Chapter 12 Supplement: Misuses of Statistics

Every day we hear or read the results obtained from research studies, surveys, and opinion polls regarding our health, our opinions, and our behavior. Statistical research is used to provide knowledge and information to help us make intelligent decisions about our health and welfare. But there is another aspect of statistics that needs to be addressed: the fact that there are people who will misuse statistics to sell us products that don’t work, to attempt to prove something that is not true, or to get our attention by using such techniques as fear, shock, and outrage supported by bad statistics.

There are some sayings that have been around for a long time that illustrate this point—for example,

“There are three types of lies—lies, damn lies, and statistics.”

“Figures don’t lie, but liars figure.”

Just because we read or hear the results of a research study or an opinion poll in the media, this does not mean that these results are reliable or that they can be applied to any and all situations. For example, reporters sometimes leave out critical details such as the size of the sample used or how the research subjects were selected. Without this information, you can’t properly evaluate the research and properly interpret the conclusions of the study or survey.

The purpose of this section is to present some ways that statistics can be misused. The point isn’t to imply that all research studies and surveys are suspect, but that there are many factors to consider when making decisions based on the results of research studies and surveys. Here are some ways that statistics can be misrepresented.

**Suspect Samples**

The first thing to consider is the sample that was used in the research study. Sometimes researchers use very small samples to obtain information. Several years ago, advertisements contained statements such as, “Three out of four doctors surveyed recommend brand such and such.” If only four doctors were surveyed, the results could have been obtained by chance alone; but if 100 doctors were surveyed, the results would be a little more meaningful.

Not only is it important to have a sample size that is large enough, but it is also necessary to see how the subjects in the sample were selected. Studies using volunteers sometimes have a built-in bias because volunteers generally don’t represent the population at large. Sometimes they are recruited from a particular socioeconomic background, and sometimes unemployed people volunteer for research studies in order to get paid. Studies that require the subjects to spend several days or weeks in an environment other than their home or workplace automatically exclude people who are employed and cannot take time off from work. Sometimes college students or retirees are used. In the past, many studies have used only men but have attempted to generalize the results to both men and women. Opinion polls that require a person to phone or mail in a response most often are not representative of the population in general since only those with strong feelings for or against the issue usually call or respond by mail.

Another type of sample that may not be representative is the convenience sample. Educational studies sometimes use students in intact classrooms since it’s convenient.
Quite often, the students in these classrooms don’t represent the student population of the entire school.

In interpreting results from studies using small samples, convenience samples, or volunteer samples, care should be used when generalizing the results to the entire population.

**Ambiguous Averages**

In this chapter, four statistical measures are loosely called “averages.” They are the mean, the median, the mode, and the midrange, and for a given set of data, the values of these averages sometimes differ quite a lot. People who know that there are several types of average can, without lying, select the one for the data that most lends evidence to support their position. For example, suppose a storeowner employs four salespeople. The number of years each has been employed by the store is 22, 10, 2, and 2. The mean number of years’ service is 9. The median is 6, and the mode is 2. Now if the storeowner wanted to advertise the fact that his employees have many years of experience, which average do you think he would use? Obviously, he would use 9. However, 9 is not very high, so since the owner sometimes doubles as a salesman and since he has owned the store for 42 years, he includes his years of service in computing the average. The mean is now 15.6 years. Quite impressive, isn’t it? (Actually the midrange, 22 years, is much more impressive.) Whenever the word “average” is used instead of mean, median, mode, or midrange, ask yourself, “What average is being used?”

Another type of statistical distortion can occur when different values are used to represent the same data. For example, one political candidate who is running for reelection might say, “During my administration, expenditures increased a mere 3%.” His opponent, who is trying to unseat him, might say, “During my opponent’s administration, expenditures have increased a whopping $6,000,000.” Here both figures are correct; however, expressing $6,000,000 as a mere 3% makes it seem like a very small increase whereas expressing a 3% increase as $6,000,000 makes it sound like a very large increase. Here again, ask yourself, “Which measure best represents the data?”

**Detached Statistics**

A claim that uses a detached statistic is one in which no comparison is made. For example, you may hear a claim such as, “Our brand of crackers has \( \frac{1}{3} \) fewer calories.” Here, no comparison is made. One-third fewer calories than what? Someone else’s crackers? A bacon double cheeseburger? Another example is a claim that uses a detached statistic such as, “Brand A aspirin works four times faster.” Four times faster than what? When you see statements such as this, always ask yourself, “As compared to what?”

**Implied Connections**

Many claims attempt to imply connections between variables that may not actually exist. For example, consider this statement: “Eating fish may help to reduce your cholesterol.” Notice the words “may help.” There is no guarantee that eating fish will definitely help you reduce your cholesterol.

“Studies suggest that using our exercise machine will reduce your weight.” Here the word *suggest* is used, and again, there is no guarantee that you will lose weight using the exercise machine advertised.

Another claim might say, “Taking calcium will lower blood pressure in some people.” Notice the word *some* is used. You may not be included in the group of “some” people. Be careful when drawing conclusions from claims that use words such as “may,” “in some people,” “might help,” etc.
Misleading Graphs

Graphs give a visual representation that enables readers to analyze and interpret data more easily than they could simply by looking at numbers; however, inappropriately drawn graphs can misrepresent the data and lead the reader to false conclusions. For example, a car manufacturer's ad stated that 98% of the vehicles it had sold in the past 10 years were still on the road. The ad then showed a graph similar to the one in Figure 12-29(a). The graph shows the percentage of the manufacturer's automobiles still on the road and the percentage of its competitors' automobiles still on the road. Is there a large difference? Not necessarily. Notice the scale on the vertical axis in Figure 12-29a. It has been cut off (or truncated) so that it starts at 95%. When the graph is redrawn using a scale that goes from 0% to 100%, as in Figure 12-29b, there is hardly a noticeable difference in the percentages. Changing the units at the starting point on the axis can convey a very different visual representation of the data.

It isn’t necessarily wrong to truncate an axis of the graph; many times it is necessary to do so. But the reader should be aware of this fact and interpret the graph accordingly. Don’t be misled if an inappropriate explanation is given.

Let’s consider another example. The percentage of the world’s total motor vehicles produced by manufacturers in the United States declined from 19.7% in 2001 to 16% in 2006, as shown by these data.

<table>
<thead>
<tr>
<th>Year</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent produced in the U.S.</td>
<td>19.7</td>
<td>20.7</td>
<td>19.7</td>
<td>18.7</td>
<td>17.7</td>
<td>16.0</td>
</tr>
</tbody>
</table>

If you draw the graph as shown in Figure 12-30a on the next page, with a scale ranging from 0% to 100%, the decrease looks pretty mild. But if you use a scale ranging from 15% to 25% as in Figure 12-30b, it looks like the U.S. auto industry will be lucky to survive another two years. Again, the choice of scale on the vertical axis has a big effect on how the information is perceived.

Another misleading graphing technique sometimes used is exaggerating a one-dimensional increase by showing it in two dimensions. For example, the average cost of a 30-second Super Bowl commercial has increased from $40,000 in 1967 to $2.5 million in 2007. The increase shown by the bar graph in Figure 12-31a on the next page represents the change by a comparison of the heights of two bars in one dimension. The same data are shown two-dimensionally with circles in Figure 12-31b. Notice that...
Faulty Survey Questions

Surveys and opinion polls obtain information by using questionnaires. There are two types of studies: interviews, and self-administered questionnaires. The interview survey requires a person to ask the questions either in person or by telephone. Self-administered questionnaire surveys require the participant to answer the questions by mail, computer, or in a group setting, such as a classroom. When analyzing the results of a survey using questionnaires, you should be sure that the questions are properly written, since the way questions are phrased can often influence the way people answer them. For example, when a group of people was asked, “Do you favor a waiting period before guns are sold?” 91% of the respondents said “Yes” and 6% said “No.” However, when the question was rephrased as, “Do you favor a national gun-registration program costing about 20% of all dollars spent on crime control?” 37% responded “Yes” and 61% responded “No.” As you can see, although the questions pertain to some form of gun control, each asks something a little different, and the responses are radically different. When reading and interpreting the results obtained from questionnaire surveys, watch out for some of these common mistakes made in the writing of the survey questions.
Asking Biased Questions. By asking a question in a certain way, the researcher can lead the respondents to answer the question the way the researcher wants them to. For example, the question, “Are you going to vote for Candidate Jones, even though the latest survey shows he will lose the election?” may lead the respondent to say, “No” since many people do not want to vote for a loser or admit that they have voted for a loser.

Using Confusing Words. Using words in a survey question that are not well defined or understood can invalidate the responses. For example, a question such as, “Do you think people would live longer if they were on a diet?” would mean many different things to people since there are many types of diets, such as low-salt diets, high-protein diets, etc.

Asking Double-Barreled Questions. Sometimes two ideas are contained in one question, and the respondent may answer one or the other in his or her response. For example, consider the question, “Are you in favor of a national health program and do you think it should be subsidized by a special tax as opposed to other ways to finance it, such as a national lottery?” Here the respondent is really answering two questions.

Using Double Negatives. Survey questions containing double negatives often confuse the respondent. For example, what is the question “Do you feel that it is not appropriate to have areas where people cannot smoke?” really asking?

Other factors that could bias a survey would include anonymity of the participant, the time and place of the survey, and whether the questions were open-ended or closed-ended.

Participants will, in some cases, respond differently to questions based on whether or not their identity is known. This is especially true if the questions concern sensitive issues such as income, sexuality, abortion, etc. Researchers try to ensure confidentiality rather than anonymity; however, many people will be suspicious in either case.

The time and place where a survey is taken can influence the results. For example, if a survey on airline safety is taken right after a major airline crash, the results may differ from those obtained in a year with no major airline disasters.

To restate the premise of this section, statistics, when used properly, can be beneficial in obtaining much information, but when used improperly, they can lead to much misinformation. Therefore, it is important to understand the concepts of statistics and use them correctly.

**EXERCISE SET**

1. According to a pilot study of 20 people conducted at the University of Minnesota, daily doses of a compound called arabinogalactan over a period of 6 months resulted in a significant increase in the beneficial lactobacillus species of bacteria. Why can’t it be concluded that the compound is beneficial for the majority of people?

2. Comment on this statement taken from a magazine advertisement: “In a recent clinical study, Brand ABC* was proved to be 1,950 percent better than creatine!”

3. In an ad for women, the following statement was made: “For every hundred women, 91 have taken the road less traveled.” Comment on this statistic.

4. In many ads for weight-loss products, under the product claims and in small print, the following statement is made: “These results are not typical.” What does this say about the product being advertised?

5. An article in a leading magazine stated that, “When 18 people with chronic, daily, whip-lash related headaches received steroid injections in a specific...
neck joint, 11% had no more headaches.” Think of a possible reason why the figure 11% was used.
6. In an ad for moisturizing lotion, the following claim is made: “... it’s the #1 dermatologist recommended brand.” What is misleading about the claim?
7. An ad for an exercise product stated that, “Using this product will burn 74% more calories.” What is misleading about that statement?
8. “Vitamin E is a proven antioxidant and may help in fighting cancer and heart disease.” Is there anything ambiguous about this claim?
9. “Just one capsule of Brand X* can provide 24 hours of acid control.” What needs to be more clearly defined in this statement?
   
   *Actual brand will not be named.
10. “... male children born to women who smoke during pregnancy run a risk of violent and criminal behavior that lasts well into adulthood.” Can we infer that smoking during pregnancy is responsible for criminal behavior in people?

For Exercises 11–13, explain why each pair of graphs is misleading.

11. A company advertises that its brand of energy pills gets into the user’s blood faster than a competitor’s brand and shows these two graphs to prove its claim.

12. The graph shows the difference in sales of pumpkins during October for the years 1999 and 2009.

13. Explain this contradiction: The two graphs were drawn using the same data, yet the first graph shows sales remaining stable, and the second graph shows sales increasing dramatically.

For Exercises 14–16, explain why each survey question might lead to an erroneous conclusion.

14. “How often do you run red lights?”
15. “Do you think gun manufacturers should put safety locks on all guns sold even though it would increase the cost of the gun by 20%?”
16. “Do you think that it is not important to give extra tutoring to students who are not failing?”
17. The results of a survey reported in USA Weekend stated that:
   “9% would drive through a toll booth without paying.”
   “13% would steal cable television or inflate their resumes.”
   Explain why these figures may not be representative of the population in general.
18. In an article in *USA Weekend*, this statement was made: “More serious seems to be coffee’s potential to raise blood pressure levels of homocysteine, a protein that promotes artery clogging. A recent Norwegian study found 20% higher homocysteine in heavy coffee drinkers (more than 9 cups a day) than in non-coffee drinkers.” Based on this statement, should we give up our daily cup of coffee?

19. An article in a newspaper with the headline, “Lead: The Silent Killer” listed the number of confirmed childhood lead poisoning cases in Allegheny County, Pennsylvania, in 1985 as 15 and in 1997 as 124. Can you conclude that the incidence of childhood lead poisoning cases is increasing in Allegheny County? Suggest a factor that might cause an increase in reported lead poisoning in children.

Source: *Pittsburgh Tribune Review*

20. In a recent article, the author states that 71% of adults do not use sunscreen. Although 71% is a large percentage, explain why it could be misleading.

21. In a book on probabilities, the author states that in the United States every 20 minutes, on average, someone is murdered. Based on this statement, can we conclude that crime is rampant in the United States? *Note*: The population of the United States is more than 300 million people.

22. For a specific year, there were 6,067 male fatalities in the workplace and 521 female deaths. A government official made this statement: “Over 90 percent of the fatal injuries the past year were men, although men accounted for only 54 percent of the nation’s employment.” Can we conclude that women are more careful on the job?