Statistics Lecture Notes

Statistics Lecture Notes

Jason D. Underdown
SLCC

May 16, 2019 15:13:39 (-06:00)

Website: Statistics Lecture Notes

Preface

As the course progresses, I will add the example problems that I present in class at the chalkboard to these lecture notes.

Studying these lecture notes alone is usually not sufficient to achieve an A grade in the course. You should prepare for each lecture by reading the section we will cover each day before coming to class. Do not feel you need to read each section extremely carefully. If you get stuck while reading, spend a few minutes trying to figure out your misunderstanding, but if you can't figure it out, make a mental note of your question (or better yet write it down) and then continue on in the section. Often when reading mathematics, you will find that your question gets answered a few sentences or paragraphs later.

Contents

Part I

Getting the Information You Need

Chapter 1

Data Collection

1.1 Introduction to the Practice of Statistics

Definition 1.1.1 statistics. how to collect, organize, summarize and analyze information so that conclusions can be drawn with a measure of confidence. ◊

Definition 1.1.2 descriptive stats. entails organizing and summarizing collected data. ◊

Definition 1.1.3 inferential stats. methods for taking sample results and extending it to the population with some measure of reliability. ◊

Definition 1.1.4 population. entire group of interest in a study ◊

Definition 1.1.5 individual. a single person or object being studied ◊

Definition 1.1.6 sample. subset of the population being studied ◊

Definition 1.1.7 statistic. numerical summary of a sample ◊

Definition 1.1.8 parameter. numerical summary of a population ◊

Definition 1.1.9 variables. characteristics of individuals in the population ◊

Definition 1.1.10 qualitative variable. responses based on attributes or characteristics ◊

Definition 1.1.11 quantitative variable. responses are numerical measures, with a natural ordering i.e. it makes sense to compare via $<$ or $>$ ◊

Definition 1.1.12 discrete variable. finite or countable number of possible values e.g. number of children ◊

Definition 1.1.13 continuous variable. infinite number of possible values e.g. height ◊

Statistic or Parameter?

For each description which follows, determine whether the numerical quantity is a *statistic* or *parameter*.

1. In a study of all 2223 passengers aboard the Titanic, it is found that 706 survived when it sank. **Answer**. parameter

2. A recent survey of a sample of MBAs reported that the average salary for an MBA is more than $82,000. (Source: The Wall Street Journal) **Answer**. statistic

3. Starting salaries for all of the 667 MBA graduates from the University of Chicago Graduate School of Business increased 8.5% from the previous year. **Answer**. parameter

4. In a recent poll of Salt Lake Community College students, 83% of students owned a vehicle. **Answer**. statistic

5. In a recent survey of 457 attendees of Jumanji, 402 would recommend the movie to a friend. **Answer**. parameter, unless you consider the population to be all people who have seen Jumanji

Discrete or Continuous Variable?

1. In a survey of 1059 adults, it is found that 39% of them have guns in their homes **Answer**. discrete

2. The number of heads obtained after flipping a coin five times **Answer**. discrete

3. The distance Tiger Woods can drive a golf ball **Answer**. continuous

4. Points scored in a basketball game **Answer**. discrete

5. Volume of water lost each day from a leaky faucet **Answer**. continuous

6. Length of a song **Answer**. continuous, but could be discrete if measured in seconds

7. Number of words in a song **Answer**. discrete

Qualitative or Quantitative Variable?

1. Gender **Answer**. qualitative

2. Temperature **Answer**. quantitative

3. Nation of origin **Answer**. qualitative

4. Number of siblings **Answer**. quantitative

5. Number of days **Answer**. quantitative

6. Grams of carbohydrates in a donut **Answer**. quantitative

7. Phone number **Answer**. qualitative

8. Value of a house **Answer**. quantitative

9. Zip code **Answer**. qualitative

2011 BMW Models

The following information relates to the 2011 model year product line of BMW automobiles. Identify the individuals in the study, the variables and if the data corresponding to the variables are qualitative or quantitative, and continuous or discrete.

Model	Body Style	Weight	Number of Seats
3 Series	Coupe	3362	4
5 Series	Sedan	4056	5
6 Series	Convertible	4277	4
7 Series	Sedan	4564	5
X3	Sport Utility	4012	5
Z4	Coupe	3505	2

Table 1.1.14: 2011 BMW Automobiles

1. Individuals: **Answer**. 3 Series, 5 Series, 6 Series, 7 Series, X3, Z4

2. Variable(s): **Answer**.

Variable	Qual./Quant.	Disc./Cont.
body style	qualitative	N/A
weight	quantitative	continuous
number of seats	quantitative	discrete

Smoker's IQ

A study was conducted in which 20,211 18--year old recruits were given an exam to measure IQ. In addition, the recruits were asked to disclose their smoking status. An individual was considered a smoker if he smoked at least 1 cigarette per day. The goal of the study was to determine whether adolescents aged 18 to 21 who smoke have a lower IQ than nonsmokers. It was found that the average IQ of the smokers was 94, while the average IQ of nonsmokers was 101. The researchers concluded that lower IQ individuals are more likely to chose to smoke, not that smoking makes people less intelligent.

1. What is the population being studied? **Answer**. adolescents aged 18 to 21

2. What is the sample? **Answer**. 20,211 18 year old recruits

3. What are the descriptive statistics? **Answer**. IQ, smoking status

1.2 Observational Studies Vs. Designed Experiments

Definition 1.2.1 response variable. The response variable is the focus of a question in a study or experiment. ◊

Definition 1.2.2 explanatory variable. An explanatory variable is one that explains changes in the response variable. ◊

Definition 1.2.3 observational study. measuring response variable without influencing the values of either the response variable or explanatory variable ◊

Definition 1.2.4 cross-sectional studies. observational studies that collect data at one particular time ◊

Definition 1.2.5 case-control studies. observational studies that require individuals to look back in time to record measurements (retrospective) ◊

Definition 1.2.6 cohort studies. observational studies that track individuals repeatedly over time (prospective) ◊

Definition 1.2.7 designed experiment. measuring response variable after intentionally changing the value of the explanatory variable ◊

Definition 1.2.8 confounding. in a study occurs when the effects of two or more explanatory variables are not separated. Therefore, any relation that may exist between an explanatory variable and the response variable may be due to some other variable or variables not accounted for in the study ◊

Definition 1.2.9 lurking variable. a variable that was not considered in the study, but affects the response variable ◊

IMPORTANT COMMENT: Because of lurking variables, observational studies *DO NOT* allow a researcher to claim causation, only *association*.

Definition 1.2.10 confounding variable. a variable in the study where you can't tell the difference between its effect and another explanatory variable's effect ◊

Definition 1.2.11 census. list of information from all individuals in a population ◊

Observational Study or Designed Experiment?

1. Rats with cancer are divided into two groups. One group received 5 milligrams of a medication that is thought to fight cancer, and the other received 10 milligrams. After 2 years, the spread of cancer is measured. **Answer**. designed experiment
2. Conservation agents netted 250 large-mouth bass in a lake and determined how many were carrying parasites. **Answer**. observational study
3. Seventh grade students are randomly divided into two groups. One group is taught math using traditional techniques; the other is taught math using a reform method. After 1 year, each group is given an achievement test to compare proficiency. **Answer**. designed experiment
4. A survey was conducted asking 400 people, "Do you prefer Coke or Pepsi?" **Answer**. observational study

Daily Coffee Consumption

Is there an association between daily coffee consumption and the occurrence of skin cancer? Reasearchers asked 93,676 women to disclose their coffee-drinking habits and also determined which of the women had nonmelanoma skin cancer. The researchers concluded that consumption of 6 or more cups of caffeinated coffee per day was associated with a reduction in nonmelanoma skin cancer.

1. Whas this an oberservational study or experiment? If observational, what type of observational study was it? **Answer**. observational, cross-section
2. What is the response variable? What is the explanatory variable? **Answer**. response: skin cancer, explanatory: coffee consumption
3. Can we conclude drinking six or more cups of coffee reduces the change of nonmelanoma skin cancer? **Answer**. No, because observational studies never allow us to determine causation, just association.

Get Married, Gain Weight

Are young couples who marry or cohabitate more likely to gian weight than those who stay single? Researchers followed 8000 men and women for 7 years. At the start of the study, none of the participants were married or living with a romantic partner. The researchers found that women who married or cohabitated during the study period gained 9 pounds more then sinlge women, and married or cohabitating men gained, on average, 6 pounds more than single men.

1. Why is this an oberational study? What type of observational study is it? **Answer**. No manipulation of explanatory variable: marital status; cohort study
2. What is the response variable in the study? What is the explanatory variable? **Answer**. response: weight, explanatory: marital status
3. Identify some of the potential lurking variables in the study. **Answer**. diet, exercise
4. Can we conclude that getting married or cohabitating causes one to gain weight? **Answer**. No, because it is an observational study, not an experiment.

1.3 Simple Random Sampling

Definition 1.3.1 random sampling. using chance (an objective device) to select individuals from a population to be included in the sample ◊

Definition 1.3.2 simple random sample. every sample of a certain size is equally likely (and every subject has an equal chance of being selected) ◊

Definition 1.3.3 N**.** population size ◊

Definition 1.3.4 n**.** sample size ◊

Definition 1.3.5 frame. list of all individuals in a population being studied ◊

Definition 1.3.6 sampling without replacement. once selected, an individual cannot be chosen again ◊

Definition 1.3.7 sampling with replacement. once selected, individuals are placed back in the population and can be chosen again ◊

1.3.1 Generating Random Numbers on the TI-84 Plus Calculator

There are several ways to generate random numbers on your calculator. One simple way is to use the `randInt` function.

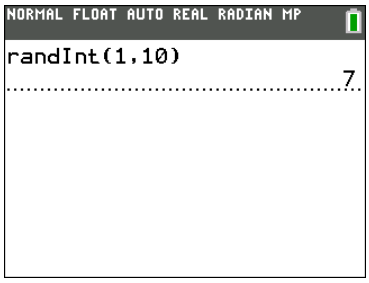
Computation 1.3.8 Generate Random Integers. To generate a random integer between 1 and 10:

1. Press the `math` button and use the cursor to highlight the `PROB` tab.
2. Select `5:randInt(`
3. Fill out the StatWizard so it looks like below

```
lower:1
upper:10
n:
Paste
```

Highlight `Paste` and press `enter`.

4. Press `enter` to execute the function you previously pasted into the command window.



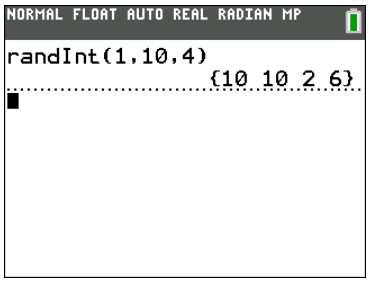
Computation 1.3.9 List of Random Integers. To generate a list of 4 random integers between 1 and 10:

1. Press the `math` button and use the cursor to highlight the `PROB` tab.
2. Select `5:randInt(`
3. Fill out the StatWizard so it looks like below

```
lower:1
upper:10
n:4
Paste
```

Highlight `Paste` and press `enter`.

4. Press `enter` to execute the function you previously pasted into the command window.



Computation 1.3.10 List of Random Integers with No Repeats. To generate a list of 10 random integers between 1 and 20 with *no repeats*:

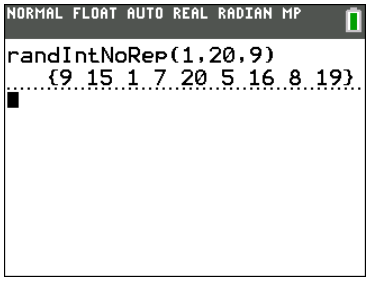
1. Press the `math` button and use the cursor to highlight the `PROB` tab.
2. Select `8:randIntNoRep(`

3. Fill out the StatWizard so it looks like below

```
lower:1
upper:20
n:10
Paste
```

   Highlight `Paste` and press `enter`.

4. Press `enter` to execute the function you previously pasted into the command window.



1.3.2 Four Tickets for Friends

Sophia has 10 friends that she would like to invite to a concert, but only 4 tickets to give away.

Mike, Jamie, Adam, Yvette, Ashley, Monica, Cherie, Julie, Willard, Bruce

1. Which of the following would produce a simple random sample of the 4 friends that she will bring with her:

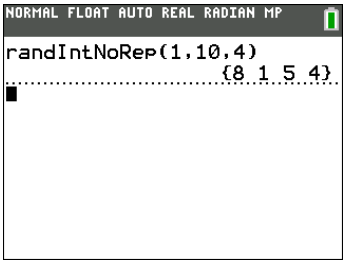
   (a) List each persons name on a piece of paper, place them in a hat and draw 4. **Answer**. yes

   (b) List the names in alphabetical order and take the first 4 names. **Answer**. no

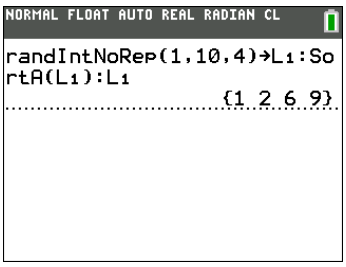
   (c) Ask one of her friends who she should bring. **Answer**. no

   (d) Number the friends from 1 to 10 and use a random number generator to produce 4 numbers between 1 and 10 which correspond to the 10 friends. **Answer**. yes

2. Use the TI graphing calculator to obtain a random sample of the 4 friends that she will bring with her and give their names. **Solution**.



Thus she would invite Mike, Yvette, Ashley and Julie.
We can also have the calculator sort the list in ascending order:



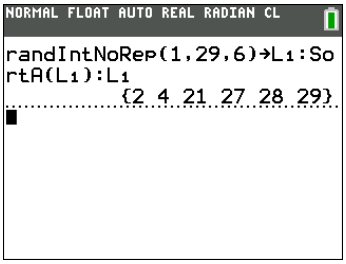
In this case, she would invite Mike, Jamie, Monica and Willard.

1.3.3 Employee Survey

The owner of a private food store is concerned with employee morale. She decides to survey the employees to learn about work environment and job satisfaction.

1 Archer	9 Foushi	16 Kemp	23 Oliver
2 Bolcerek	10 Gow	17 Lathus	24 Orsini
3 Bryant	11 Grove	18 Lindsey	25 Salazar
4 Carlisle	12 Hall	19 Massie	26 Ullrich
5 Cole	13 Hills	20 McGuffin	27 Vaneck
6 Dimas	14 Houston	21 Musa	28 Weber
7 Ellison	15 Kats	22 Nickas	29 Zavodny
8 Everhart			

1. Obtain a simple random sample of size 6 from the above table using the TI graphing calculator. **Solution**.



Thus the randomly selected employees are: Bolcerek, Carlisle, Musa, Vaneck, Weber, and Zavodny.

1.4 Other Effective Sampling Methods

Definition 1.4.1 stratified sampling. a simple random sample is drawn from each nonoverlapping subgroup (or **stratum**) into which the population has been separated.Works best when individuals in each stratum are similar in some way and we want to make sure each stratum type is represented. The size of each simple random sample is proportional to the number of individuals in the stratum from which it is selected. (e.g. separating individuals into age groups) ◊

Definition 1.4.2 systematic sampling. a starting point, p, is randomly selected and then every kth subject is included in the sample, where

$$k = \lfloor N/n \rfloor$$

The brackets, $\lfloor x \rfloor$, mean round x down to the nearest integer, or take the **floor** of x. Subjects in the sample are numbered:

$$p, p+k, p+2k, p+3k, \ldots, p+(n-1)k$$

Works best when individuals are arranged sequentially but you're not sure how many total individuals there are because you don't have a frame of the population (files in a drawer, people coming out of a building e.g. exit polling, etc.). ◊

Definition 1.4.3 cluster sampling. the population is divided into sections or **clusters**; all individuals from randomly selected clusters are included in the sample.

Works best when individuals are grouped geographically so it makes logistical sense to sample this way, but location does not affect your variable of interest; clusters are heterogeneous like the population. ◊

Definition 1.4.4 convenience sampling. sampling subjects who are easy to get (bad sampling method).

Works best when: NEVER!! (not random) ◊

Definition 1.4.5 self-selected. a convenience sample where individuals are self-selected (another bad sampling method).

Works best when: NEVER!! (not random, favors those with stronger opinions) ◊

Definition 1.4.6 multi-stage sample. a sample that is selected in multiple stages, each of which might use a different method of sampling.

Works best when: The population is organized hierarchically but no list exists. ◊

1.5 Bias in Sampling

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce tincidunt lectus nec gravida mattis. Donec viverra bibendum felis, in egestas neque commodo vel. Duis id consectetur sapien. Etiam vel ante ut quam fringilla faucibus. Pellentesque a dictum enim, sed molestie quam. Morbi finibus facilisis orci, a tempor quam hendrerit sit amet. Phasellus eget maximus enim. Pellentesque fermentum nulla lectus, sed blandit risus tempus sed. Praesent bibendum facilisis maximus. Fusce vestibulum nulla in tincidunt vulputate. Sed at velit non orci molestie vulputate. Donec auctor risus vel purus auctor laoreet quis a mi. In eget tempor quam.

Part II

Descriptive Statistics

Chapter 2

Summarizing Data in Tables and Graphs

2.1 Organizing Qualitative Data

Placeholder text

2.2 Organizing Quantitative Data

Placeholder text

2.3 Graphical Misrepresentations of Data

Placeholder text

Chapter 3

Numerically Summarizing Data

3.1 Measures of Central Tendency

Placeholder text

3.2 Measures of Dispersion

Placeholder text

3.3 Measures of Central Tendency and Dispersion from Grouped Data

Placeholder text

3.4 Measures of Position and Outliers

Placeholder text

3.5 The Five--Number Summary and Boxplots

Placeholder text

Chapter 4

Describing the Relation between Two Variables

Part III

Probability and Probability Distributions

Chapter 5

Probability

Chapter 6

Discrete Probability Distributions

Chapter 7

The Normal Probability Distribution

Part IV

Inference: From Samples to Population

Chapter 8

Sampling Distributions

Chapter 9

Estimating the Value of a Parameter

Chapter 10

Hypothesis Tests Regarding a Parameter

Chapter 11

Inferences on Two Samples