1.2 Sample versus Population

Luis Talavera

2022-11-27

Exercises

1.6 Description and inference a. Distinguish between description and inference as reasons for using statistics. Illustrate the distinction using an example. b. You have data for a population, such as obtained in a census. Explain why descriptive statistics are helpful but inferential statistics are not needed.

Ans.

a. We use descriptive statistics when we want to summarize and illustrate features of our data. By the other hand, inferential statistics is used when we need to make a prediction or use estimate population characteristics based on sample data. For example, we can collect data about school education levels and summarized how many people have a high school diploma, drop college, have a bachelor's degree or PhD, these data can be plotted in a bar plot, that is descriptive statistics; inferential statistics will look the data to make conclusions.

b. Descriptive statistics will help us summarizing the characteristics of the census answers, giving us plots easy to understand and letting us know the big picture about the sample. Inferential is not necessary as long as we don't want to make conclusiones about related topics in the census.

1.7 Ebook use During the spring semester in 2014, an ebook survey was administered to students of Winthrop University. Of the 170 students sampled, 45% indicated that they had used ebooks for their academic work. Identify (a) the sample, (b) the population, and (c) the sample statistic reported.

Ans.

a. 170
b. All registered students in Winthrop University
c. 45%

1.8 Concerned about global warming? The Institute for Public Opinion Research at Florida International University has conducted the FIU/Florida Poll of about 1200 Floridians annually since 1988 to track opinions on a wide variety of issues. In 2006 the poll asked, "How concerned are you about the problem of global warming?" The possible responses were very concerned, somewhat concerned, not very concerned, and haven't heard about it. The poll reported percentages (44, 30, 21, 6) in these categories. a. Identify the sample and the population. b. Are the percentages quoted statistics or parameters? Why?

Ans.

a. Sample: 1200, Population: People living in Florida.

b. The percentages are statistics because they are summaries of a sample taken from a population.

1.9 Preferred holiday destination Suppose a student is interested in knowing the preferred holiday destinations of the faculty members in his university. He is affiliated to the college of business and interviews a few of the faculty members of this college about their preferred holiday destination. In this context, identify the (a) subject, (b) sample, and (c) population.

Ans.

a. Each faculty member in the college of business.
b. The faculty members interviewed.
c. All faculty members in the college of business.

1.10 Is globalization good? The Voice of the People poll asks a random sample of citizens from different countries around the world their opinion about global issues. Included in the list of questions is whether they feel that globalization helps their country. The reported results were combined by regions. In the most recent poll, 74% of Africans felt globalization helps their country, whereas 38% of North Americans believe it helps their country. (a) Identify the samples and the populations. (b) Are these percentages sample statistics or population parameters? Explain your answer.

Ans.

a. Samples: Random citizens from different countries. Populations: People living in a country.
b. Statistics because they are numerical summaries of a sample taken from a population.

1.11 Graduating seniors' salaries The job placement center at your school surveys all graduating seniors at the school. Their report about the survey provides numerical summaries such as the average starting salary and the percentage of students earning more than $30,000 a year. a. Are these statistical analyses descriptive or inferential? Explain. b. Are these numerical summaries better characterized as statistics or as parameters?

Ans.

a. The statistic analysis is descriptive, the reason is that is describing characteristics about the data.
b. Parameters because they are numerical summaries of all **population**.

1.12 At what age did women marry? A historian wants to estimate the average age at marriage of women in New England in the early 19th century. Within her state archives she finds marriage records for the years 1800– 1820, which she treats as a sample of all marriage records from the early 19th century. The average age of the women in the records is 24.1 years. Using the appropriate statistical method, she estimates that the average age of brides in early 19th-century New England was between 23.5 and 24.7. a. Which part of this example gives a descriptive summary of the data? b. Which part of this example draws an inference about a population? c. To what population does the inference in part b refer? d. The average age of the sample was 24.1 years. Is 24.1 a statistic or a parameter?

Ans.

a. "The average age of the women in the records is 24.1 years"
b. "she estimates that the average age of brides in early 19th-century New England was between 23.5 and 24.7."
c. All married woman in 19th century.

d. 24.1 is a statistic. The numerical summary is of a sample taken from a population.

1.13 Age pyramids as descriptive statistics The figure shown is a graph published by Statistics Sweden. It compares Swedish society in 1750 and in 2010 on the numbers of men and women of various ages, using age pyramids.
Explain how this indicates that
a. In 1750, few Swedish people were old.
b. In 2010, Sweden had many more people than in 1750.
c. In 2010, of those who were very old, more were female than male.
d. In 2010, the largest five-year group included people born during the era of first manned space flight.

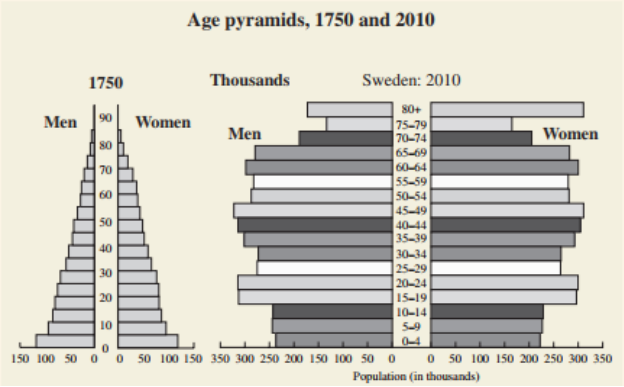

Figure 1: Sweden Age Pyramids

Ans.

a. The top of the pyramid are small bars, the bar length indicates the people which belong to that age range
b. The maximum bars' length in 2010 is greater than 1750 in all age ranges
c. The top bar indicates the very old people, females' bar is greater than males' bar.
d. From the plot, the largest bars are from people between 45-49.

1.14 Gallup polls Go to the website www.galluppoll.com for the Gallup poll. From reports listed on its home page, give an example of (a) a descriptive statistical analysis and (b) an inferential statistical analysis.

Ans.

a. 21% of sampled people is satisfied with the way things are going in the U.S.
b. Between 41% and 46% of U.S. adults approve of the way the president is handling

1.15 Graduate studies Consider the population of all undergraduate students at your university. A certain proportion have an interest in graduate studies. Your friend randomly samples 40 undergraduates and uses the sample proportion of those who have an interest in graduate studies to predict the population proportion at the university. You conduct an independent study of a random sample of 40 undergraduates and find the sample proportion of those are interested in graduate studies. For these two statistical studies,
a. Are the populations the same?
b. How likely is it that the samples are the same? Explain.
c. How likely is it that the sample proportions are equal? Explain.

Ans.

a. Yes, they are the same
b. Since the sample were randomly taken, is very unlikely that they are the same.
c. Thanks to randomness the variability of sample proportions will be small.

1.16 Samples vary less with more data We'll see that the amount by which statistics vary from sample to sample always depends on the sample size. This important fact can be illustrated by thinking about what would happen in repeated flips of a fair coin.
a. Which case would you find more surprising—flipping the coin five times and observing all heads or flipping the coin 500 times and observing all heads?
b. Imagine flipping the coin 500 times, recording the proportion of heads observed, and repeating this experiment many times to get an idea of how much the proportion tends to vary from one sequence to another. Different sequences of 500 flips tend to result in proportions of heads observed which are less variable than the proportion of heads observed in sequences of only five flips each. Using part a, explain why you would expect this to be true.

Ans.

a. The second since both heads and tails have the same probability.
b. The approximated margin error will be smaller in 500 flips sequences than 5 flips sequences

1.17 Comparing polls The following table shows the result of the 2012 presidential election along with the vote predicted by several organizations in the days before the election. The sample sizes were typically about 1000 to 2000 people. The percentages for each poll do not sum to 100 because of voters who indicated they were undecided or preferred another candidate.

	Predicted Vote	
Poll	**Obama**	**Romney**
Rasmussen	48	49
CNN	49	49
Gallup	48	49
Pew Research	50	47
Rand	54	43
Fox	46	46
Actual Vote	50.6	47.8

Source: http://ncpp.org/files/Presidential%20National%20Polls%202012%200103%20Full.pdf

Figure 2: Sweden Age Pyramids

a. Treating the sample sizes as 1000 each, find the approximate margin of error.

b. Do most of the predictions fall within the margin of error of the actual vote percentages? Considering the relative sizes of the sample, the population, and the undecided factor, would you say that these polls had good accuracy?

Ans.

a.

$$\frac{1}{\sqrt{1000}} \times 100\% = 3.2\%$$

b. Sure, most of the predictions fall within margin of error therefore We can consider them as accurate predictions.

1.18 Margin of error and n A poll was conducted by Ipsos Public Affairs for Global News. It was conducted between May 18 and May 20, 2016, with a sample of 1005 Canadians. It was found 73% of the respondents "agree" that the Liberals should not make any changes to the country's voting system without a national referendum first. (http://globalnews.ca/news/). Find the approximate margin of error if the poll had been based on a sample of size (a) n = 900, (b) n = 1600, and (c) n = 2500. Explain how the margin of error changes as n increases.

Ans.

a.
$$\frac{1}{\sqrt{900}} \times 100\% = 3.33\%$$

b.
$$\frac{1}{\sqrt{1600}} \times 100\% = 2.5\%$$

c.
$$\frac{1}{\sqrt{2500}} \times 100\% = 2\%$$

As long as n increases the approximate margin of error decreases.

1.19 Smoking cessation A study published in 2010 in The New England Journal of Medicine investigated the effect of financial incentives on smoking cessation. As part of the study, 878 employees of a company, all of whom were smokers, were randomly assigned to one of two treatment groups. One group (442 employees) was to receive information about smoking cessation programs; the other (436 employees) was to receive that same information as well as a financial incentive to quit smoking. The outcome of interest of the study was smoking cessation status six months after the initial cessation was reported. After implementation of the program, 14.7% of individuals in the financial incentive group reported cessation six months after the initial report, compared to 5.0% of the information-only group. Assume that the observed difference in cessation rates between the groups (14.7% - 5.0% = 9.7%) is statistically significant.

a. What does it mean to be statistically significant? (choose the best option from (i)–(iv))
   i. The financial option was offered to 9.7% more smokers in the study than the nonsmokers who were employees of the company.
   ii. 9.7% was calculated using statistical techniques.
   iii. If there were no true impact of the financial incentive, the observed difference of 9.7% is unlikely to have occurred by chance alone.
   iv. We know that if the financial incentive were given to all smokers, 9.7% would quit smoking.

b. Is the difference between the groups attributable to the financial incentive?

Ans.

a. iii
b. Yes. Because the workers were separated by groups randomly.