


reactions. Again, the author stated that subjective truth is more important. The interview questions were loosely presented, and the place (aside from an outdoor cafe) was not specified, both of which could affect the nature of the narratives. Moreover, the interviewer was a female, and the male participant may have felt uneasy about reporting his homosexuality to a female. Most important, interpretations of this narrative report are influenced by the theoretical leanings of the interpreter. It may be that this particular case supported her views. And, because a complete transcript was not presented, the particular narratives might have been selected because they supported the author's views.

SUGGESTED READINGS

- Gubrium, J. F., & Holstein, J. A. (2009). *Analyzing narrative reality*. Thousand Oaks, CA: Sage.
- Herman, L., & Vervaeck, B. (2005). *Handbook of narrative analysis*. Lincoln: University of Nebraska Press.
- Huberman, A. M., & Miles, M. B. (2002). *The qualitative researcher's companion*. Thousand Oaks, CA: Sage.
- Josselson, R., & Lieblich, A. (1993). *The narrative study of lives* (Vol. 1). Newbury Park, CA: Sage.
- Josselson, R., & Lieblich, A. (1995). *Interpreting experience: The narrative study of lives* (Vol. 3). Thousand Oaks, CA: Sage.
- McMillan, J. H., & Schumacher, S. (2000). *Research in education: A conceptual introduction* (5th ed.). White Plains, NY: Longman.
- Riessman, C. K. (2007). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.
- Webster, L., & Mertova, P. (2007). *Using narrative inquiry as a research method: An introduction to using critical event narrative analysis in research on learning and teaching*. New York: Routledge.

Chapter 4

Surveys



The case study is the cruelest of the descriptive methods for studying behavior. The focus is on a single individual or group of individuals studied as a unit. As such, if results can be considered internally valid, they can be generalized only to other similar individuals with like characteristics and backgrounds who would be measured under similar situations. In brief, case studies and most analytic studies lack external validity: Their results cannot be generalized to the population of which the participant is a sample. On the other hand, **surveys** are conducted with the specific intent of generalizing the results—almost always quantitative—to the population of interest. (Populations are always defined in terms of a target characteristic that is shared by all of its members.)

Surveys can be better controlled than case studies. They obtain information that is otherwise inaccessible. On occasion,

surveys are conducted with the intent of obtaining a general sense of what people feel. Politicians, for example, may want to assess what people feel are important election issues. So, generalizations in terms of estimations of population means, standard deviations (SDs), and the like are not considered. There are two broad types of surveys.

Attitude surveys measure likes, dislikes, and so on. *Research surveys* test hypotheses. With both, responses are related to demographic information.

Two crucial aspects of surveys are the development of a valid and reliable questionnaire and selection of the sample. Because the intent is to generalize the results to the population, the sample has to be representative. **Probability samples** are representative: There is a certain probability that every member can be included in the sample. The process of sample selection is the key to valid results. Sample selection depends on the **sample frame**, the population that has

a chance to be selected. It can be an entire list of members, a list that is created as its members attend something (a clinic, a hospital, a meeting, etc.), or a list of members of units (such as houses). It should be comprehensive (all members have a chance of being selected) and exclude few. It should provide a determination of the probability of selecting each member. Also, it should be efficient (include only those members of interest). Researchers should note the criteria of inclusion and exclusion of members of the sample frame.

The sampling procedure can result in a random or a **stratified sample**. With the latter, the population is segmented on some basis (e.g., age, college level), and samples of each segment are in the same ratio as exists in the population. In either case of probability sampling, it is possible to calculate estimations of population values, usually means, SDs, and standard errors (SEs) of the mean. Be leery of surveys that report these values for nonprobability samples.

Nonprobability samples are not representative, because they unintentionally exclude members of the population. For example, telephone directories exclude people without telephones, with unlisted numbers, or with phones installed after the directory was published. **Quota samples** are nonrepresentative samples with certain characteristics in the same ratio that exists in the population but only from a particular location, and the participants are selected at the discretion of the interviewer. The most accessible may be selected and will have characteristics that

administration, however, is likely to yield almost 100% returns. But all members of the group may not be present at administration time. It is equally important to know who gathered the data, how the data were gathered, and where the data were gathered.

Results of surveys are often reported in terms of percentages. In such cases, tests of significance typically involve the chi-square distribution. **Chi-square tests** determine whether frequencies in certain categories differ from what would be expected on the basis of chance. When categories differ along a single dimension, the test is called **goodness of fit**. For example, all faces of an unbiased die should occur equally often when it is tossed many times. If you toss it 120 times, you would expect each face to appear 20 times ($120/6 = 20$). Suppose you obtain frequencies of 18, 20, 22, 30, 16, and 14 for faces 1 through 6, respectively. Is there evidence to suggest a biased die? The $\chi^2 = 8.00$, and you enter the chi-square table with $6 - 1 = 5$ *df*. A chi-square of at least 11.07 can occur by chance 5 times out of 100 ($p = .05$). A value of 8.00 can be expected more often, and you conclude that the obtained frequencies do not differ reasonably from the expected frequencies of 20. There is not sufficient evidence to consider the die biased.

Contingency tests determine whether frequencies in one category depend on membership in another. For example, does preference for rare, medium, or well-done steak depend on

gender? Males and females are surveyed about their preference. You end up with a 3 (steak preference) $\times 2$ (gender) table of frequencies. Expected frequencies are based on marginal totals across rows and columns for each cell ($[\text{row total} \times \text{column total}] / N$). Then, chi-square is calculated and compared with a tabled value for $df = (\text{rows} - 1)(\text{columns} - 1)$. Here, $df = (3 - 1)(2 - 1) = 2$. If the obtained value is lower than the tabled χ^2 , you conclude that steak preference does not depend on gender. If the obtained value is significant, then steak preference depends on gender. Then, further analysis is required to pinpoint the source of the significance of chi-square. There are, however, some requirements for using the chi-square test. The data are categorical, the basis for frequencies. The sample size should be sufficiently large that you end up with an expected frequency of at least 10 in a 2×2 table or 5 in a larger table. Most important, every entry is independent of every other entry. That means that frequencies refer to unique individuals (i.e., you can't have two entries for one individual in two categories).

To avoid threats to internal validity, the questionnaire should be reliable and valid for the target population; any measuring instrument, or scale, used also should be reliable and valid; testing should be conducted under uniform conditions; and administrators (for questionnaires that are not mailed and for interviews) should be naive with respect to any research hypothesis being tested, and they should be trained and

supervised. To eliminate the possibility of a biased sample, some step should be taken to contact a sample of nonresponders to determine that they do not differ from the sample of responders. The

caution factors associated with evaluation of a survey are in Box 4.1. We begin our critical review by evaluating a survey together. Then you will evaluate another survey by yourself.

Box 4.1 Caution Factors When Evaluating Surveys

- The sample frame
- The percentage of participants selected
- The criterion of exclusion
- The response return rate (is it less than 70%?)
- The method used to contact nonresponders
- The likelihood that characteristics of nonresponders introduced bias
- The nature of the questionnaire
- The reliability and validity of the questionnaire
- The number and characteristics of interviewers
- The training and supervision of interviewers
- The uniformity of the interview settings

STUDY EXAMPLE 4.1: "PHYSICAL PUNISHMENT BY MOTHERS AND FATHERS IN BRITISH HOMES"

This survey, conducted in England, determined whether mothers or fathers differed in the administration of punishment to their children.

The Study

Nobes, G., Smith, M., Upton, P., & Heverin, A. (1999). Physical punishment by mothers and fathers in British homes. *Journal of Interpersonal Violence, 14*(8), 887–902. Copyright © 1999 by Sage.

Recent evidence indicates that harsh physical punishment of children leads to the development of emotional and behavioral problems. . . . Furthermore, it is likely that much abuse begins as physical punishment that goes too far. . . . To understand parental violence against children and to identify the characteristics of potential abusers, it is important to address the question of whether mothers or fathers administer most physical punishment. . . .

Two British studies have reported that mothers administer physical punishment considerably more frequently than fathers. According to the mothers interviewed . . . 11.1% of mothers and 3.3% of fathers smacked [or spanked] their 3-year-old children daily or more frequently. Similarly . . . 15% of mothers of 11-year-olds reported smacking weekly or more often, compared with 5% who said their partners smacked this frequently. Furthermore, 79% of the mothers of 7-year-olds . . . reported that they smacked children more than did fathers, whereas only 10% of mothers said that fathers smacked most. . . .

► (Note that greater frequency of spanking by mothers is consistent with the mothers spending more time with their children during the day, when children are most likely to misbehave.)

However, mothers tend to underestimate fathers' use of physical punishment (and fathers underestimate mothers). . . . Inasmuch as the . . . studies were based only on mothers' reports of their own and their partners' actions, it is likely that their findings are misleading. . . . In contrast, studies based on self-reports . . . and on people's reports of their own parents . . . all indicate little or no disparity between mothers' and fathers' rates of physical punishment.

► (Note that self-reports are not always reliable, especially if the respondent does not want to be seen in a bad light.)

Although this comparison of findings strongly suggests that because they have relied on mother reports the British studies have underestimated the father's role in administering physical punishments, another explanation is that British families differ from those in countries in which most other studies have been conducted. . . . For these reasons, there is a need for information based on mothers' and fathers' self-reports. In addition, . . . it is important to assess how severely children are punished and to record parents' use of forms of physical punishment other than smacking.

Another issue that has rarely been addressed is the degree to which the extent of parents' physical punishments is a function of the amount of time they spend with children. Because most fathers have less contact with their children than do mothers, a finding of parity between mothers and fathers would mean that fathers are more physically punitive than mothers while with children.

In this article, we report a community self-report study of British mothers' and fathers' accounts of their use of physical punishments. Parents were asked about the frequency and severity of a number of forms of physical punishment. In addition, they reported who they considered to be responsible for most child care in their families and the activities they engaged in with their children. It was predicted

that mothers and fathers would administer physical punishment to similar extents in terms of frequency and severity. In families in which parents were equally responsible for child care, however, fathers would use more physical punishments than mothers.

1. What was the rationale for the survey?

Harsh punishment can lead to behavioral and emotional problems. Because abuse can start as punishment that goes too far, it is important to examine punishment of children by parents. Earlier British studies that relied on reports by mothers found that they punished their children more than did fathers. Other studies that relied on self-reports found that both parents punished their children equally. The disparity may be due to the method of gathering data or the origin of the studies. To clarify the discrepancies, it is important to obtain self-reports of both parents' use of punishment, its severity, and its form other than spanking. It is equally important to consider the amount of time each parent spends with the children. If fathers and mothers equally punish their children but fathers spend less time with the children, they are more punitive.

2. What was the specific purpose of the survey?

Mothers and fathers were asked to report frequency and severity of different forms of punishment and to indicate which parent had primary care of the child. It was predicted that both parents would be similar in their reports, but when both parents had equal responsibility for child care, the father would be more punitive.

Method

Sample

The sample for this study was composed of 465 parents from 366 two-parent families. There were 362 mothers (all primary caretakers), 99 of their partners (secondary caretaking fathers), and 4 primary caretaking fathers.

These parents were from a larger original sample . . . of 434 families, each with a child aged 1.0, 4.0, 7.6, or 11.6 years. The original sample was randomly selected from health authority lists in an urban area outside London and an area of South London. The sampling was stratified so that, in the original and present sample, there were approximately equal numbers of children of each gender and each age group. All the primary caretakers in the original sample (425 mothers, 9 fathers) were interviewed, and in every fourth family the secondary caretaker was also invited to participate. If that family was headed by a lone parent, the secondary caretaker in the next eligible family was asked to take part.

► (Note that the basis for stratification does not mean that the families had only one child. It means that a particular male or female child of the desired age was part of the family. Nonetheless, we do have an excellent example of stratified random sampling.)

The 68 lone parents were excluded from the present sample because all but 5 were mothers, and their inclusion would have therefore risked confounding the actions of mothers with those of lone parents. . . . Also excluded were ineligible families (e.g., in which the respondents did not speak English) and 18 fathers who refused to participate.

► (Note that the original sample included 434 families [425 primary care mothers and 9 primary care fathers]. The new sample excluded 68 of these, 63 mothers and 5 fathers, leaving 425 – 63 = 362 primary care mothers and 9 – 5 = 4 primary care fathers. The further exclusions [non-English-speaking and 18 fathers who refused to be interviewed] had to have occurred with the original sample. Keep in mind that the 18 refusers may have punished their children more than those who agreed to participate.)

The response rate of primary caretakers from the original sample was 80.9%. Because no information about family structure or responsibility for refusal was available for most of the families who refused, we can only estimate the relative response rates of mothers and fathers. Exclusion of families in which both parents were approached, and of those in which the father was the primary caretaker, leads to an estimated response rate of 85.3% in families in which only the mother was asked to participate. Among confirmed two-parent families in which both were invited to take part, the figure fell to 65.8%. These figures indicate that approximately 15% of primary caretaking mothers and 23% of secondary caretaking fathers refused to be interviewed.

► (Keep in mind that 65.8% acceptance is low but not unreasonably so, considering that the study involved families. But the 15% of primary care mothers and 23% of secondary care fathers who refused to participate may represent harsher punishers than those who participated. It is also possible, however, that refusers did not have time to participate.)

All but 20 mothers and 9 fathers were White. Thirty-two mothers and 10 fathers were born outside the United Kingdom. There were 8 stepfathers and 2 stepmothers, and five children were adopted. Of the fathers, 8 were unemployed and 2 worked part-time, whereas 166 (45.5%) mothers were not employed and 101 (27.7%) worked part-time. The remaining parents worked at least 25 hours

per week. For the purpose of this research, "fathers" are defined as men who had a [cohabiting] relationship of at least a year's duration with the mother.

► *(It is worth keeping in mind that we don't know how many "fathers" were birth fathers, were always involved in the lives of their children, and had cohabited with the mother as a stepfather for at least 1 year.)*

3. How was the sample of parents obtained?

Through stratified random sampling, families were selected from health authority lists to ensure equal numbers of male and female children at ages 1.0, 4.0, 7.6, or 11.6 years. From the 434 parents, 68 lone parents were excluded, leaving 366 two-parent families. Non-English-speaking individuals also were excluded, and 18 fathers refused to participate.

4. What was the response rate of confirmed two-parent families in which both parents were asked to participate?

It is estimated to be 65.8%, with about 15% of primary care mothers and 23% of secondary care fathers refusing to participate.

5. What can be said about the refusers?

They may have been parents who severely punish their children. Because there were more fathers than mothers who refused, a finding of no differences in this survey may be misleading. On the other hand, the refusers may have been unable to participate.

6. Describe the final sample.

There are 362 primary care mothers, 99 secondary care partners, and 4 primary care fathers. Most were White, most of the "fathers" (men who had lived with the mother for at least 1 year) were employed full-time, whereas only 45.5% of the mothers stayed at home full-time.

Procedure

Parents . . . were sent a letter about the study. Unless they declined at this stage, parents then were visited to explain the research and, if they were willing, to arrange the interviews. Interviews took place in the family home. When both parents participated, mothers and fathers were interviewed separately and, in all

but eight families, simultaneously, by different members of the research team. Interviews lasted approximately 3 hours.

► *(Note that interviews were not taped, for practical reasons. Therefore, all subsequent measures relied on interviewer notes as well as their accuracy.)*

7. What was the general procedure?

Letters initially were mailed to potential participants. Positive responders were visited, and those who agreed later were interviewed for about 3 hours. Most parents who were both included were interviewed separately at the same time by different interviewers. (A footnote thanks three individuals for their assistance. It is difficult to know whether they or the authors conducted the interviews.)

Measures

The interviews were interview-led and semistructured. Mothers' and fathers' interviews were directly comparable within an age group. There were some small differences between the interview schedules for different child-age groups, so that the content was appropriate to the child's age. Parents were asked about the nature and frequency of their own use of physical punishments of children. The interview also covered a number of other topics relating to parental control, family members, and relationships.

Types of punishment. Physical punishments were grouped into one of four categories, these being

"smacking/hitting," including spanking, slapping, and beating. "Smacking" usually involves use of the open hand. . . .

"physical restraint," including pushing, shoving, shaking, throwing, and holding. . . .

"punishment by example," consisting of punishments such as squeezing, hair pulling, biting, and pinching, usually administered with the intention of demonstrating to children the consequences of their actions.

"ingestion," including washing the child's mouth out with soap and water and forced feeding. . . .

► *(Note that each category is clearly defined.)*

For each category, parents were asked whether they had administered to the child any of the different punishment types included in the category. . . .

Frequency of punishment. Unless parents replied that they had never used a punishment of a particular type, frequency of use of that type was established by

asking such questions as, "When was the last time? And the time before?" and so forth, until the interviewer was confident [he or she] could make an estimate of the frequency of use in the last year (3 months for 1-year-old children). . . . Frequencies of use of the four physical punishment categories were coded on a seven-category ordinal scale, from *never to daily or more often*. Reliability of these measures, assessed by two researchers interviewing and coding 12 mothers 2 weeks apart, was high ($\alpha = .98$ for smacking/ hitting).

► (Note that reliability of scoring was established, but a similar procedure was not reported for the remaining categories of punishment, nor are there reports of validity measures.)

A derived measure of the frequency of use of physical punishment, of any kind, was also calculated. This measure ("any physical punishment") takes into account the frequency of all four categories of physical punishment. It is based on the frequency of the category of punishment that the parent uses most often: If a parent smacks once or twice a week, pushes a few times a year, and does not use other physical punishments, the frequency of use of any punishment will effectively equal that of smacking ("weekly or more often"). . . .

Severity of punishment. To assess severity of punishments, parents were asked to describe actual events. . . . The intensity and consequences of the most recent incident in each category were recorded from parents' replies to such questions as, "[on the last occasion,] how many times did you smack? Where on her body? Did it leave a mark? Did you use your hand (which part)? Or something else?" Information about the most severe punishments ever administered by the parent was sought. . . .

On the basis of all the information available, interviewers judged the severity of the most serious punishment in each category. . . . "Mild" punishments were non-prolonged and nonrepeated actions that did not involve the use of implements, had no prolonged effect, and were not intended to harm. "Moderate" punishments were intermediate between mild and severe. "Severe" punishments were those that were intended to, had the potential to, or actually did cause harm to the child and included actions that were repeated, prolonged, or involved the use of implements.

► (Note clear definitions of severity.)

All incidents that were rated either moderate or severe were cross-validated on completion of fieldwork, at which point it was possible to collate and directly compare descriptions of all these events. . . .

Parental involvement. . . . The frequency during the previous week of joint engagement in six activities (e.g., reading, playing indoors and out, and household tasks) was measured on a scale of 0 (*no joint activities*) to 18 (*all six activities daily or more often*). . . . Parents were also questioned about who took most responsibility for looking after the children, and fathers were asked about the amount of time that they spent alone with their children. This was rated on a 6-point scale, from *none to more than 2 whole days a week* (or equivalent).

Punishment roles. Parents were asked who they considered to be the main punisher. . . . The possibility that, in some families, fathers administer punishments on behalf of the mother was investigated by asking both parents whether one would sometimes execute a punishment that was initiated by the other.

8. What general measures of punishment were made?

Four types of punishment, their frequency (including all forms of punishment as a single score) and severity, as well as parental involvement with the child and parental roles in administering punishment. In most instances, responses were converted to category scales.

Results

Frequency of Physical Punishments Reported by Mothers and Fathers

The great majority of parents had physically punished their children at some time in some way. . . . Approximately equal proportions of fathers and mothers reported having never done so (24 mothers [6.6%] vs. 9 fathers [8.7%], ns). . . . No overall difference was found between mothers' and fathers' frequencies of administration of physical punishments. A total of 26.8% of mothers, compared with 26.2% of fathers, reported using some sort of physical punishment at least as often as weekly.

Almost all parents had smacked, slapped, spanked, or hit their child (see Table 4.1). Mothers and fathers were almost equally likely to report never having punished in any of these ways (8.8% mothers, 9.7% fathers, ns). There was no overall difference between mothers' and fathers' frequency. . . . in this smacking/hitting category. However, a somewhat greater proportion of mothers than fathers smacked weekly or more often (25.1% mothers, 17.5% fathers. . . .)

In the category of physical restraints as means of punishment or control. . . . fathers' frequency of use was significantly higher than mothers' ($p < .005$). Almost 4 times the proportion of fathers as mothers reported using such methods with a frequency that was rated as at least monthly (4.7% mothers, 18.6% fathers). . . . Fathers who frequently used physical restraints were also more likely than other fathers to smack frequently ($r = .35, p < .001$).

Approximately equal proportions of mothers and fathers reported ever having used punishments in the "punishment by example" category ... and no difference was apparent between mothers' and fathers' rates.

The small number of parents who had used some form of ingestion as a punishment were mostly mothers, although this difference was not significant (9.7% mothers, 3.9% fathers. ...).

There were no significant differences between mothers' and fathers' rates of punishment either of boys or of girls: 46.4% mothers and 40.7% fathers smacked their sons at least monthly, whereas 33.1% mothers and 30.6% fathers smacked their daughters at least monthly.

There were few differences between mothers and fathers relating to children's ages. Among parents of 1-year-olds, fathers reported significantly higher rates of use of any punishment than did mothers ($z = 2.35, p = .02$). This difference was largely accounted for by fathers' markedly more frequent use of physical restraints at this age ($z = 3.32, p < .001$). ... Mothers of 4-year-olds smacked more than did fathers of 4-year-olds ($z = 2.00, p = .05$). ... There were no apparent differences between mothers and fathers of 7- or 11-year-olds in their frequency of use of punishment.

Table 4.1 Percentage of Parents Using Physical Punishments

Punishment Category	N	Ever	In Last Year	Monthly or More Often		z	p
				Ever	More Often		
Smacking/hitting							
By mothers	362	91.2	76.8	39.8			
By fathers	103	90.3	76.7	35.9	1.07	.29	
Physical restraint							
By mothers	362	39.5	25.4	4.7			
By fathers	102	50.0	42.2	18.6	3.00	.003	
Punishment by example							
By mothers	361	38.8	12.7	1.1			
By fathers	103	30.1	14.6	2.9	1.20	.23	
Ingestion							
By mothers	267	9.7	5.2	0.0			
By fathers	77	3.9	3.9	0.0	1.57	.12	
Any physical punishment							
By mothers	362	93.4	80.4	42.5			
By fathers	103	91.3	80.6	45.6	.17	.87	

Table 4.2 Most Severe Punishment Administered by Parent

Most Severe	Mother		Father	
	n	%	n	%
None/not applicable	24	6.7	9	9.1
Mild	163	45.5	52	52.5
Moderate	115	32.1	17	17.2
One or more severe incidents	56	15.6	21	21.2
Total	358	100.0	99	100.0

9. What were the basic results regarding frequency of punishment by mothers and fathers?

More mothers than fathers smacked their children on a weekly basis, but more fathers used physical constraint as punishment. There were no differences in the use of punishment by example between mothers and fathers, nor were there differences in the frequency of punishment of all kinds of male or female children. But fathers punished 1-year-old children most frequently, and more often than did mothers, with physical restraint, and mothers smacked 4-year-old children more frequently than did fathers.

Severity of Fathers' and Mothers' Physical Punishments

The proportions of mothers and fathers who reported ever having used punishment rated as severe (Table 4.2) were 15.6% and 21.2%, respectively ($\chi^2(1) = 1.72, p = .19$).

The proportion of mothers who reported having used an implement in punishing their children was also less than that of fathers (35 mothers [10.7%] vs. 14 fathers [15.2%], $\chi^2(1) = 1.56, p = .21$). The majority of implements used were either slippers or wooden spoons ... sticks or belts ... [hairbrushes], folded newspapers, or a flyswat. ...

Fathers had inflicted significantly more punishments rated as severe on sons than daughters (30.8% vs. 10.6%, $\chi^2(1) = 5.98, p = .01$). ... The difference between mothers' and fathers' use of severe punishments on boys was marginally significant ($\chi^2(1) = 3.40, p = .07$) but not on girls ($\chi^2(1) = .07, p = .78$). There were no significant differences between mothers' and fathers' use of severe punishments at any of the four age groups.

10. What were the basic results regarding severity of punishment by mothers and fathers?

There was a tendency, although it was not significant, for more fathers than mothers to have reported using severe punishment at some time. There also was a tendency for more fathers than mothers to use implements other than a hand to punish their children. Fathers used severe punishment more often on sons than on daughters and did so more often than did mothers. There were no differences in frequencies of severe punishment by fathers and mothers as a function of age of the child.

► (Here is an opportunity to check some statistics. We have a 4 [severity] by 2 [parent] contingency table and can perform a chi-square test to determine whether the degree of severity of punishment depends on whether the parent was a mother or father. I performed a straightforward test and obtained $\chi^2(3) = 8.851$, $p < 0.5$. Therefore, the degree of severity of punishment depends on whether the punisher is a mother or a father. Now, to isolate the source of the contingency, we can look at the individual cells and compare the extent of difference between observed and expected frequencies by $z = (\text{observed frequency} - \text{expected frequency}) / \sqrt{\text{expected frequency}}$. There is only one outstanding difference: Fewer fathers administered moderate punishment [17] than would be expected [28]—derived from the performance of the test—if both parents were equally likely to administer this degree of punishment. This also means that more mothers than fathers administered moderate punishment. [An excellent discussion of isolating the source of a significant chi-square is in Hays [1988, pp. 775–779].]

Involvement with child. In the families in which mothers reported themselves to be the main caretakers (81.4%), there were no significant differences either between mothers' and fathers' frequency of smacking/hitting or in the proportion of mothers and fathers who reported having used severe punishments. . . . In the remaining families in which fathers were described by mothers as taking at least as much responsibility as the mother for the children, all 67 mothers and 25 of the fathers were interviewed. There was no evidence . . . of differences either in self-reported rates of smacking/hitting or in the use of severe punishments by mothers and fathers. . . . Furthermore, the parents' role as either main caretaker or secondary/equal caretaker was not associated with mothers' or fathers' frequency of punishment or with mothers' or fathers' use of severe punishment.

The extent of mothers' and fathers' involvement in activities with their children was positively correlated with their frequency of use of smacking/hitting ($r = .25$, $p < .001$ and $r = .24$, $p = .04$, respectively) but was associated with neither mothers'

nor fathers' use of severe physical punishments. According to self-reports, there was no significant difference between mothers' and fathers' involvement in activities with their children. . . . Within the group of parents . . . who were described as relatively highly active . . . with their children, mothers and fathers smacked or hit their children equally frequently. . . . Similar proportions had used a severe physical punishment. . . . Mothers . . . and fathers . . . who engaged in relatively few activities with their children also smacked or hit at similar rates and were almost equally likely to have used a severe physical punishment. . . .

Fathers' frequency of smacking/hitting was not correlated with the amount of time they were alone with their children during each week. . . . Comparisons of fathers who had and had not used one or more severe physical punishments showed no difference in the amounts of time they spent alone with their children, and they reported being responsible for similar proportions of the caretaking of their children.

11. What were the results regarding role of caregivers?

Although frequency of punishment was related to amount of time spent with a child, there were no reported differences in that frequency between mothers and fathers, even when mothers were the primary caretakers. Moreover, mothers and fathers reported being equally involved with their children, and regardless of whether this involvement was high or low, mothers and fathers reported punishing their children equally often. Punishment by fathers was not related to time spent alone with their children nor to severity of punishment.

Punishment Roles

According to mothers and fathers, fathers' punishments were at least as effective as mothers'. About half of the mothers (48.0%) and a third of the fathers (36.8%) reported that children responded best to punishments by fathers, compared with about 1 in 6 parents (18.7% mothers and 16.2% fathers) who thought that mothers' punishments were more effective.

Most of the mothers (76.3%) and fathers (61.0%) said that parents did not administer punishments for each other in their families. . . .

Discussion

The data indicate [near parity] between mothers' and fathers' frequencies of use of physical punishment of their children. Very similar proportions of mothers and fathers reported ever having administered a physical punishment, and almost identical proportions did so weekly or more often. An exception . . . is that fathers

more often physically restrained (pushed or forcefully held) their children than did mothers. This was most evident in parents of 1-year-olds. . . .

. . . Those fathers who frequently physically restrained their children were also more likely to smack/hit more frequently than other fathers. This suggests that fathers typically sometimes choose to smack or hit and at other times choose to restrain. In contrast, when mothers physically punish their children, they almost always resort to smacking/hitting.

We can speculate that the greater tendency for fathers than mothers sometimes to restrain their children occurs because . . . fathers are better equipped successfully to restrain the child and are less willing to hit, fearing that they might hurt or injure the child . . . supported by the fact that this difference between mothers' and fathers' rates of use of physical restraint was particularly evident among parents of 1-year-old children.

Although nonsignificant, the finding that higher proportions of mothers reported smacking/hitting their children weekly or more often than did fathers concurs with those reported by other researchers who recorded rates of smacking/hitting. . . .

No difference in incidence was evident in the present study when all physical punishments were considered. . . .

Regarding the severity of punishments, although mothers' and fathers' use of severe punishment showed no statistically significant differences. . . . the present data suggest that fathers were more severely punitive. Greater proportions of fathers than of mothers reported having used implements and were judged to have administered a punishment rated as severe. These differences solely reflect fathers' greater tendency than mothers to punish sons severely. . . .

These data concur with those of studies of official and clinical reports of abuse that adjusted for father absence. . . . They also are consistent with the findings of studies of victims' reports of their own parents. . . . All indicate that fathers are more often the perpetrators of violence to, and abuse of, children. However, . . . according to other community self-report studies, greater proportions of mothers than fathers have severely punished their children. . . .

The present study, then, is alone among self-report studies in indicating that fathers are no less, and may be *more*, severely punitive than mothers. . . . An explanation is that the fathers in the present study were willing to report their administration of severe punishments, because the relevant questions were asked in the context of a long interview concerning many aspects of family life. . . .

. . . Mothers were the main caretakers in the large majority of families. The finding that, in the families in which this was the case, mothers and fathers administered approximately equal amounts of physical punishment, suggests that fathers are more punitive relative to the amount of time spent with children. However, in

the small number of families in which the father was as responsible for caretaking as the mother, again no difference was found. . . .

A possible explanation of these findings is that fathers who take equal responsibility in caretaking, and spend as much time with their children as do mothers, are an exceptional group who are also similar to their partners in their use of physical punishment. The large majority of fathers, though, have a more traditional father's role that involves less caretaking and more physical punishment relative to the amount of time they spend with their children. . . .

In summary . . . British mothers and fathers reported using physical punishment with approximately equal frequency. No significant difference in the proportion of mothers and fathers who had used physical punishments was found. A slightly higher proportion of fathers than of mothers reported having inflicted severe punishment. Because most mothers spend more time with their children than do fathers, these findings are consistent with the view that, relative to the amount of time they spend with their children, most fathers are somewhat more physically punitive than mothers. However, in families in which mothers' and fathers' caretaking roles are similar, so too are their roles as disciplinarians.

12. What were the major conclusions?

British fathers and mothers administer punishment with about equal frequency, although there are some differences in type of punishment administered. Notably, mothers used smacking and hitting; fathers used smacking and hitting as well as physical restraint. And although the difference was not significant, there was a greater tendency for fathers than mothers to use more severe punishment, especially with their sons. Because fathers and mothers were found to be equal in frequency of punishment, and because fathers spend less time with their children, fathers are relatively more punitive than mothers.

13. Are the conclusions justified?

Tentatively, the conclusions are justified if the self-reports are accurate and if nonresponders would have responded in the same way. Even here, the conclusions would apply only to parents of similar socioeconomic status, with similar families and work histories of each parent. Self-reports, however, are not always accurate. Parents had to rely on their memory of how frequently they administered punishment to a particular child (if there was more than one child in the family) during the last week, month, and year. There is room for error in memory the longer the time span. Moreover, parents would not want to present themselves in a bad light, and this, too, could result in underestimation of frequency, severity, or both.

Indeed, no parent would want to be accused of child abuse. Finally, all figures were based on rating scales, some of whose reliability were not reported, and interpretations were based on notes by the researchers, who may or may not have been the authors. There is a possibility of error here, too.

A more important problem is with nonresponders. A relatively large percentage of fathers from two-parent families refused to participate, and one has to wonder why. If they were too busy to participate, then they probably would have responded as did the participants. But if they did not want to discuss their punishment behavior, they may have been more frequent (and possibly more severe) punishers, and study results may have been very different. Related to this, we don't know how many of the "fathers" who cohabited with the mothers actually fathered the children. Behavior of these two groups may well have differed.

STUDY EXAMPLE 4.2: "BAD DATES OR EMOTIONAL TRAUMA? THE AFTERMATH OF CAMPUS SEXUAL ASSAULT"

The following article reports the results of a survey conducted on female college seniors regarding experiences of being forcibly raped or being subjected to nonconsensual sex while intoxicated.

The Study

Schwartz, M. D., & Leggett, M. S. (1999). Bad dates or emotional trauma? The aftermath of campus sexual assault. *Violence Against Women*, 5(3), 251–271. Copyright © 1999 by Sage.

The public image of campus sexual assault has come full circle, at least in the viewpoints expressed in the mass media... It was not until the 1980s that the media generally became aware that sexual assault on the college campus was more than a problem of stranger rape... Study after study began to document that men the women knew and with whom they were often friends or even lovers were the most common sexual aggressors on campus....

In the 1990s, however, there has been an extensive movement to minimize the problem... The "backlash" has been a strong tactical offensive by men and a few women who argue that campus rape either does not occur or else what we have been calling campus rape is not particularly important... What researchers call

acquaintance rape is often simply the result of a woman regretting her consensual sexual encounters and consequently overreacting....

Roiphe's most popular claim is that the women who say that they are victims of acquaintance rape are commonly not affected by the sexual encounter.... Roiphe believes that... women [who were drunk when supposedly raped] know that they could have had less to drink and maintained some more control, but they did not, so the morning after they have to face up to the fact that they were too drunk and went a little too far with a man. Because the women were in control of their own drinking, they have no reason to blame the man for the sexual activity....

... What is important is that the backlash message has gotten into the general discourse. Do women who have had too much to drink and then find that a man has engaged in sexual intercourse with them without their permission become emotionally upset by this incident? Or, is Roiphe right that most of them just see this as a bad date? Who do these women blame for such incidents? Do they indeed see themselves as rape victims? The data in this article will speak at least preliminarily to these three concerns.

Sexual Assault on Campus

... Because these women often did not define their experiences as *rape*, rape must not be what happened. Researchers are making an issue out of sexual experiences that women simply engaged in but then regretted.... These women would have at least talked to a rape crisis center had they remotely thought they had been raped.... Because rape crisis centers do not automatically report rapes to the authorities, any woman would feel safe about reporting the rape and discussing her experiences.

... Because these women did not report that they had been raped, then... researchers are using a different definition than the one in the criminal law—they are using the radical feminist critique.... Because the women questioned were in college, they were too educated not to realize they had been raped....

Another controversial area deals with the problem of sexual intercourse while the woman is too drunk or high to give consent.... The issue is whether a woman who has had too much to drink and is physically incapable of giving her consent has been raped or if she has simply not refused.... But the law is clear... that the crime of rape consists of sexual acts without a woman's consent and that the act is rape if the woman is incapable of giving consent (unconscious, of low mental capacity, too intoxicated to give consent)....

... In our study we tried to make the connection [between drinking and consent] more clear by asking if the woman had sexual intercourse when she did not want

to because she was unable to give her consent or stop the man because of being intoxicated or on drugs. This is an act that is without doubt a felony crime. . . .

Self-Blame

The importance of these questions is that many women believe that if they are too intoxicated or drugged to say no, then they are to blame for the sexual assault. . . . There has not been much study of this question, but the few studies that have compared stranger rape victims with acquaintance rape victims have found more self-blame among the latter group. . . .

As critics have noted, some respondents . . . do not claim that they are rape victims. It is not uncommon for women who have been attacked to be unable to fully understand that the incident that is bothering them is actually an event defined by criminal law as felony rape. . . .

Emotional Effects

An area closely related to self-blame is how affected emotionally and psychologically the woman is after a sexual assault. . . . The reaction of her friends and the amount of self-blame she takes upon herself can be relevant to this distress. There is a reasonably large literature in psychology that divides the victims of sexual assault into two groups: stranger rape and acquaintance rape. . . . Very little work had been done comparing acknowledged versus unacknowledged rape victims. . . . Women who do not admit to themselves that they are rape victims are just as likely to suffer from psychological distress as those who name their experiences as rape.

. . . Critics have claimed that women who engaged in unwanted sexual intercourse or other sexual acts while too intoxicated to give consent or protest were not as psychologically distressed as women who were victims of rape by force. . . . Thus, a goal of this research was to investigate the claims of critics that women are not as affected by acquaintance rape as feminists claim, by looking directly at the differences between those women who were assaulted while drunk or high and those women who were forced into sexual intercourse or acts.

Hypotheses

It is an important part of backlash arguments that women who experience unwanted sexual intercourse because of intoxication are relatively unaffected by the sexual experience. . . . Women physically forced to have sex are the victims of "real rape" . . . and . . . "real rape" victims are more emotionally distressed by what happened. Thus, the first hypothesis is

1. The women who were raped due to intoxication are significantly less affected by the event than the women who were raped due to physical force. Second, there is no question that an important psychological factor in rape is self-blame and that a large number of rape victims blame themselves after being raped. . . . Thus, the second hypothesis is
2. The women raped because of alcohol or drugs blame themselves for the event more than do the women raped by physical force. . . . Women whose experiences meet the legal definition of rape generally do not believe that they were raped. . . . Here, . . . we further specify that even fewer women who were raped because of intoxication will report that their experience was a rape.
3. The women who have been raped because of force will be more likely to label their experiences as rape than will women who had unwanted sex because they were too intoxicated to resist.

Method

. . . We created a questionnaire using previously used surveys, but mainly Koss's Sexual Experiences Survey (SES). . . . We added open-ended questions at the end discussing the aftermath of the experience. . . . We decided to limit our sample entirely to seniors. In this way, we thought, we might be maximizing the amount of time available for sexual experiences while at college.

▲ (Regarding the selection of seniors, the authors have noted that the opposite effect could have occurred—i.e., that fewer victims were available because some victims might have quit college or transferred to another college.)

The university used requires senior-year integrative classes. By sampling these classes [on the first day of classes], we obtained usable questionnaires from 388 females in 25 classes.

▲ (Data collectors were female.)

Although data were simultaneously collected from men, in this study we used data only from the female respondents.

▲ (The authors approached all 36 instructors, but 11 refused because they could not afford the class time. The 11 did not seem to be associated with a particular area of study.)

The students were given a human subjects research consent form, which was read verbatim by the administrator, then signed, dated, and collected from each participating student. Each anonymous survey was completed voluntarily, and students were informed verbally and in writing that they were free to stop answering questions at any time with no questions asked and no penalties. Only five females did not complete usable questionnaires.

Given the criticisms that SES questions were loosely worded, we tightened up several questions to remove any misunderstanding. For example, . . . we asked, "Have you engaged in sexual intercourse when you didn't want to but were so intoxicated or under the influence of alcohol or drugs that you could not stop it or object?" . . . It is clear that the reason the woman had sex with the man was because she was physically or mentally unable to resist. It is clear the act was rape under virtually all state statutes. . . .

. . . To see if the two types of victims under discussion here were different on our questions, we created a variable called *rapetype*. This variable divided attempted and completed rape into two categories: (a) rape (and attempted rape) while unable to give consent because of alcohol or drug intoxication, and (b) rape (or attempted rape) because the male used or threatened physical force. . . . Of course . . . we did not use the word *rape* with the respondents; they were asked about their behavior as described in the statute.

Self-blame was measured by a simple variable. Women who reported that they were the victims of events measured by *rapetype* were asked to report for the most serious event whether they blamed themselves for what happened, whether they blamed the man for what happened, or whether they blamed both themselves and the man. . . .

Finally, the women in this survey who reported that they were the victims of unwanted sexual experience were asked to self-report on how much they were psychologically and emotionally affected by that experience. We asked the women to categorize their response using the following choices:

1. It was not very important to me; I was not much affected.
2. It did not bother me for very long; I bounced back fairly quickly.
3. It affected me. I changed as a person (e.g., not as trusting, depressed, unhappy, or some other reaction).
4. It deeply affected me and caused emotional pain.

Findings

Of the 388 women who filled out the questionnaires in this survey, 65 reported that they were victims of an event that would under Ohio law be considered a

felony rape. Thirty-five reported that they had been the victims of unwanted sexual intercourse when they were helpless to resist or stop the man, whereas 30 reported that they were overcome with force or a threat of force.

In Table 4.3, we see that of the women raped due to physical force, there were 16.7% who claimed to be unaffected versus only 5.7% of women victimized because of alcohol or drugs. . . . Of the women raped because they were unable to give consent due to intoxication, 94.2% claimed to be affected by the event versus 83.4% of the physically forced women. . . . There were similar numbers of women affected or deeply affected by the experiences in each category. . . . There is no statistically significant difference in the reported emotional outcome between the two groups of women. . . . These data do not support the first hypothesis prediction that women who were raped by intoxication were less likely to report psychological or emotional distress.

Table 4.3 Whether Women Were Affected by the Type of Assault

How Affected	Type of Rape		Total
	Alcohol	Physical Force	
Unaffected	5.7	16.7	10.8
Somewhat affected	51.4	26.7	40.0
Affected	31.4	36.7	33.8
Deeply affected	11.4	20.0	15.4
Total	53.8	46.2	100.0

Note: Total affected: alcohol = 94.2%; physical force = 83.4%.

Chi-square = 5.18

df = 3

Significance = 0.159

Table 4.4 shows *rapetype* by whom the women blame for the event. Once again, it is the lack of statistical significance in this table that is interesting. . . . 79.3% of the women who were raped while intoxicated put all or part of the blame on themselves. All of these women were the victims of a felony crime, and all said specifically that the only reason that they had unwanted sexual intercourse was that they were unable to resist or fight back. What might be surprising to many is that 50% of the women raped by force or threat of force also took on some degree of self-blame. . . . Slightly more than one fourth of all of the women raped while intoxicated completely blamed themselves for what happened. However, virtually one quarter of the women raped by force took on all of the blame also.

Table 4.4 Whom the Women Blame, by the Type of Assault

Who to Blame	Type of Rape		Total
	Alcohol	Physical Force	
Myself	27.6	22.7	25.5
Man	20.7	50.0	33.3
Both	51.7	27.3	41.2
Total	56.9	43.1	100.0

Chi-square = 5.16

df = 2

Significance = 0.076

At the end of the questionnaire, removed from the other questions, we asked the simple question of whether the woman answering the survey had been raped since coming to college. . . . All of the 51 women who chose to answer this question had already stated in response to the questions used to develop Tables 4.3 and 4.4 that they had in fact been victims of rape since coming to college. However, Table 4.5, for the first time, introduces the use of the word *rape*. . . . All of the women being asked if they had been raped had in fact been raped. Yet, in response to the question that specifically asked "Have you ever been raped?" only 1 victim (3.3%) raped while too intoxicated to give consent answered affirmatively, and only 5 victims (23.8%) raped because of physical force said that they had. The latter figure generally agrees with . . . researchers who have found that about 25% or so of rape victims label the event as rape.

Table 4.5 Whether Woman Says She Was Raped, by Type of Assault

Have Ever Been Raped	Type of Rape		Total
	Alcohol	Physical Force	
Yes	3.3	23.8	11.8
No	96.7	76.2	88.2
Total	58.8	41.2	100.0

Chi-square = 4.99

df = 1

Significance = 0.026

Phi = .31

What is important in Table 4.5 is that what Koss called the problem of the "hidden rape victim" is much worse in the case of the woman raped while intoxicated. Hidden rape victims who do not define what happened to them as rape do not seek out the services of rape counselors, do not attend to various mental health services, and often do not understand why they are suffering from various symptoms of emotional pain. . . . These hidden rape victims do not seek care because they do not see that they are victims. They do not make this recognition because they live in a society that makes women responsible as gate-keepers for sexual relations, rewards men for sexual aggressiveness, and blames women for their own victimization. What we have found in this society is that a tremendous number of women engage in self-blame for being victimized by predatory rapists.

Discussion and Conclusions

The first decision made in this study was to reject Hypothesis 1; there is no evidence that women raped due to intoxication were less affected emotionally than women who were raped by force. The numbers show that virtually all victims of rape are affected, regardless of the circumstances surrounding the act. . . . Virtually everyone who has done research in this field argues that the methodology we used is likely to be conservative in terms of eliciting admissions from women that they were victims of rape. . . . Many women report on anonymous questionnaires that they have never told anyone of their experience, and it is at least logical to assume (if we cannot prove it) that some of these women would continue to keep their secret, even in an anonymous questionnaire.

The second hypothesis specifies that the women who were raped while intoxicated would blame themselves more than would women who were raped by force. This hypothesis was also rejected. Still, although the table did not reach statistical significance, it is instructive to look at the percentages in the table. As mentioned earlier, about one quarter of all rape victims blame themselves entirely. . . . Women continue to blame themselves even when they are the victims of a rape accomplished through the use of force. . . .

. . . So many women are affected emotionally and psychologically by it but still blame themselves and do not report it to the police. . . . Here, we found . . . of 43 women who answered this question, only 2 (4.7%) had reported their sexual assault to the police. Women generally do not report their victimization, in part because of self-blame and embarrassment. . . .

Technically, our third hypothesis was confirmed. There is a statistically significant difference in that women who were raped by force or threat of force are more likely to classify their experience as rape than are women raped while intoxicated. The problem with this interpretation, however, is that it is only technically correct. . . . Although

the hypothesis is correct, actually very few women of any experience labeled their experience as rape.

The results of this study are simple enough. . . . The percentage of women who report being victimized by sexual assaults remains high. At the same time, the strong tendency shown here is for women to commonly take the blame upon themselves, either fully or partially, for the behavior of male rapists and sexual aggressors. This continues to create hidden victims and to keep women from seeking the help they need. . . .

The implications of this study are clear. . . . Rape programming [on] college campuses needs to center clearly on blame and self-blame. . . . Clearly, those women who have been victimized by acquaintances in a situation without force but where they were too drunk to resist also are in need of various forms of support. Furthermore, another form of rape programming must be to educate people on how to react to friends who have survived an unwanted sexual experience. . . . Universities and colleges need to actively work to search out hidden victims. . . .

CRITIQUE OF STUDY EXAMPLE 4.2

1. What was the rationale for the survey?
2. What was the specific purpose of the survey?
3. What was the general procedure?
4. What are some sources of selective losses?
5. How valid and reliable is the questionnaire?
6. What are some potential weaknesses in using this self-report questionnaire?
7. What were the reported effects of being raped and "rapetype?"
8. In Table 4.3, are degrees of freedom (*df*) for the chi-square test accurate?
9. What is the source of 53.8%? 46.2%?
10. What were the major results regarding self-blame and type of rape?
11. Look at the total percentages reported in Table 4.4. They differ from those (53.8% and 46.2%) of Table 4.3. What does this tell you?
12. What were the results regarding the question that specifically asked if the women had been raped since entering college?
13. Only 51 out of 65 women answered the rape question. What might you suppose about the 14 women (5 in the alcohol group and 9 in the physical force group) who did not answer it?

14. Examine Table 4.5, whose statistical analysis reveals a significant $\chi^2(1) = 4.99$ and an interpretation that a yes or no response depends on the group to which the rape victim belonged. Frequencies are easy to obtain: Each percentage is multiplied by its total *N*, to yield $51 \times .588 = 30$ and $51 \times .412 = 21$. Then $.033 \times 30 = 1$ and $.238 \times 21 = 5$ for the yes responses and the remaining frequencies are 29 and 16 for the no responses. You now have enough information to verify that if chi-square is computed in the usual fashion, it is indeed 4.99. However, two of the frequencies (1 and 5) are very small and require that a correction (of continuity) be applied to all cell values in calculating chi-square. A value of .50 must be subtracted from each absolute difference between obtained and expected frequencies before that difference is squared. For the first cell (yes, alcohol), this is $(1 - 3.53) - .50)^2 = (2.53 - .5)^2 = 4.1209$. Then 4.1209 is divided by 3.53, and so forth for each cell. Recalculate the chi-square with the correction. What is the result? What do you conclude?

15. What was the conclusion regarding the first hypothesis? Is it justified?
16. What was the conclusion regarding the second hypothesis? Is it justified?
17. What was the conclusion regarding the third hypothesis? Is it justified?

For answers to these questions, see page 356.

BIBLIOGRAPHY

Hays, W. L. (1988). *Statistics* (4th ed.). New York: Holt, Rinehart & Winston.

SUGGESTED READINGS

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. K. (2009). *Introduction to research in education* (8th ed.). Orlando, FL: Harcourt Brace.

Fink, A. (2009). *How to conduct surveys: A step-by-step guide* (4th ed.). Thousand Oaks, CA: Sage.

Fowler, F. J., Jr. (2009). *Survey research methods* (4th ed., Applied Social Research Methods Series, Vol. 1). Thousand Oaks, CA: Sage.

Frey, J. H. (1989). *Survey research by telephone* (2nd ed.). Newbury Park, CA: Sage.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley.

McMillan, J. H., & Schumacher, S. (1997). *Research in education: A conceptual approach*. White Plains, NY: Longman.

- Rea, L. M., & Parker, R. A. (2005). *Designing and conducting survey research: A comprehensive guide* (3rd ed.). San Francisco, CA: Jossey-Bass.
- Selltiz, C., Wrightsman, L. S., & Cook, S. W. (1981). *Research methods in social relations* (4th ed.). New York: Holt, Rinehart & Winston.
- Weisberg, H. F., Krosnick, J. A., & Bowen, B. D. (1989). *An introduction to survey research and data analysis* (2nd ed.). Glenview, IL: Scott, Foresman.

Chapter 5

Correlation Studies



Correlation studies attempt to establish relationships between two or more quantitative variables. Participants with the characteristics of interest are selected and measured, and relationships are assessed via correlation coefficients. It is difficult—if not impossible—to establish causal statements from simple correlations alone.

Observed correlations may be **spurious correlations**—i.e., due to unmeasured variables. A classic example is the positive correlation between ice cream sales and number of drownings. In this case, both ice cream sales and number of children swimming (and thus drowning) increases with daily temperature. The observed correlation disappears when temperature is held constant. Additionally, the direction of any presumed causal relationship may be difficult to establish because the variables may influence each other (e.g., loneliness and depression).

If only two variables are involved, we typically assess the relationship with a Pearson's product moment correlation coefficient r or **Pearson's r** . It is considered a **zero-order correlation** because it does not control for any other variable. The magnitude of the coefficient describes the strength of the relationship, and the sign (positive or negative) describes the direction of the relationship. **Positive relationships** indicate that both characteristics covary in the same direction (e.g., amount of exercise and fatigue). **Negative relationships** indicate that both characteristics covary in opposite directions; an increase in one is accompanied by a decrease in the other (e.g., hours of sleep and fatigue). The magnitude ranges from zero (no relationship) to one (perfect relationship). The closer the value is to 1.00, the stronger the linear relationship or degree of covariation.

One of the most important assumptions made when r is computed is that the relationship between the two variables