Contents

Part I Exploring and Understanding Data

Chapter 1	*Stats Starts Here*	*1*
Chapter 2	*Displaying and Describing Categorical Data*	*5*
Chapter 3	*Displaying and Summarizing Quantitative Data*	*20*
Chapter 4	*Understanding and Comparing Distributions*	*39*
Chapter 5	*The Standard Deviation as a Ruler and the Normal Model*	*55*
Review of Part I	*Exploring and Understanding Data*	*74*

Part II Exploring Relationships Between Variables

Chapter 6	*Scatterplots, Association, and Correlation*	*102*
Chapter 7	*Linear Regression*	*113*
Chapter 8	*Regression Wisdom*	*138*
Chapter 9	*Re-expressing Data: Get It Straight!*	*157*
Review of Part II	*Exploring Relationships Between Variables*	*168*

Part III Gathering Data

Chapter 10	*Understanding Randomness*	*194*
Chapter 11	*Sample Surveys*	*204*
Chapter 12	*Experiments*	*215*
Review of Part III	*Gathering Data*	*233*

Part IV Randomness and Probability

Chapter 13	*From Randomness to Probability*	*249*
Chapter 14	*Probability Rules!*	*262*
Chapter 15	*Random Variables*	*284*
Chapter 16	*Probability Models*	*305*
Review of Part IV	*Randomness and Probability*	*334*

Part V From the Data at Hand to the World at Large

Chapter 17	*Sampling Distribution Models*	*358*
Chapter 18	*Confidence Intervals for Proportions*	*392*
Chapter 19	*Testing Hypotheses About Proportions*	*408*
Chapter 20	*More About Tests and Intervals*	*427*
Chapter 21	*Comparing Two Proportions*	*444*
Review of Part V	*From the Data at Hand to the World at Large*	*469*

Part VI Learning About the World

Chapter 22	*Inferences About Means*	*497*
Chapter 23	*Comparing Means*	*523*
Chapter 24	*Paired Samples and Blocks*	*553*
Review of Part VI	*Learning About the World*	*573*

Part VII Inference When Variables are Related

Chapter 25	*Comparing Counts*	*611*
Chapter 26	*Inferences for Regression*	*642*
Review of Part VII	*Inference When Variables are Related*	*669*
Chapter 27	*Analysis of Variance*	*707*
Chapter 28	*Multiple Regression*	*718*

Chapter 1 – Stats Starts Here

1. **Voters.** The response, party affiliation, is a categorical variable.

2. **Job growth.** The response, change in workforce size, is a categorical variable.

3. **Medicine.** The company is measuring minutes, a quantitative variable.

4. **Stress.** The researcher is measuring heart rate, a quantitative variable.

5. **The news.** Answers will vary.

6. **The Internet.** Answers will vary.

7. **Bicycle and pedestrian safety.** *Who* – pedestrians and bicyclists killed or severely injured in New York City between 2010 and 2014. *What* – proportion of pedestrians and bicyclists killed or injured by left-turning and right-turning vehicles. *Population of interest* – Answers may vary. Perhaps: All pedestrians and bicyclists killed or severely injured in New York City between 2010 and 2014 *or* all pedestrians and bicyclists killed or severely injured in New York City in any year *or* all pedestrians and bicyclists killed or severely injured in major U.S. cities. Discuss with students why these and other solutions may be problematic.

8. **Investments.** *Who* – 30 companies with similar retirement plans. *What* – 401(k) employee participation rates. *Population of interest* – All companies with similar retirement plans.

9. **Fake news.** *Who* – Middle school, high school, and college students in 12 states. *What* – Ability to evaluate the quality of information found in different online resources. *Population of interest* – All U.S middle school, high school, and college students.

10. **Biological instinct.** *Who* – 40 undergraduate women. *What* – Whether or not the women could identify the sexual orientation of men based on a picture. *Population of interest* – All women.

11. **Blindness.** *Who* – 24 patients. *What* – Whether or not stem cell therapy was effective in treating Stargardt's disease and/or dry age-related macular degeneration. *Population of interest* – All people with these eye conditions.

12. **Molten iron.** *Who* – 10 castings at Cleveland Casting. *What* – The pouring temperature (in degrees Fahrenheit) of molten iron. *Population of interest* – All castings at Cleveland Casting.

13. 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.

14. 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?

15. 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.

16. Party and the environment. *Who* – American voters. *What* – Gender, age (in years), race, party affiliation, education, whether or not the person was "worried a great deal" about climate change, air pollution, and pollution of waterways. *When* – 2017. *Where* – United States. *Why* – The information was gathered for presentation in a Gallup public opinion poll. *How* – Poll. *Variables* – There are eight variables. Gender, race, party affiliation, education, whether or not the person was "worried a great deal" about climate change, air pollution, and pollution of waterways are categorical variables; age is a quantitative variable.

17. 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, 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 is not specifically stated. *Variables* – There are seven variables. Type of birth, level of prenatal care, gender of baby, and baby's health problems are categorical variables; mother's age, length of pregnancy, and birth weight of baby are quantitative variables.

18. Flowers. *Who* – 385 species of flowers for 47 years. 385(47) = 18,095 cases. *What* – Date of first flowering (in days). *When* – An unspecified 47 year period. *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. The number of years, 47, is also a variable.

19. Herbal medicine. *Who* – Patients. *What* – Herbal cold remedy or sugar solution, and cold severity on a scale of 0-5. *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 conducted an experiment. *Variables* – There are two variables. Type of treatment (herbal or sugar solution) is a categorical variable, and severity rating is a quantitative variable. The subjectivity of "cold severity" is a concern that should be raised about this study.

20. 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 grapes. *How* – Not specified. *Variables* – There are seven variables. State and variety of grapes grown are categorical variables; size of vineyard, number of years in existence, average case price, gross sales, and percent profit are quantitative variables.

21. Streams. *Who* – Streams. *What* – A number of variables including: 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; acidity, temperature, and BCI are quantitative variables.

22. Fuel economy. *Who* – Every model of automobile. *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. Manufacturer and type of car are categorical variables; weight, horsepower, city mileage, and highway mileage are quantitative variables.

23. Refrigerators. *Who* – 148 models of French door style refrigerators. *What* – Brand, price (probably in dollars), temperature performance, temperature uniformity, energy efficiency, noise, ease of use (the five previous variables are measured as Excellent, Very Good, Good, Fair, or Poor), number of doors, capacity (cu. ft.), exterior height (in.), exterior width (in.), and exterior depth (in.). *When* – 2017. *Where* – Not stated. *Why* – The information was compiled to provide information to the readers of *Consumer Reports*. *How* – Not specified. *Variables* – There are 11 variables. Brand, temperature performance, temperature uniformity, energy efficiency, noise, and ease of use are categorical variables; price, number of doors, capacity, exterior height, exterior width, and exterior depth are quantitative variables.

24. Walking in circles. *Who* – 32 people. *What* – Sex, height, handedness, the number of yards walked before going out of bounds, and the side of the field on which the person walked out of bounds. *When* – Not specified. *Where* – Not specified. *Why* – The researcher was interested in whether people naturally walk in circles when lost. *How* – Data were collected by observing the people on the field, as well as by measuring and asking the participants. *Variables* – There are 5 variables. Sex, handedness, and side of the field are categorical variables; height and number of yards walked are quantitative variables.

25. Kentucky Derby 2016. *Who* – Kentucky Derby races. *What* – Year, winner, jockey, trainer, owner, and time (in minutes, seconds, and hundredths of a second. *When* – 1875 – 2016. *Where* – Churchill Downs, Louisville, Kentucky. *Why* –Not specified. *How* – Official statistics are kept for the race each year. *Variables* – There are 6 variables. Winner, jockey, trainer and owner are categorical variables; year and time are quantitative variables.

26. Indy 2016. *Who* – Indy 500 races. *What* – Year, winner, time (in minutes, seconds, and hundredths of a second), and average speed (in miles per hour). *When* – 1911 – 2016. *Where* – Indianapolis Motor Speedway. *Why* – Not specified. *How* – Official statistics are kept for the race every year. *Variables* – There are 4 variables. Winner is a categorical variable, while year, time, and average speed are quantitative variables.