.103+(7-1)80.952}{13+7-2} = 103.053$$

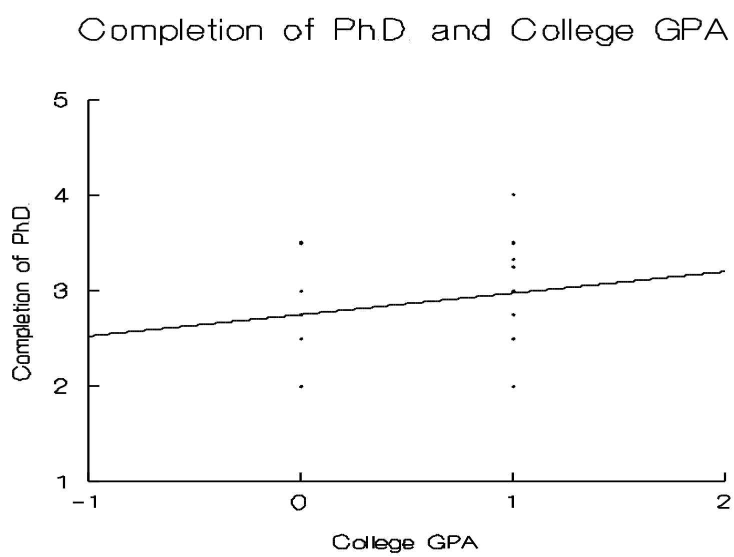
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{61.538-48.571}{\sqrt{103.053\left(\frac{1}{13}+\frac{1}{7}\right)}} = 2.725$$

$[t_{.025(18)} = \pm 2.101]$ Reject H_0

The t calculated here (2.725) is equal to the t calculated to test the significance of the r calculated in Exercise 10.1.

10.6 Relationship between college GPA and completion of Ph.D. program:

a. Plot of data with regression line:



b.

$s_X = 0.503$

$s_Y = 0.476$

$cov_{XY} = 0.051$

$$r_{pb} = \frac{cov_{XY}}{s_X s_Y} = \frac{0.051}{(0.503)(0.476)} = .213$$

c.

$$r_b = \frac{r_{pb}\sqrt{p_1 p_2}}{y} = \frac{.213\sqrt{(.32)(.68)}}{.358} = .278$$

d. Yes, it is reasonable to consider r_b because there really is a continuum of College Grade Point average, and the distribution is roughly normal.

10.7 Regression equation for relationship between college GPA and completion of Ph.D. program:

$$b = \frac{\text{cov}_{XY}}{s_X^2} = \frac{0.051}{.503^2} = .202$$

$$a = \frac{\Sigma Y - b\Sigma X}{N} = \frac{17 - .202(72.58)}{25} = .093$$

$$\hat{Y} = bX + a = .202X + .093$$

When $X = \overline{X} = 2.9032$, $\hat{Y} = .202(2.9032) + .093 = .680 = \overline{Y}$.

10.8 They represent nothing meaningful because (1) the values (0,1) for Ph.D. are arbitrary, and (2) no one would be admitted to graduate school with a GPA even approaching 0.00.

10.9 Establishment of a GPA cutoff of 3.00:

a.

Ph.D. (Y):	0	0	0	0	0	0	0	0	1	1	
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1									
GPA (X):	0	1	0	1	1	0	0	0	1	0	
1	1	1	1	0	1	1	1	0	0	0	1
1	1	0									

b.

$$s_X = 0.507$$

$$s_Y = 0.476$$

$$\text{cov}_{XY} = 0.062$$

$$\phi = \frac{0.062}{(0.507)(0.476)} = .256$$

c.

$$t = \frac{r\sqrt{(N-2)}}{\sqrt{1-r^2}} = \frac{(.256)\sqrt{23}}{\sqrt{.934}} = \frac{1.228}{.967} = 1.27 \quad \text{[not significant]}$$

10.10 Exercise 10.9 as a contingency table:

		Completed Ph.D. 0	1	
GPA > 3	No	5 (3.52)	6 (7.48)	11
	Yes	3 (4.48)	11 (9.52)	14
		8	17	25 = *N*

a.

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

$$= \frac{(5-3.52)^2}{3.52} + \frac{(6-7.48)^2}{7.48} + \frac{(3-4.48)^2}{4.48} + \frac{(11-9.52)^2}{9.52}$$

$$= .6223 + .2928 + .4889 + .2301$$

$$= 1.6341 \quad [\chi^2_{.05(1)} = 3.84]$$

b.

$$\phi = \sqrt{\chi^2 / N}$$

$$.256 = \sqrt{1.6341/25}$$

$$.256 = .256$$

10.11 Alcoholism and childhood history of ADD:

a.

$$s_X = 0.471$$
$$s_Y = 0.457$$
$$\text{cov}_{XY} = 0.135$$

$$\phi = \frac{0.135}{(0.471)(0.457)} = .628$$

b. $\chi^2 = N\phi^2 = 32\left(.628^2\right) = 12.62 \quad [p < .05]$

10.12 Development ordering of language skills using Spearman's r_S :

a.

$$s_X = 4.472$$

$$s_Y = 4.472$$

$$cov_{XY} = 19.429$$

$$r_s = \frac{19.429}{(4.472)(4.472)} = .972$$

b. The correlation between the two judges is very high, indicating substantial agreement about the order of the skills.

10.13 Development ordering of language skills using Kendall's τ

a.

$$\tau = 1 - \frac{2(\#\text{ inversions})}{\#\text{ pairs}} = 1 - \frac{2(6)}{15(14)/2} = 1 - \frac{23}{105} = .886$$

b.

$$z = \frac{\tau}{\sqrt{\frac{2(2N+5)}{9N(N-1)}}} = \frac{.886}{\sqrt{\frac{2(30+5)}{9(15)(14)}}} = \frac{.886}{\sqrt{.037}} = 4.60 \quad [p < .05]$$

10.14 Ranking of videotapes of children's behaviors by clinical graduate students and experienced clinicians using Spearman's *r*:

$$s_X = 3.028$$

$$s_Y = 3.028$$

$$cov_{XY} = 8.1667$$

$$r_s = \frac{8.1667}{(3.028)(3.028)} = .891$$

10.15 Ranking of videotapes of children's behaviors by clinical graduate students and experienced clinicians using Kendall's τ:

Experienced	**New**	**Inversions**
1	2	1
2	1	0
3	4	1
4	3	0
5	5	0
6	8	2

7	6	0
8	10	2
9	7	0
10	9	0

$$\tau = 1 - \frac{2(\#\text{ inversions})}{\#\text{ pairs}} = 1 = \frac{2(6)}{10(9)/2} = 1 - \frac{12}{45} = .733$$

10.16 Ranking of videotapes of children's behaviors by clinical graduate students and experienced clinicians using Kendall's W and $\bar{r}_S$

Column totals: (T_j): 10 22 8 28 26 13 46 43 34 45
$K = 5$
$N = 10$

$$W = \frac{12\Sigma T_j^2}{K^2N(N^2-1)} - \frac{3(N+1)}{N-1}$$

$$= \frac{12(9423)}{5^2(10)(99)} - \frac{3(11)}{9} = \frac{113076}{24750} - \frac{33}{9} = 4.459 - 3.667 = .902$$

$$\bar{r}_S = \frac{KW-1}{K-1} = \frac{5(.902)-1}{4} = .878 = .88$$

The average pairwise correlation among judges' rankings = 0.88.

10.17 Verification of Rosenthal and Rubin's statement

	Improvement	No Improvement	Total
Therapy	66 (50)	34 (50)	100
No Therapy	34 (50)	66 (50)	100
Total	100	100	200

a.

$$\chi^2 = \Sigma\frac{(O-E)^2}{E} = \frac{(66-50)^2}{50} + \frac{(34-50)^2}{50} + \frac{(34-50)^2}{50} + \frac{(66-50)^2}{50}$$
$$= 20.48$$

b. An $r^2 = .0512$ would correspond to $\chi^2 = 10.24$. The closest you can come to this result is if the subjects were split 61/39 in the first condition and 39/61 in the second (rounding to integers.)

10.18 Point-biserial correlation from Mireault's (1990) data.

Correlation between Gender and DepressT

$r_{pb} = -.1746$ [$p = .0007$]

10.19 ClinCase against Group in Mireault's data

	ClinCase 0	ClinCase 1
Loss	69	66
Married	108	73
Divorced	36	23

a. $\chi^2 = 2.815$ [$p = .245$]
$\phi_C = .087$

c. This approach would be preferred over the approach used in Chapter 7 if you had reason to believe that differences in depression scores below the clinical cutoff were of no importance and should be ignored.

10.20 ClinCase against Gender in Mireault's data

		ClinCase 0	ClinCase 1
Gender	Male	65	75
	Female	148	87

a. $\chi^2 = 9.793$ [p = .002]
$\phi_C = .162$

b. The answer to this exercise and exercise 10.17 are very close. Both techniques are addressing the same question except that here we have dichotomized the depression score.

10.21 Small Effects:

a. If a statistic is not significant, that means that we have no reason to believe that it is reliably different from 0 (or whatever the parameter under H_0). In the case of a correlation, if it is not significant, that means that we have no reason to believe that there is a relationship between the two variables. Therefore it cannot be important.

b. With the exceptions of issues of power, sample size will not make an effect more important than it is. Increasing N will increase our level of significance, but the magnitude of the effect will be unaffected.