

Chapter 1

Introduction: Variables and Processes in Statistics

***1.1**

Categorical: Students can fall in three different categories in their rating of the instructor's preparation for class.

***1.2**

- a. Individuals' weights and blood pressures involve two quantitative variables because weight and blood pressure take numerical values.
- b. Weight is treated as a categorical variable (whether or not an individual is overweight), and blood pressure is treated as a quantitative variable which can take any value over an entire range of (numerical) possibilities.
- c. Blood pressure is treated as a categorical variable (whether or not an individual has high blood pressure), and weight is treated as a quantitative variable which takes numerical values.
- d. Blood pressure is treated as a categorical variable (whether or not an individual has high or low blood pressure), and weight is treated as a categorical variable (whether or not an individual is overweight).

1.3

- a. The variable of interest is quantitative because blood pressure is recorded with numerical values for each individual.
- b. The results for a quantitative variable are best summarized with a mean.

1.4

- a. The individuals are people around the world; the variable is whether they approve or disapprove of the Iraq war, and it is categorical.
- b. The individuals are countries; the variable is what percentage of a country's people disapprove of the Iraq war, and it is quantitative (the percentage for each country takes a numerical value).
- c. The individuals are countries; the variable is whether the country as a whole approves or disapproves of the Iraq war, and it is categorical.

1.5

The reporter is mainly concerned with statistical inference because conclusions about a country's citizens in general are being made.

***1.6**

- a. The individuals are adults, the variable is marital status, and it is categorical.
- b. The individuals are states, the variable is divorce rate, and it is quantitative, because for each state a numerical value is recorded.
- c. The word “Census” suggest the information refers to the population.

1.7

The reporter is mainly concerned with displaying and summarizing data because the divorce rate for each individual state is displayed on the map.

***1.8**

It is categorical because it classifies the amounts of alcohol consumed into four groups.

1.9

Ticket price is being treated as a categorical variable because the values are classified as belonging within certain price ranges that don’t follow an ordered numerical progression.

***1.10**

- a. The variable is categorical (whether or not the individual is able to cope without a cell phone).
- b. The variable is summarized with a percentage (nearly 40 percent).

***1.11**

- a. It is a quantitative variable because amount of time spent in the sun is recorded with numerical values.
- b. They would summarize by reporting the mean amount of time spent in the sun for each of the two groups.

1.12

- a. The variable of interest is the location of first sexual encounter for the individual.
- b. The variable is categorical because it takes qualitative values (locations).
- c. The variable is being summarized with a percentage (56%).

1.13

The sociologists are mainly concerned with performing statistical inference to make conclusions about the population of all sexually active teenagers.

***1.14**

We consider the workers to be a population because of the word “Census”.

1.15

The numbers arose from Census data, referring to entire populations, not samples.

1.16

- a. The Americans polled are the individuals comprising the sample.
- b. Would you like to live to be at least 100? (Other phrasings are possible.)
- c. How old would you like to live to be? (Other phrasings are possible.)

***1.17**

- a. (2) For each individual (appliance), a numerical value is recorded (percentage owning that appliance), and so there is one quantitative variable involved.
- b. (3) There is a categorical explanatory variable (which year – 1987 or 2001) and a quantitative response variable (what percentage of households owned the appliance in that year).
- c. (2) There is one quantitative variable for which numerical values are recorded (how many television sets are owned).

1.18

- a. The individuals being studied are appliances.
- b. The variable of interest is for each appliance the percent of homes that contained it.
- c. The variable is quantitative because percentages take numerical values.

1.19

- a. The explanatory variable is year (1987 or 2001).
- b. The explanatory variable is categorical because the two year values are used to classify the responses into two groups; the year numbers are not treated as quantities.
- c. The response variable is what percentage of homes had the appliance.
- d. The response variable is quantitative because percentages are recorded with numerical values.
- e. We'd expect the percentages to be higher in 2001, due to the decrease in cost of appliances and/or American's increasing tendency to rely on "gadgets".

1.20

- a. The variable of interest is quantitative because the number of television sets per household is recorded by numerically.
- c. The reported summary is a mean (average number of televisions per individual).

1.21

- a. $Q \rightarrow C$: Age in years is taken as a quantitative explanatory variable and whether or not a viewer likes the ad is taken as a categorical explanatory variable.
- b. $C \rightarrow C$: Age is taken as a categorical variable (whether or not an individual falls into a certain age range) for the response variable of whether or not the individual likes the ad.
- c. $Q \rightarrow Q$: Age in years is treated as the explanatory variable and the numerical rating of the ad is the response variable.
- d. $C \rightarrow Q$: Age group is treated as a categorical explanatory variable and numerical ratings of the ad is the quantitative response variable.

1.22

Television advertisers are concerned with data production because they are discussing ways to treat age and people's response to ads as either categorical or quantitative variables.

***1.23**

- a. Deciding on the details of the survey questions is data production.
- b. Finding the average is part of the displaying and summarizing phase.
- c. Finding the unlikelihood of a certain sample average involves probability.
- d. Drawing a more general conclusion about all the professor's students is a statistical inference task.

1.24

The variable of interest is quantitative because a numerical value is being recorded for each city.

1.25

The magazine is concerned with displaying and summarizing the data gained in their survey.

1.26

- a. The explanatory variable is gender.
- b. Gender is a categorical variable because an individual can fall into only two possible categories (male or female).
- c. The response variable is the type of graffiti in public restrooms.
- d. Type of graffiti is being treated as a categorical variable because the two possible categories are whether the writings are competitive and derogatory or advisory and sympathetic.
- e. The type of writings for each gender would be summarized by proportions because it is a categorical variable.

1.27

Smoking is the response variable to whether or not a person is an alcoholic.

1.28

Smoking is the explanatory variable, considering if whether or not a person smokes plays a role in whether or not they are an alcoholic.

1.29

Carlos

1.30

The researchers are concerned with probability because they are discussing the chance of a certain outcome (more than half of patients given feeding tubes) occurring.

1.31

Responses will vary.

1.32

Responses will vary.

