

The Truth (and Lies) About Statistics and Evaluation

Nicole P. Yuan
University of Arizona
Rape Prevention and Education Program
May 2006

Part I: The Truth (and Lies) About Statistics

- What statistics do you use in your programs?
- How do you use them?

Objective of Presentation

- To improve abilities to understand, critique, and use statistics appropriately
- To identify main sources for rape statistics

What Are Statistics?

- The science of collecting, organizing, and tabulating data (Holmes, 1990).
- Two types of statistics:
 - 1) Descriptive statistics (mean, median, mode, range, percentiles, standard deviation)
 - 2) Inferential statistics

DESCRIPTIVE STATISTICS are used to describe a sample.

MEAN = the average value (add up all values and divide by number of values)

MEDIAN = list scores from lowest to highest and select the one that is in the middle

MODE = most typical score/response

RANGE = highest score minus the lowest score

PERCENTILES = describe data ordered in hundredths. For example, a score that is equal or greater than 95 percent of those that took the exam is in the 95th percentile

STANDARD DEVIATION = relationship of a set of scores to the mean (if normal curve, 68% fall under 1 SD of the mean, 95% within 2 SD, and 99% within 3 SD)

INFERENTIAL STATISTICS involve using a sample to make judgments or conclusions about a particular population. An example is conducting a study about college students' sexual experiences to make judgments about college students in general.

How Can Statistics Be Used to Misinform an Audience?

- Numbers are used to purposely deceive the audience

Example: Among all households that were surveyed, 75% reported that they waited for marriage before having sexual intercourse

- Numbers are accidentally misused or poorly explained

- 1) How might this be a dishonest representation of the survey results?

The sample size might be really small. Maybe the study consisted of 4 households. By saying 75%, you may assume that the researchers used a big sample.

- 2) What data collection methods may have influenced the findings above?

Administered the survey to all individuals who attended a Sunday service at a Catholic church during July 4th weekend. Thus, most of the attendees were older and more conservative.

What Are Important Things to Consider When Using Statistics?

- Sample representativeness
 - Influenced by sample size
 - The number of participants in the study
 - Influenced by random sampling
 - When every person has an equal chance of being included in a sample
- Sample biases

SAMPLE REPRESENTATIVENESS – The goal is to obtain a sample that is close approximation of a population as a whole.

SAMPLE SIZE is the number of participants in the study. There are statistical procedures to determine what sample size you need to detect an effect. In general, the larger the sample, the more likely it will represent a population. It will capture the diversity inherent in populations (diversity in background characteristics, experiences, beliefs)

RANDOM SAMPLING is when every person has an equal chance of being included in the sample. Techniques include picking every nth person, randomly drawing names from a hat, using a random numbers list, or using a computer to randomly select individuals from a list.

SAMPLE BIASES – A sample is biased when it does not accurately reflect a population.

Q: What biases may exist among studies focused on sexual experiences?

Q: Are individuals who participate different from those who do not?

(1) May attract individuals who have had more or fewer unwanted sexual experiences, (2) those who respond may be more sensitive, interested, and open to discussing sexual violence issues, (3) they may be more educated and more interested in research.

What Are Important Things to Consider When Using Statistics?

- Measurement
 - How is the topic defined?
 - How are the questions asked?
- Replication of findings

MEASUREMENT

Q: What are some concerns about how rape is/used to be defined? How do rape definitions vary?

1. Use of such terms as “rape” or “sexual assault” instead of behaviorally-oriented definitions
2. Acknowledgement that rape can be committed by someone you know
3. Ack of rape within marital relationships
4. Ack of women-to-woman rape
5. Ack of oral and anal sex as forms of rape
6. Ack of degree of force, threat of force, and coercion

HOW QUESTIONS ARE ASKED may contribute to different responses. Different responses to questionnaires versus face-to-face interviews based on level of anonymity. In-person versus telephone interviews. Gender of interviewer.

REPLICATION – Because results can be influenced by biases or chance, studies should be replicated. You should avoid putting too much value in one single study until it has been replicated.

What Are the Main Sources of Rape Statistics?

See Dean Kilpatrick online article, *Rape and Sexual Assault*
(www.musc.edu/vawprevention/research/sa.shtml)

Government Sources

1) *FBI Uniform Crime Reports*

- .03% of individuals in Arizona reported experiencing forcible rapes in 2004 (FBI, 2005)
- Limitations in the data

FBI UNIFORM CRIME REPORTS – Characteristics include:

- 1) number of rapes and attempted rapes REPORTED to law enforcement each year,
- 2) definition is limited to vaginal penetration by a penis using force or threat of force

Rape Statistics Government Sources

2) *National Crime Victimization Survey*

- In 2004, the rates of rape/sexual assault were 1.4% for women and 0.1% for men per 1,000 persons or 1,000 households (Catalano, 2005)
- Male sample is based on 10 or fewer cases

Will discuss in more detail during group exercise!

Rape Statistics Research Studies

1) *The National Women's Study*

- Kilpatrick, Edmunds, & Seymour, 1992
- 13% of women reported being victims of completed rape during lifetime

2) *The National Survey of Adolescents*

- Kilpatrick & Saunders, 1996
- 8% of adolescents reported being a victim of at least one sexual assault

THE NATIONAL WOMEN'S STUDY

Characteristics include: 1) national sample of 4,008 adult women

2) conducted in 1990

3) did not assess attempted rape

4) used similar definitions to NVAW Survey

THE NATIONAL SURVEY OF ADOLESCENTS

Characteristics include: 1) national sample of 4,023 adolescents ages 12-17

Rape Statistics Research Studies

3) *The National Violence Against Women Survey*

- Tjaden & Thoennes, 2006
- 17.6% of women and 3% of men reported experiencing a completed/attempted rape in lifetime

4) Small research studies

SMALL STUDIES – Conducted with various populations, such as college students, community samples, clinical samples. Sometimes are convenient samples.

Exercise: Critique of Rape Statistics

Divide into three groups. Read the article assigned to your group. Answer the following questions:

- Who are the participants?
- How was the data collected?
- What are the strengths of the statistics?
- What are the weaknesses of the statistics?

National Crime Victimization Survey

Study : (1) sample is approximately 50,000 randomly selected households, (2) participants are 12 years or older, (3) completed rape is sexual intercourse involving physical force or psychological coercion, including vaginal, anal, and oral penetration, including penetration by object, (4) sexual assault includes other forms of sexual behavior forced on the victim.

ii. Strengths: 1) broader definition of rape used

iii. Weaknesses: 1) uses only 2 questions to assess rape victimization, 2) excludes households without phones

b. National Violence Against Women Survey

i. Characteristics: 1) national sample of 8,000 women and 8,005 men, 2) ages 18 and older, 3) computer telephone interviews conducted, and 4) used 5 behaviorally specific questions

ii. Strengths: 1) separate questions for completed and attempted rape, 2) very descriptive questions, 3) used female interviewers with female respondents, 4) used bilingual interviewers,

iii. Weaknesses: 1) excludes individuals who are homeless, live in prisons or other group settings, and those without telephones, 2) lumped all AI/AN individuals in one category, 3) small number of victims from diverse racial/ethnic backgrounds.

Exercise: Critique of Rape Statistics

- In general, what are some of the best ways to use statistics in your programs?

c. Mohler-Kuo college study

- i. I. Characteristics: 1) 119 schools from Harvard School of Public Health College Study, 2) sample collected in 1997, 1999, 2001, 3) used a mail survey, 3) 3 questions to assess rape
- ii. Strengths: 1) use ii. Strengths: 1) definitions similar to those used in other studies, 2) large national sample of women
- iii. Weaknesses: 1) s iii. Weaknesses: 1) sexual intercourse is not defined, 2) not all forms of penetration, 3) self-report and non-response bias

d. Discussion: *What is the best way to use statistics in your programs?*

- a. Use studies that have samples most similar to your target audience.
- b. Provide information about the study and not only cite the percentages.
- c. Use data that has been replicated across studies.
- d. Do not single out one study, especially if the study is small.

Part II: The Truth (and Lies) About Evaluation

- How are you evaluating your programs?
- What are some of the challenges and problems that you face?

Objective of Presentation

- To improve understanding of evaluation and research design
- Identify and develop potential solutions for common challenges

What is Evaluation?

- It is research that focuses on the aims of a program and determines the degree to which goals are met (Hall & Hall, 2004).

Designs for Evaluation

(Weiss, 1998)

1) After-Only Design

- Collect data from participants only after received the program

2) Before-and-After Design

- Collect data before and after participants have received the program

DESIGN is very important to an evaluation because it helps determine whether what happens after a program is due to the program itself (Weiss, 1998). Are the effects due to the program or would they happen anyway?

Designs for Evaluation

3) After-Only with Control/Comparison Group

4) Before-and-After Designs with Control/Comparison Group

- This is the most common design

Control and Comparison Groups

- Both consist of individuals who do not receive the program
- Control groups
 - Uses random selection
- Comparison groups
 - Does not use random selection
 - Involves finding a group that has similar characteristics to the program group

COMPARISON GROUPS – Example: finding a school that is similar to the one receiving your program

Comparison groups help strengthen your ability to make inferences about the effects of your program. If the groups are different, you do not know if the effects are due to the program. Otherwise, the better scores may reflect some other factors, such as age, education, etc.

Measurement

- Many evaluations measure attitudes, values, knowledge, skills, and behavior (Weiss, 1998)
- Short-term versus long-term effects
- Definition of success

In the area of rape and sexual assault, what are some examples of:

- 1) Rape-related attitudes? = rape myth acceptance, victim blaming, rape empathy, sex role stereotyping, attitudes toward women
- 2) Rape-related values?
- 3) Knowledge? = knowledge about sexual assault
- 4) Skills? = resistance skills, assertiveness, communication
- 5) Behavior? = intentions to rape or engage in certain dating behaviors, rape awareness of behaviors, sexual aggressive behaviors

Ideally, measure both short-term and long-term effects to determine that changes persist over time.

Success is often defined by program staff and/or funding agency.

Measurement

- Interviews and questionnaires
 - Close-ended questions
 - Open-ended questions
 - Validity
 - Reliability
 - Sensitivity to change
 - Sensitivity to individual differences

Interviews and questionnaires should consist of good questions. Developing a questionnaire is complex, difficult, and takes time (Hall & Hall, 2—4).

CLOSE-ENDED QUESTIONS include answers to choose from. The answer choices should cover the whole range of responses that participants may have.

OPEN-ENDED QUESTIONS allow individuals to provide responses in their own words. This is ideal if you want suggestions or do not know the whole range of possible responses.

Assessing the QUALITY of an instrument involves assessing validity and reliability.

- 1) Validity is accuracy in measuring the concept of interest (example is assessing knowledge of cooking – you ask what times are the worst times to call for delivery – NOT valid question of cooking knowledge).
- 2) Reliability is consistency of responses over time. How you respond to a set of questions is relatively consistent over time.
- 3) Sensitivity to change. Want to avoid questions that will not change before and after receive the program (such as “is rape a bad behavior?”)
- 4) Sensitivity to individual differences. Want to avoid questions that most people select the same response (such as “is rape a bad behavior?” with yes/no answers instead of “what do you believe is the worse consequence of rape?” with economic costs, physical consequences among survivors, mental health impact among survivors, effects of family and communities).

Exercise: Evaluation of Rape Prevention Programs

Divide into two groups. Read the article assigned to your group. Answer the following questions:

- What did they measure?
- What questions/questionnaires did they use?
- How many times was the data collected?
- What are strengths of the evaluation design?
- What are the weaknesses of the evaluation design?

a. Anderson et al. (1998)

Data collected: rape-supportive attitudes

How it was collected: Rape Myth Acceptance Scale and Attitudes Toward Rape Scale-Revised

How many times data was collected: three times (pretest, posttest, and follow-up at 7 weeks after posttest)

Strengths: 1) random assignment of course sections, 2) control group, 3) short-term and long-term assessment, and 4) 4 weeks between pre-test and post-test

Weaknesses: 1) drop out by the follow-up assessment

Exercise: Evaluation of Rape Prevention Programs

- Are these evaluation designs practical in the real world?
- How do they compare to what you do?
- What challenges do you face?
- What are some possible solutions?

a. Avery-Leaf et al. (1997)

Data collected: attitudes justifying the use of dating violence

How it was collected: Conflict Tactics Scale, Justification of Interpersonal Violence questionnaire, Justification of Dating Jealousy and Violence scale, and Social Desirability scale.

How many times it was collected: two times (however, control group was assessed before the experimental group)

Strengths: 1) control group, 2) two assessment periods, 3) staggered the assessment so that the control group would not inadvertently be exposed to the program, and 4) random assignment of health classes

Weaknesses: 1) As indicated on page 15, some of the measures were not sensitive enough. From the start, several participants said that pushing, slapping, or punching were never acceptable. No room for this attitude to improve after receiving the program.

Conclusion

Two recommendations:

- 1) Use statistics wisely
- 2) Design evaluations so that they accurately examine the effects of your programs

The statistics on sanity are that one out of every four Americans is suffering from some form of mental illness.

Think of your three best friends. If they are okay, then it's you.

--Rita Mae Brown

Rita May Brown is a best-selling author.

References

- Anderson, L., Stoelb, M. P., Duggan, P., Hieger, B., Kling, K. H., & Payne, J. P. (1998). The effectiveness of two types of rape prevention programs in changing the rape-supportive attitudes of college students. *Journal of College Student Development, 39*, 131-142.
- Avery-Leaf, S., Cascardi, M., O'Leary, K. D., & Cano, A. (1997). Efficacy of a dating violence prevention program on attitudes justifying aggression. *Journal of Adolescent Health, 21*, 11-17.
- Bachman, R., & Saltzman, L. E. (1995). *Violence against women: Estimates from the redesigned survey*. Washington, DC: Bureau of Justice Statistics, National Institute of Justice.
- Catalano, S. M. (2005). *Criminal Victimization, 2004*. Washington, DC: Bureau of Justice Statistics, National Institute of Justice.

References

Federal Bureau of Investigation. (2005). *Crime in the United States 2004, Uniform Crime Reports*. Washington, DC: U.S. Department of Justice.

Hall, I., & Hall, D. (2004). *Evaluation and social research: Introducing small-scale practice*. New York, NY: Palgrave Macmillan.

Holmes, C. B. (1990). *The honest truth about lying with statistics*. Springfield, IL: Charles C. Thomas Publisher.

Kilpatrick, D. G. *Rape and sexual assault*. Retrieved April 20, 2006, from <http://www.musc.edu/vawprevention/research/sa.shtml>

References

- Kilpatrick, D. G., Edmunds, C., & Seymour, A. (1992). *Rape in America: A report to the nation*. Charleston, SC: National Victim Center & the Crime Victims Research and Treatment Center, Medical University of South Carolina.
- Kilpatrick, D. G., & Saunders, B. E. (1996). *Prevalence and consequences of child victimization: Results from the National Survey of Adolescents*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.
- Mohler-Kuo, M., Dowdall, G. W., Koss, M. P., & Wechsler, H. (2004). Correlates of rape while intoxicated in a national sample of college women. *Journal of Studies on Alcohol, 65*, 37-45.
- Tjaden, P., & Thoennes, N. (2006). *Extent, nature, and consequences of rape victimization: Findings from the National Violence Against Women Survey*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Weiss, C. H. (1998). *Evaluation, 2nd edition*. Upper Saddle River, NJ: Prentice-Hall, Inc.