

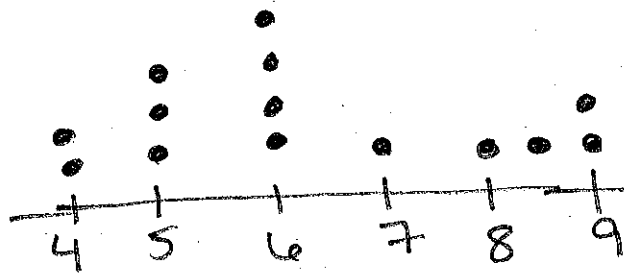
Terms

Statistics: deals with the collection, presentation and analysis of data.

Example: We are interested in the average number of hours of sleep a De Anza student in this class slept last night.

Collection

Presentation Dot plot



Analysis: compute the mean (avg.)

$$4+4+5+5+5+6+6+6+6+7+8+8.5+9+9$$

14

$$\approx 6.3 \text{ hrs}$$

Chapter 1 Key Statistics Terms

Population: The entire collection of people or things ~~is~~ under study

Sample: A subset of the population

A sample must be representative

A representative sample has the same characteristics as the population

Parameter: A number that is a property of the population

ex) Avg. # of hrs of sleep for a DA student

Statistic: a number that is a property of the sample

ex) Avg. # of hrs of sleep for the class

Variable: The characteristic of interest from each member of the population *i.e. What do you have to collect to compute the param/stat.
*usually represented by

ex) # of hrs you each slept last night

There are two types of variables: a cap. letter X or Y

- Measured or numerical var. → we can do math with these
- Categorical var. → put things into categories, but we can't do math on these

Data: actual values of the variables, actual info. collected

ex) 4, 6, 8, 6, ...

Example: We are interested in the average number of hours of sleep a De Anza student in this class slept last night. A sample of 12 students from our class is selected.

Population: All the students in this class

Sample: The 12 students whose cards were chosen

Parameter: Avg. hrs. of sleep for the whole class

Statistic: Avg. hrs of sleep last night for the 12 people in our sample

Variable: X = number of hrs. of sleep ~~one~~ each person got last night

Data: 9, 4, 6, 8, 7, 6, 5, 9, 6, 5, 6, 8.5

Example: You are interested in the proportion of registered voters in California who support California's greenhouse gas emissions law (AB32) that requires the state to reduce its greenhouse gas emissions by at least 17% over the next decade. A survey was taken February 5-17, 2013 among 834 registered voters in California.

Population:

Sample:

Parameter:

Statistic:

Q Variable: $X = \#$ of people who support the bill
 ~~$X = \text{whether a person supports}$~~
~~AB32~~

A Data: ~~i.e. we ask "Do you support AB32"~~
Yes or No or Undecided

Example: You want to determine the average number of glasses of milk college students drink per day. In your English class, you asked five students how many glasses of milk they drank the day before. The answers were 1, 0, 1, 3, and 4.

Population: All college students

Sample: 5 students in your Eng. class

Parameter: Avg. # of glasses of milk
a college student
drinks in 1 day

Statistic: Avg. # of glasses of milk
those 5 students drank
yesterday

Variable: X = # of glasses of milk
a student drank yesterday

List the Data: 1, 0, 1, 3, 4

Ch 1 Key Terms (continued)

Proportion Example 2.

Political pollsters are interested in the proportion of people that will vote for a particular cause.

Population: All registered voters

Sample: The people who are polled
For example, 500 reg. voters
who we call.

Parameter: Proportion of all reg.
voters who support the
cause

Statistic: Proportion of people polled
who support the cause

Variable: X = # of people who support
the cause

Data: Yes, no, unsure

Types of Data

- Qualitative Data = result of describing a quality or attribute
ex) hair color → brown, black, etc...
party affiliation → R, D, I, G...

- Quantitative Data = numerical data

- Discrete data → the result of counting
* These data can only have certain values

- Continuous data → the result of measuring

* Conts data usually has to be rounded,
+ discrete data doesn't

- ex) • Number of bks in your backpack
→ Discrete quantitative
- Weight of your backpack
→ Continuous quantitative
 - Brand of your backpack
→ Qualitative

Note: We can collect data from a sample
or from the entire population

Determine the type of data:

1. The number of pairs of shoes you own

Discrete quantitative

2. The type of car you drive

Qualitative

3. Where you go on vacation

Qualitative

4. The distance from your house to the nearest grocery store

Continuous quantitative

5. The number of classes you take per school year

6. The tuition for your classes

7. The type of calculator you use

8. Movie ratings

9. Political party preferences

10. The weight of sumo wrestlers

11. The amount of money won playing poker

12. The number of correct answers on a quiz

13. People's attitudes toward the government

14. IQ scores

Relative Frequency Tables

We are interested in the number of days per week that a De Anza students works.

Data	Frequency	Relative Frequency	Cumulative Rel Freq
0	6	$\frac{6}{17} = 0.3529$	0.3529
1	0	0	0.3529
2	0	0	0.3529
3	3	$\frac{3}{17} = 0.1765$	0.5294
4	2	$\frac{2}{17} = 0.1176$	0.6470
5	4	$\frac{4}{17} = 0.2353$	0.8823
6	2	$\frac{2}{17} = 0.1176$	0.9999
7	0	0	0.9999
Total	17		

Sample size

What percent work:

- 0 days per week? 35%
- 1 to 3 days per week? 18%
- Less than 5 days per week? (CRF for 4)
65%
- At least 5 days per week?
• Work 5, 6, or 7 days $\rightarrow 0.2353 + 0.1176 + 0$
35%
- At most 5 days per week?
• Not less than 5 $\rightarrow 0.9999 \rightarrow 0.6470$
88%

Sampling Methods Notes

- A sample must be representative
- If sampling is poorly done, it can lead to sampling errors

Types of Sampling

Simple Random Sample / Random Sample / SRS

1. List and number each element of the pop.
2. Generate random #s on calculator
3. Select the members of the pop. w/ those #s

Stratified Sample

1. Divide pop. into subgroups called strata
 2. Do SRS within each subgroup
- ex) Pop = US residents, subgps. = each state, choose 10 random people in ea. state

Cluster Sample

1. Divide pop. into subgroups
 2. Use SRS to choose some of the subgroups
 3. Severy every member of those subgroups
4. ex) Pop = SJ residents, subgroups = zip codes,
Sample = choose 2 zip codes & poll every person in those zip codes

Systematic Sample

1. List & # every member of the pop.
2. Randomly select one # as a starting point
3. Select every n^{th} person from there

Convenience Sample = using a sample that is readily available

- * This is not a random sample
- ex) People exiting the Cupertino Safeway between 5 and 6 pm today

Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience).

1. A soccer coach randomly selects 6 players from a group of boys aged 8 to 10, 7 players from a group of boys aged 11 to 12, and 3 players from a group of boys aged 13 to 14 to form a recreational soccer team.

stratified

2. From a list of high tech companies, a pollster randomly selects 5 of them and then interviews all human resource personnel in those companies.

cluster

3. An engineering researcher interviews 50 randomly selected women engineers and 50 randomly selected men engineers.

stratified

4. A medical researcher interviews every third cancer patient from a list of cancer patients at a local hospital.

SYS.

5. A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers.

SRS

6. To determine how many pairs of jeans a college student owns, a student interviews class mates in his algebra class.

convenience