

1 The Art and Science of learn form data

Luis Talavera

2022-12-01

Exercises

1.28 UW Student survey In a University of Wisconsin (UW) study about alcohol abuse among students, 100 of the 40,858 members of the student body in Madison were sampled and asked to complete a questionnaire. One question asked was, “On how many days in the past week did you consume at least one alcoholic drink?”

- Identify the population and the sample.
- For the 40,858 students at UW, one characteristic of interest was the percentage who would respond “zero” to this question. For the 100 students sampled, suppose 29% gave this response. Does this mean that 29% of the entire population of UW students would make this response? Explain.
- Is the numerical summary of 29% a sample statistic or a population parameter?

Ans.

- Population: Student body in Madison. Sample: 100 students of student body in Madison
- No. A sample of 100 is not enough to generalize conclusions about the population.
- A sample statistic.

1.29 Euthanasia The General Social Survey asked, in 2012, whether you would commit suicide if you had an incurable disease. Of the 3112 people who had an opinion about this, 1862, or 59.8%, would commit suicide.

- Describe the population of interest.
- Explain how the sample data are summarized using descriptive statistics.
- For what population parameter might we want to make an inference?

Ans.

- All American adults
- The percentage of people who will commit suicide.
- The population proportion who would commit suicide

1.30 Mobile data costs A study is conducted by the Australian Communications and Media Authority. Based on a small sample of 19 mobile communications plans offered, the average cost per 1000 MB of free monthly mobile data allowance is found to be \$5.40, with a margin of error of \$2.16. Explain how this margin of error provides an inferential statistical analysis.

Ans. The margin of error provides a range of accepted communications plans cost. Future plans releases costs should be between \$5.4 +-\$2.16

1.31 Breaking down Brown versus Whitman Example 2 of this chapter discusses an exit poll taken during the 2010 California gubernatorial election. The administrators of the poll also collected demographic data, which allows for further breakdown of the 3889 voters from whom information was collected. Of the 1633 voters registered as Democrats, 91% voted for Brown, with a margin of error of 1.4%. Of the 1206 voters registered as Republicans, 10% voted for Brown, with a margin of error of 1.7%. And of the 1050 Independent voters, 42% voted for Brown, with a margin of error of 3.0%.

- Do these results summarize sample data or population data?
- Identify a descriptive aspect of the results.
- Identify an inferential aspect of the results.

Ans.

- The results summarize sample data.
- Percentages reported
- Margin of error reported.

1.32 Online learning Your university is interested in determining the proportion of students who would be interested in completing summer courses online, compared to on campus. A survey is taken of 100 students who intend to take summer courses.

- Identify the sample and the population.
- For the study, explain the purpose of using (i) descriptive statistics and (ii) inferential statistics.

Ans. a. Sample: 100 students who intend to take summer courses. Population: Students enrolled at university. b.

- How many students prefer completing courses online and how many students prefer completing courses on campus.
- How many students will take the summer courses if those are online.

1.33 Marketing study For the marketing study about sales in Example 5, identify the (a) sample and population and (b) descriptive and inferential aspects.

Ans.

- Population: all customers; Sample: 1000 customers
- Descriptive: average sales per person in sample; Inferential: estimate of average sales per person in population.

1.34 Support of labor unions The Gallup organization has asked opinions about support of labor unions since its first poll in 1936, when 72% of the American population approved of them. In its 2014 poll, it found that support of labor unions had fallen to 53% of Americans, based on a sample of 1,540 adults.

- Calculate an estimated margin of error for these data.
- What is the range of likely values for Americans who support labor unions in 2014?
- This analysis is an example of i. descriptive statistics ii. inferential statistics iii. a data file iv. designing a study

Ans.

- 2.54%

- b. 50.46% - 55.54%
- c. Inferential statistics

1.35 Multiple choice: Use of inferential statistics? Inferential statistics are used

- a. to describe whether a sample has more females or males.
- b. to reduce a data file to easily understood summaries.
- c. to make predictions about populations by using sample data.
- d. when we can't use statistical software to analyze data.
- e. to predict the sample data we will get when we know the population

Ans. c

1.36 True or false? In a particular study, you could use descriptive statistics, or you could use inferential statistics, but you would rarely need to use both.

Ans. False, usually you first use descriptive statistics in a study and after that you use inferential statistics.

1.37 Statistics in the news Pick up a recent issue of a national newspaper, such as The New York Times or USA Today, or consult a news website, such as msnbc.com or cnn.com. Identify an article that used statistical methods. Did it use descriptive statistics, inferential statistics, or both? Explain

Ans. The article "The Rams super bowl afterparty turned into a historic hangover" uses descriptive statistics to show how the effectiveness of the offense and defense of the reigning NFL champion Los Angeles Rams has changed. It points out that the defense has maintained the same statistics and that even the special teams have improved compared to last year, but on the other hand, the offense has only worsened its numbers compared to last year. It is descriptive statistics because descriptive statistics summarize and show information of the phenomenon in this case the performance of the team.

1.38 What is statistics? On a final exam that one of us recently gave, students were asked, "How would you define 'statistics' to someone who has never taken a statistics course?" One student wrote, "You want to know the answer to some question. There's no answer in the back of a book. You collect some data. Statistics is the body of procedures that helps you analyze the data to figure out the answer and how sure you can be about it." Pick a question that interests you and explain how you might be able to use statistics to investigate the answer.

Ans. How likely are Cincinnati Bengals to make playoffs? I first use current season and previous season data to analyze the probability of winning against their remaining rivals, a brief summary about their statistics in winning seasons vs losing seasons.

1.39 Surprising suicide data? In Exercise 1.29, of 3112 people who responded, 59.8% or 1862 people, said they would commit suicide if they had an incurable disease. Suppose that 50% of the entire population shares this view about suicide. If the sample of 3112 people were a random sample, would this sample proportion of 0.598 be surprising? Investigate, using the web app described in Activity 2, setting the samples size to 3112 and the population proportion to 0.5.

Ans. The likelihood of getting this result is extremely small.