

**SOLUTIONS MANUAL**



# Stats

*Data & Models*



*De Veaux Velleman Bock*

Includes **ANSWERS** by Paul Velleman

## Chapter 2 - Data

- 1. The News.** Answers will vary.
- 2. Investments.** *Who* – 30 similar companies. *What* – 401(k) employee participation rates (in percent). *When* – Sometime after reading the 1992 *Fortune* article. *Where* – United States. *Why* – the company in question is concerned that its employee participation rate is lower than the rates of similar companies. *How* – Companies were “sampled”, using an unspecified method. *Variables* – There is one quantitative variable, 401(k) participation rate. *Concerns* – How was the sample of companies selected?
- 3. Oil spills.** *Who* – 50 recent oil spills. *What* – Spillage amount (no specified unit) and cause of puncture. *When* – Sometime between 1990 (Oil Pollution Act passed) and 1995 (report made in *Marine Technology*). *Where* – United States. *Why* – To determine whether or not spillage amount per oil spill has decreased since Congress passed the 1990 Oil Pollution Act and use that information in the design of new tankers. *How* – Not specified. *Variables* – There are 2 variables. Spillage amount is a quantitative variable, and cause of puncture is a categorical variable.
- 4. Oscars.** *Who* – Winners of Best Actor or Best Actress Oscars. *What* – Age (in years) at the time they won Best Actor or Best Actress Oscars. *When* – From the beginning of the Academy Awards to the present. *Where* – United States. *Why* – To compare age distributions of actors and actresses. *How* – It appears that this information was collected from a census of all Best Actor and Actress Oscar winners. *Variables* – There is one quantitative variable, age of the actor or actress.
- 5. Weighing bears.** *Who* – 54 bears. *What* – Weight, neck size, length (no specified units), and sex. *When* – Not specified. *Where* – Not specified. *Why* – Since bears are difficult to weigh, the researchers hope to use the relationships between weight, neck size, length, and sex of bears to estimate the weight of bears, given the other, more observable features of the bear. *How* – Researchers collected data on 54 bears they were able to catch. *Variables* – There are 4 variables; weight, neck size, and length are quantitative variables, and sex is a categorical variable. No units are specified for the quantitative variables. *Concerns* – The researchers are (obviously!) only able to collect data from bears they were able to catch. This method is a good one, as long as the researchers believe the bears caught are representative of all bears, in regard to the relationships between weight, neck size, length, and sex.
- 6. Molten iron.** *Who* – 10 crankshafts. *What* – The pouring temperature (in degrees Fahrenheit) of molten iron. *When* – Sometime before the *Quality Engineering* report in 1995. *Where* – Cleveland Casting Plant in Cleveland, Ohio. *Why* – Cleveland Casting is interested in quality control. They want know how close to the desired pouring temperature of 2,550 degrees Fahrenheit the machine is able to keep the molten iron. *How* – Random sampling. *Variables* – Pouring temperature is a quantitative variable.

## 2 Part I Exploring and Understanding Data

7. **Arby's menu.** *Who* – Arby's sandwiches. *What* – type of meat, number of calories (in calories), and serving size (in ounces). *When* – Not specified. *Where* – Arby's restaurants. *Why* – These data might be used to assess the nutritional value of the different sandwiches. *How* – Information was gathered from each of the sandwiches on the menu at Arby's, resulting in a census. *Variables* – There are three variables. Number of calories and serving size are quantitative variables, and type of meat is a categorical variable.
8. **Firefighters.** *Who* – Shipboard Firefighters. *What* – Pulling force (in Newtons), weight (probably in pounds), and gender. *When* – Not specified, but sometime prior to the 1982 report. *Where* – Not specified. *Why* – The researchers wanted to compare the abilities of men and women. *How* – Not specified. *Variables* – There are three variables. Pulling force and weight are quantitative variables. Gender is a categorical variable.
9. **Babies.** *Who* – 882 births. *What* – Mother's age (in years), length of pregnancy (in weeks), type of birth (caesarean, induced, or natural), level of prenatal care (none, minimal, or adequate), birth weight of baby (unit of measurement not specified, but probably pounds and ounces), gender of baby (male or female), and baby's health problems (none, minor, major). *When* – 1998-2000. *Where* – Large city hospital. *Why* – Researchers were investigating the impact of prenatal care on newborn health. *How* – It appears that they kept track of all births in the form of hospital records, although it is not specifically stated. *Variables* – There are three quantitative variables: mother's age, length of pregnancy, and birth weight of baby. There are four categorical variables: type of birth, level of prenatal care, gender of baby, and baby's health problems.
10. **Flowers.** *Who* – 385 species of flowers. *What* – Date of first flowering (in days). *When* – Not specified. *Where* – Southern England. *Why* – The researchers believe that this indicates a warming of the overall climate. *How* – Not specified. *Variables* – Date of first flowering is a quantitative variable. *Concerns* – Hopefully, date of first flowering was measured in days from January 1, or some other convention, to avoid problems with leap years.
11. **Fitness.** *Who* – 25,892 men. *What* – Fitness level and cause of death. It is not clear what categories were used for fitness level. *When* – Over a 10-year period prior to the article being published in May 2002. *Where* – Not specified. *Why* – To establish an association between fitness level and death from cancer. *How* – Researchers tracked the group of men over a 10-year period. *Variables* – There are two categorical variables: fitness level and cause of death.
12. **Schools.** *Who* – Students. *What* – Age (probably in years, though perhaps in years and months), race or ethnicity, number of absences, grade level, reading score, math score, and disabilities/special needs. *When* – This information must be kept current. *Where* – Not specified. *Why* – Keeping this information is a state requirement. *How* – The information is collected and stored as part of school records. *Variables* – There are seven variables. Race or ethnicity, grade level, and disabilities/special needs are categorical variables. Number of absences, age, reading test score, and math test score are quantitative variables. *Concerns* – What tests are used to measure reading and math ability, and what are the units of measure for the tests?

- 13. Herbal medicine.** *Who* – experiment volunteers. *What* – herbal cold remedy or sugar solution, and cold severity. *When* – Not specified. *Where* – Major pharmaceutical firm. *Why* – Scientists were testing the efficacy of an herbal compound on the severity of the common cold. *How* – The scientists set up a controlled experiment. *Variables* – There are two variables. Type of treatment (herbal or sugar solution) is categorical, and severity rating is quantitative. *Concerns* – The severity of a cold seems subjective and difficult to quantify. Also, the scientists may feel pressure to report negative findings about the herbal product.
- 14. Tracking sales.** *Who* – Customers of a start-up company. *What* – Customer name, ID number, region of the country, date of last purchase, amount of purchase (probably in dollars), and item purchased. *When* – Present time. *Where* – United States. *Why* – The company is building a data base of sales information. *How* – Presumably, the company records the information from each new customer. *Variables* – There are six variables. Name, ID number, region of the country, and item purchased are categorical variables. Date and amount of purchase are quantitative variables. *Concerns* – Region is a categorical variable, and it is potentially confusing to record it as a number.
- 15. Cars.** *Who* – Automobiles. *What* – Make, country of origin, type of vehicle and age of vehicle (probably in years). *When* – Not specified. *Where* – A large university. *Why* – Not specified. *How* – A survey was taken in campus parking lots. *Variables* – There are three categorical variables and one quantitative variable. Make, country of origin, and type of vehicle are categorical variables, and age of vehicle is a quantitative variable.
- 16. Vineyards.** *Who* – Vineyards. *What* – Size of vineyard (in acres), number of years in existence, state, varieties of grapes grown, average case price (in dollars), gross sales (probably in dollars), and percent profit. *When* – Not specified. *Where* – United States. *Why* – Business analysts hoped to provide information that would be helpful to producers of American wines. *How* – Not specified. *Variables* – There are five quantitative variables and two categorical variables. Size of vineyard, number of years in existence, average case price, gross sales, and percent profit are quantitative variables. State and variety of grapes grown are categorical variables.
- 17. Streams.** *Who* – Streams. *What* – Name of stream, substrate of the stream (limestone, shale, or mixed), acidity of the water (measured in pH), temperature (in degrees Celsius), and BCI (unknown units). *When* – Not specified. *Where* – Upstate New York. *Why* – Research is conducted for an Ecology class. *How* – Not specified. *Variables* – There are five variables. Name and substrate of the stream are categorical variables, and acidity, temperature, and BCI are quantitative variables.
- 18. Age and party.** *Who* – 1180 Americans. *What* – Region, age (in years), political affiliation, and whether or not the person voted in the 1998 midterm Congressional election. *When* – First quarter of 1999. *Where* – United States. *Why* – The information was gathered for presentation in a Gallup public opinion poll. *How* – Phone Survey. *Variables* – There are four variables. Region, political affiliation, and whether or not the person voted in 1998 are categorical variables, and age is a quantitative variable.

#### 4 Part I Exploring and Understanding Data

- 19. Air travel.** *Who* – All airline flights in the United States. *What* – Type of aircraft, number of passengers, whether departures and arrivals were on schedule, and mechanical problems. *When* – This information is currently reported. *Where* – United States. *Why* – This information is required by the Federal Aviation Administration. *How* – Data is collected from airline flight information. *Variables* – There are four variables. Type of aircraft, departure and arrival timeliness, and mechanical problems are categorical variables, and number of passengers is a quantitative variable.
- 20. Fuel economy.** *Who* – Every model of automobile in the United States. *What* – Vehicle manufacturer, vehicle type, weight (probably in pounds), horsepower (in horsepower), and gas mileage (in miles per gallon) for city and highway driving. *When* – This information is collected currently. *Where* – United States. *Why* – The Environmental Protection Agency uses the information to track fuel economy of vehicles. *How* – The data is collected from the manufacturer of each model. *Variables* – There are six variables. City mileage, highway mileage, weight, and horsepower are quantitative variables. Manufacturer and type of car are categorical variables.
- 21. Refrigerators.** *Who* – 41 refrigerators. *What* – Brand, cost (probably in dollars), size (in cu. ft.), type, estimated annual energy cost (probably in dollars), overall rating, and repair history (in percent requiring repair over the past five years). *When* – 2002. *Where* – United States. *Why* – The information was compiled to provide information to the readers of *Consumer Reports*. *How* – Not specified. *Variables* – There are 7 variables. Brand, type, and overall rating are categorical variables. Cost, size, estimated energy cost, and repair history are quantitative variables.
- 22. Lotto.** *Who* – States in the US. *What* – State name, whether or not the state sponsors a lottery, the number of numbers in the lottery, the number of matches required to win, and the probability of holding a winning ticket. *When* – 1998. *Where* – United States. *Why* – It is likely that this study was performed in order to compare the chances of winning the lottery in each state. *How* – Although not specified, the researchers probably simply gathered data from a number of different sources, like state lottery websites and publications. *Variables* – There are 5 variables. State name and whether or not the state sponsors a lottery are categorical variables, and number, matches, and probability of winning are quantitative.
- 23. Sleep.** *Who* – Days. *What* – Hours of sleep, whether or not he suffered from “early awakening”, whether or not he watched TV in the morning, whether or not he watched TV in the evening, the number of hours spent standing that day, and mood (measured on an arbitrary scale 10 – 90.) *When* – Prior to the 2001 article. *Where* – Not specified. *Why* – The professor wanted to analyze sleep patterns, possibly to examine the relationships between variables related to sleep. *How* – The professor recorded data relating to sleep each day. *Variables* – There are 6 variables. “Early awakening”, TV in the morning, and TV in the evening are categorical variables, and hours of sleep, hours standing, and mood are quantitative variables.

- 24. Indy.** *Who* – Indy 500 races. *What* – Year, winner, pole position, average speed (in miles per hour), pole winner, and average pole speed (in miles per hour). *When* – 1911 – 2003. *Where* – Indianapolis, Indiana. *Why* – It is interesting to examine the trends in Indy 500 races. *How* – Official statistics are kept for the race every year. *Variables* – There are 6 variables. Winner and pole winner are categorical variables. Year, pole position, average speed, and average pole speed are quantitative variables.
- 25. Horse Race.** *Who* – Kentucky Derby races. *What* – Date, winner, margin (in lengths), jockey, winner's payoff (in dollars), duration of the race (in minutes and seconds), and track condition. *When* – 1875 – 2004. *Where* – Churchill Downs, Louisville, Kentucky. *Why* – It is interesting to examine the trends in the Kentucky Derby. *How* – Official statistics are kept for the race each year. *Variables* – There are 7 variables. Winner, jockey, and track condition are categorical variables. Date, margin, winner's payoff, and duration are quantitative variables.
- 26. Stat students.** *Who* – Students in a statistics class. *What* – Height (units not specified, but presumably in inches), shoe size, sex, degree program, and birth order. *When* – Not specified. *Where* – Not specified. *Why* – The information was collected for use in classroom illustrations. *How* – An online survey was conducted. Presumably, participation was required for all members of the class. *Variables* – There are 5 variables. Sex and degree program are categorical variables. Height and shoe size are quantitative variables. Birth order can be either categorical or quantitative, depending on how it is used. For example, we could find the proportion of the students who were second born. In this case, we are treating the variable as categorical. If we found the average birth order, we would be using the variable as a quantitative variable. *Concerns* – Shoe sizes for men and women are not measured on the same scale.