variable(s) *caused* changes in the dependent variable. To causally link variables we look beyond a statistical model to identify and eliminate other possible causes of an outcome. We discuss this in Chapter 11.

CONCLUDING OBSERVATIONS

In deciding how to analyze and present data you need to consider your research question, how your data are measured, and your audience. Contingency tables are appropriate for nominal and ordinal data. They also have the advantage of being easily understood. We have found that inexperienced researchers ignore the power of percentages: a clear, uncluttered table reporting relative frequencies is usually easier to interpret than one that only reports the raw numbers. After many years of teaching, we have learned that not all people interpret contingency tables—even ones with relative frequencies—accurately. If you use contingency tables in your report, it's wise to include a sentence summarizing the findings.

Regression, especially multiple regression, is a powerful tool that can handle ratio variables efficiently. It has the distinct advantage of being able to include a number of independent variables in the same equation. Probably the greatest potential misuse of regression is to use it to analyze a small data set, where the addition of a few cases can greatly alter the findings.

RECOMMENDED RESOURCES

Arlene, Fink, *How To Conduct Surveys: A Step By Step Guide*. Fourth edition (Thousand Oaks, CA: Sage Publications, Inc., 2009). See especially Chapter 6.

Lee, E. S., *Analyzing Complex Survey Data*, Second Edition (Thousand Oaks, CA: Sage Publications, Inc., 2006).

O'Sullivan, E., G. Rassel, and M. Berner, *Research Methods for Public Administrators*, Fifth Edition (New York: Pearson/Longman, 2008). Chapter 13.

Rea, Louise, and Richard Parker, *Conducting and Designing Survey Research: A Comprehensive Guide*, Third Edition (San Francisco, CA: Jossey Bass, 2005).

Tufte, Edward R., *Data Analysis for Politics and Policy* (Englewood Cliffs, NJ: Prentice Hall, 1974) has a clear discussion with excellent examples on the use and interpretation of linear correlation and regression.

CHAPTER 8 EXERCISES

Analyzing Data to Find Relationships Exercises

There are four sets of exercises for this chapter.

- Exercise 8.1 Nonprofit Participation in Experimental Financial Assistance Program is designed to give you practice in creating and interpreting contingency tables.
- Exercise 8.2 Do Public Service Announcements Yield More Volunteers? asks you to apply your knowledge of linear regression.
- Exercise 8.3 Physicians for Access also asks you to apply your knowledge of linear regression.
- Exercise 8.4 Fresh Start Center revisits the data presented in Table 5.8, and asks you to suggest strategies for identifying relationships.

EXERCISE 8.1 Nonprofit Participation in Experimental Financial Assistance Program

Scenario

One hundred (100) statewide nonprofits are asked to participate in an experimental program to deliver financial assistance. The local board of directors of each agency decides whether the organization will participate in the program. Table 8.4 contains the data on each nonprofit's region and whether it will participate in the program.

TABLE 8.4

Database on Participation in Experimental Program

Agency Number	Region of State	Experimental Program	Agency Number	Region of State	Experimental Program
1	Plain	No	51	Mountain	No
2	Plain	No	52	Mountain	No
3	Mountain	No	53	Coast	No
4	Plain	No	54	Mountain	Yes
5	Mountain	Yes	55	Plain	Yes
6	Mountain	No	56	Mountain	No
7	Coast	No	57	Plain	No
8	Coast	No	58	Plain	Yes
9	Coast	No	59	Coast	No
10	Coast	No	60	Coast	Yes
11	Coast	Yes	61	Coast	No
12	Mountain	No	62	Coast	No
13	Plain	Yes	63	Plain	Yes
14	Mountain	Yes	64	Coast	No
15	Coast	No	65	Coast	No
16	Coast	No	66	Coast	Yes
17	Plain	No	67	Coast	No
18	Coast	No	68	Plain	Yes
19	Plain	Yes	69	Coast	Yes
20	Mountain	No	70	Mountain	No
21	Coast	No	71	Plain	No
22	Mountain	No	72	Plain	No
23	Coast	No	73	Coast	Yes
24	Coast	No	74	Plain	No
25	Coast	Yes	75	Plain	No
26	Coast	No	76	Coast	No
27	Coast	No	77	Coast	No
28	Coast	No	78	Coast	No
29	Plain	No	79	Plain	No

TABLE 8.4

Database on Participation in Experimental Program

Agency Number	Region of State	Experimental Program	Agency Number	Region of State	Experimental Program
30	Coast	No	80	Plain	No
31	Coast	No	81	Plain	No
32	Plain	Yes	82	Mountain	No
33	Coast	No	83	Mountain	Yes
34	Coast	Yes	84	Coast	No
35	Plain	No	85	Plain	No
36	Coast	No	86	Plain	No
37	Coast	No	87	Plain	Yes
38	Mountain	No	88	Plain	No
39	Plain	No	89	Coast	No
40	Coast	No	90	Mountain	No
41	Plain	Yes	91	Coast	No
42	Coast	No	92	Mountain	Yes
43	Coast	No	93	Coast	No
44	Mountain	No	94	Plain	No
45	Mountain	No	95	Mountain	No
46	Coast	No	96	Coast	Yes
47	Coast	No	97	Coast	No
48	Coast	No	98	Plain	No
49	Plain	No	99	Coast	No
50	Mountain	Yes	100	Coast	No

Section A: Getting Started

1. Organize the data in a contingency table with three columns and two rows. Make Region of State the independent variable and Experimental Welfare Program the dependent variable.
   a. Tally the number of agencies in each region that are in the program and the number that are not. Report the number in each column. Fully label the table.
   b. In a second table, calculate the percentage of organizations in each region that are in the program and the percentage that are not. Enter only the percentages in the cells of the second table with the number of cases in each column at the head of the column. Fully label the table.
   c. What can you tell the state director about the relationship between the decision to participate in the Experimental Program and Region of the State?
2. What percentage of the agencies are in each region of the state?
3. What percentage of the agencies will participate in the experimental program? What percentage will not? What percentage of the agencies in the Plains region are participating?

4. Calculate and report the percentage difference between the percentage of agencies in the Mountains and the percentage in the Plains that are participating in the experimental program.

EXERCISE 8.2 Do Public Service Announcements Yield More Volunteers?

Scenario

A radio station runs public service announcements for an organization that has volunteers build houses for low-income families. The database representing the number of volunteers and the number of public service announcements is in Table 8.5.

Section A: Getting Started

1. Prepare a scatterplot to show the relationship between the two variables. Remember that "number of volunteers" is the dependent variable. Describe the direction of the relationship. Is it linear?
2. Compare your scatterplot for public service announcements to the scatterplot in Figure 8.1. How are they similar? How do they differ? Which relationship do you think is the stronger? Why?
3. The regression and correlation statistics for the relationship between number of volunteers and public service announcements are $a = 63.9$; $b = 8.5$; $r = 0.74$.
   a. Use this information to draw a regression line in your scatterplot.
   b. Use this information to write the regression equation and estimate the number of volunteers expected if the number of public service announcements is 4.

TABLE 8.5

Database on Number of Volunteers and Public Service Announcements

Week	Number of Volunteers	Public Service Announcements
1	130	7
2	98	2
3	123	6
4	138	10
5	132	8
6	150	9
7	83	5
8	115	6
9	76	5
10	149	7

Section A: Getting Started

1. The following data were reported in answer to the question "How big a problem is racism in our society today?"

   Of all respondents: 26% a "big problem," 22% a "small problem"

   Of African American respondents: 44% a "big problem," 11% a "small problem"

   Of White respondents: 22% a "big problem," 23% a "small problem"

   a. Use the reported statistics, for example, 26%, to estimate p and $1-p$. Compute the sampling error for the entire sample, the African American sample, and the White sample. (Assume that the number of Whites is the same as the number of non–African Americans.)
   b. Use the sampling error and report the confidence interval for the percentage of African Americans and Whites who believed that racism was a big problem.
   c. What is the probability that your estimates in 1b are wrong? How did you arrive at this estimate?
2. Assume maximum variability (50-50 split)
   a. Compute the sampling errors for the entire sample and for the 204-member African American sample.
   b. As a general practice, would you analyze survey data using a 50-50 split or would you use the statistical analysis (question 1) to come up with more precise estimates? Justify your answer.
3. The survey also reported that among African American respondents 60 percent had personally felt that "a shopkeeper or sales clerk was trying to make" them feel unwelcome. You are curious if the same thing is true in your community. In trying to decide a value of p would you use 0.60, 0.50, or something else? Justify your answer.
4. The community action group considers replicating parts of a survey. What size sample is needed to have 2 percent, 5 percent, or 10 percent accuracy? (Note that accuracy is the same as sampling error.)

EXERCISE 9.2 Attitudes toward Corporal Punishment: Are Men and Women Different?

Scenario

A child care organization commissioned a random survey to identify attitudes toward child-raising. A topic of interest was the difference between the beliefs of men and women regarding discipline. The follow table reports data on men's and women's attitudes toward spanking.

Attitudes toward Spanking by Respondent Gender

Favor Spanking	Male	Female
Strongly	115	107
Somewhat	212	221
Oppose	73	109
Strongly oppose	18	42

Chi-square = 13.1, degrees of freedom = 3, significance = 0.004.

Section A: Getting Started

1. State the alternative hypothesis and the null hypothesis that could be tested with the data in this table.
2. Identify the independent and the dependent variables.
3. Calculate percentages and include in them in a table. Write a sentence to describe the relationship shown in the table. Do the data in the table support or contradict your hypothesis? Explain.
4. Based on the chi-square evidence what would you do, that is, would you reject the null hypothesis?

Section B: Small Group Exercise

1. What are the implications of the findings presented in exercise 9.2? Do you consider the table an interesting observation, a question for further study, or something else?
2. A finding of statistical significance can be persuasive. What other evidence should the child care organization present so that parents and other stakeholders are able to evaluate the findings?

EXERCISE 9.3 What Is Going On in the Schools?

Scenario

A community action group has heard complaints that African American and Hispanic students are more likely to be suspended (either in-school or out-of-school) than other students. The superintendent of schools offers to review the files of students in grades 9–12. The school system has 39,000 students in grades 9–12: 58 percent African American, 12 percent Hispanic, and 25 percent White.

Section A: Getting Started

1. What target population would you recommend the superintendent use? Why did you recommend this population? (Note that target population refers to the specific population that the data will represent.)
2. State the alternative hypothesis and the null hypothesis the superintendent should test.
3. If the superintendent tests the hypothesis and makes a Type I error, explain what has happened.

4. If the superintendent tests the hypothesis and makes a Type II error, explain what has happened.
5. Should the superintendent be more concerned about a Type I error or a Type II error? Justify your answer.
6. The superintendent originally set $\alpha = 0.05$. How can she further decrease the probability of a Type I error? How can she further decrease the probability of a Type II error?

Section B: Small Group Exercise

1. Decide on a target population and suggest a sample size for the superintendent's study. Justify your recommendation.
. Discuss what actions the superintendent might take if the null hypothesis is rejected.
3. List arguments for the position
   a. Committing a Type I error is the more serious concern.
   b. Committing a Type II error is the more serious concern.
4. Based on what you have observed in Exercises 9.2 and 9.3 draft a memo "What you want to know about statistical significance: A guide for citizens."

EXERCISE 9.4 How Groups Work Together

The following table was created as part of a study of collaborations formed around women's issues and environmental issues. Members answered a series of questions to see if the coalitions were different. Each question was answered along a scale ranging from 1 = Not at all true to 7 = To a great extent true.

Indicator	Women's Issues Mean(s)	Environmental Issues Mean(s)	*t*-value
• I understand my organization's roles and responsibilities	6.6 (.7)	5.9 (1.2)	1.7
• Partners agree about goals	6.6 (.5)	5.3 (0.8)	4.6***
• Meetings accomplish what is necessary for the collaboration to function	6.9 (.3)	6.1 (0.8)	3.1**
• Tasks are well coordinated	6.4 (.8)	4.8 (1.5)	3.1**
• Collaboration hinders my organization from achieving its mission	1.6 (1.2)	2.1 (1.5)	.9
• Collaboration affects my organization's independence	2.1 (1.5)	2.6 (1.6)	.8
• I feel pulled between meeting my organization's expectations and the collaboration's expectations	2.5 (1.9)	4.0 (1.5)	2.0**
• Partner organizations have combined and used resources so that all partners benefit	6.8 (.4)	5.5 (1.0)	4.0***
• My organization shares information that will strengthen the partners' organization	6.7 (.5)	6.1 (0.7)	2.3*

Indicator	Women's Issues Mean(s)	Environmental Issues Mean(s)	*t*-value
• Partner organizations appreciate and respect what my organization contributes	6.7 (.5)	5.7 (1.)	3.0**
• My organization achieves its goals better through collaboration	6.7 (.6)	5.9 (1.4)	1.7
• My organization believes it is better to stay and work with partners rather than leave	6.9 (0.3)	6.0 (1.0)	2.9**
• Representatives of partner organizations are trustworthy	6.9 (0.3)	5.4 (1.0)	4.8***
• My organization can count on each partner to meet its obligations	6.5 (0.5)	4.8 (0.8)	6.0***

Two-tailed *t*-test: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Section A: Getting Started

1. In plain English explain what information the table contains.
2. In plain English interpret the statistical information for the last line ("My organization can count on each partner to meet its obligations.)
3. How do collaborations focused on women's issues differ from collaborations focused on environmental issues? What criteria did you use to make your choices?