INSTRUCTOR'S SOLUTIONS MANUAL

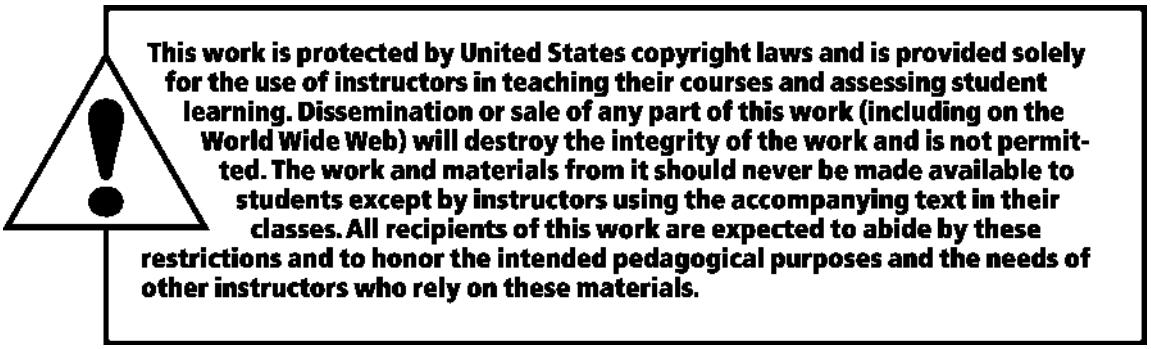
SONGFENG ZHENG
Missouri State University

PROBABILITY & STATISTICS WITH R FOR ENGINEERS AND SCIENTISTS

Michael Akritas
The Pennsylvania State University

PEARSON

Boston Columbus Hoboken Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo



Reproduced by Pearson from electronic files supplied by the author.

Publishing as Pearson, 501 Boylston Street, Boston, MA 02116.

ISBN-13: 978-0-321-85308-0
ISBN-10: 0-321-85308-3

www.pearsonhighered.com

PEARSON

Contents

Chapter 1
Basic Statistical Concepts

1.2 Populations and Samples

1. (a) The population consists of the customers who bought a car during the previous year.

   (b) The population is not hypothetical.

2. (a) There are three populations, one for each variety of corn. Each variety of corn that has been and will be planted on all kinds of plots make up the population.

   (b) The characteristic of interest is the yield of each variety of corn at the time of harvest.

   (c) There are three samples, one for each variety of corn. Each variety of corn that was planted on the 10 randomly selected plots make up the sample.

3. (a) There are two populations, one for each shift. The cars that have been and will be produced on each shift make up the population.

   (b) The populations are hypothetical.

   (c) The characteristic of interest is the number of nonconformances per car.

4. (a) The population consists of the all domestic flights, past or future.

   (b) The sample consists of the 175 domestic flights.

   (c) The characteristic of interest is the air quality, quantified by the degree of staleness.

5. (a) There are two populations, one for each teaching method.

   (b) The population consists of all students who took or will take a statistics course for engineering using one of each teaching methods.

   (c) The populations are hypothetical.

   (d) The samples consist of the students whose scores will be recorded at the end of the semester.

1.3 Some Sampling Concepts

1. The second choice provides a closer approximation to simple random sample.

2. (a) It is not a simple random sample.

   (b) In (a), each member of the population does not have equal chance to be selected, thus it is not a simple random sample. Instead, the method described in (a) is a stratified sampling.

3. (a) The population includes all the drivers in the university town.

   (b) The student's classmates do not constitute a simple random sample.

   (c) It is a convenient sample.

   (d) Young college students are not experienced drivers, thus they tend to use seat belts less. Consequently, the sample in this problem will underestimate the proportion.

4. We identify each person with a number from 1 to 70. Then we write each number from 1 to 70 on separate, identical slips of paper, put all 70 slips of paper in a box, and mix them thoroughly. Finally, we select 15 slips from the box, one at a time, without replacement. The 15 selected numbers specify the desired sample of size $n = 15$ from the 70 iPhones. The R command is

   $$y = sample(seq(1,70),\ size=15)$$

   A sample set is 52 8 14 48 62 6 70 35 18 20 3 41 50 27 40.

5. We identify each pipe with a number from 1 to 90. Then we write each number from 1 to 90 on separate, identical slips of paper, put all 90 slips of paper in a box, and mix them thoroughly. Finally, we select 5 slips from the box, one at a time, without replacement. The 5 selected numbers specify the desired sample of size $n = 5$ from the 90 drain pipes. The R command is

   $$y = sample(seq(1,90),\ size=5),$$

   A sample set is 7 38 65 71 57.

6. (a) We identify each client with a number from 1 to 1000. Then we write each number from 1 to 1000 on separate, identical slips of paper, put all 1000 slips of paper in a box, and mix them thoroughly. Finally, we select 100 slips from the box, one at a time, without replacement. The 100 selected numbers specify the desired sample of size $n = 100$ from the 1000 clients.

(b) Using stratified sampling: Get a simple random sample of size 80 from the sub-population of Caucasian-Americans, a simple random sample of size 15 from the sub-population of African-Americans, and a simple random sample of size 5 from the sub-population of Hispanic-Americans. Then combine the three subsamples together.

(c) The R command for part (a) is

y = sample(seq(1,1000), size=100)

and the R command for part (b) is

y1 = sample(seq(1,800), size=80)
y2 = sample(seq(801,950), size=15)
y3 = sample(seq(951,1000), size=5)
y = c(y1, y2, y3)

7. One method is to take a simple random sample of size n from the population of N customers (of all dealerships of that car manufacturer) who bought a car the previous year.

   The second method is to divide the population of the previous year's customers into three strata according to the type of car each customer bought and perform stratified sampling with proportional allocation of sample sizes. That is, if N_1, N_2, N_3 denote the sizes of the three strata, take simple random samples of approximate sizes (due to round-off) $n_1 = n(N_1/N)$, $n_2 = n(N_2/N)$, $n_3 = n(N_3/N)$, respectively, from each of the three strata. Stratified sampling assures that the sample representation of the three strata equals their population representation.

8. It is not a simple random sample because products from facility B have a smaller chance to be selected than products from facility A.

9. No, because the method excludes samples consisting of n_1 cars from the first shift and $n_2 = 9 - n_1$ from the second shift for any (n_1, n_2) different from $(6, 3)$.

1.4 Random Variables and Statistical Populations

1. (a) The variable of interest is the number of scratches in each plate. The statistical population consists of 500 numbers, 190 zeros, 160 ones, and 150 twos.

   (b) The variable of interest is quantitative.

   (c) The variable of interest is univariate.

2. (a) Statistical population: If there are N undergraduate students enrolled at PSU, the statistical population is a list of length N and the i-th element in the list is the major of the i-th student. The variable of interest is qualitative. Another possible variable: gender.

(b) Statistical population: If there are N restaurants on campus, the statistical population consists of a list of N numbers, and the i-th element is the capacity of the i-th restaurant. The variable of interest is quantitative. Another possible variable: food type.

(c) Statistical population: If there are N books in Penn State libraries, the statistical population consists of a list of N numbers, and the i-th element is the check-out frequency of the i-th book in the library. The variable of interest is quantitative. Another possible variable: pages of the book.

(d) Statistical population: If there are N steel cylinders made in the given month, the population consists of a list of N numbers, and the i-th element is the diameter of the i-th steel cylinder made in the given month. The variable of interest is quantitative. Another possible variable: weight.

3. (a) The variable of interest is univariate.

(b) The variable of interest is quantitative.

(c) If N is the number cars of available for inspection, the statistical population consists of N numbers, $\{v_1, \cdots, v_N\}$, where v_i is the total number of engine and transmission nonconformances of the ith car.

(d) If the number of nonconformances in the engine and transmission are recorded separately for each car, the new variable would be bivariate.

4. (a) The variable of interest is the degree of staleness. Statistical population consists of a list of 175 numbers, and the i-th number is the degree of staleness of the air in the i-th domestic flight.

(b) The variable of interest is quantitative.

(c) The variable of interest is univariate.

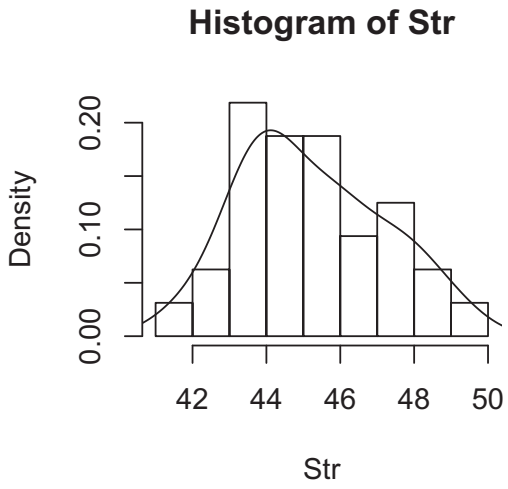
5. (a) The variable of interest is the type of car a customer bought and his/her satisfaction level. Statistical population: If there are N customers who bought a new car in the previous year, the statistical population is a list of N elements, and the i-th element is the car type the i-th customer bought along with his/her satisfaction level, which is a number between 1 to 6.

(b) The variable of interest is bivariate.

(c) The variable of interest has two components. The first is qualitative and the second is quantitative.

1.5 Basic Graphics for Data Visualization

1. The histogram produced by the commands is shown as following:



The stem and leaf plot is as following:

The decimal point is at the |

41 | 5

42 | 39

43 | 1445788

44 | 122357

45 | 1446

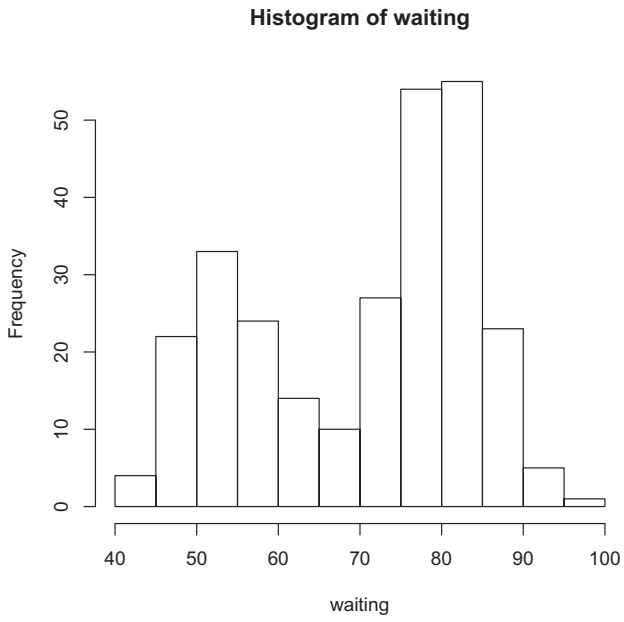
46 | 00246

47 | 3577

48 | 36

49 | 3

2. The histogram on the waiting time is as following



The corresponding stem and leaf plot is given below. It is clear that the shape of the stem and leaf plot is similar to that of the histogram.

The decimal point is 1 digit(s) to the right of the |

4 | 3

4 | 55566666777788899999

5 | 00000111111222223333333444444444

5 | 555555666677788889999999

6 | 00000022223334444

6 | 555667899

7 | 00001111123333333444444

7 | 55555555666666666777777777777888888888888888999999999

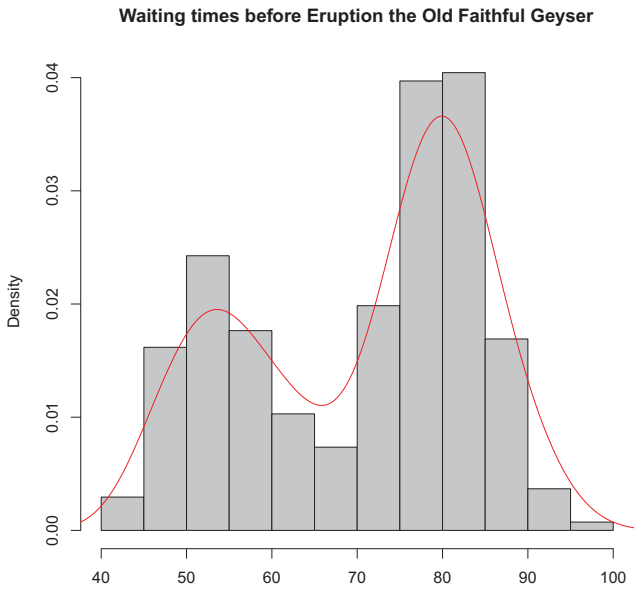
8 | 000000001111111111112222222222223333333333333334444444444

8 | 55555566666677888888999

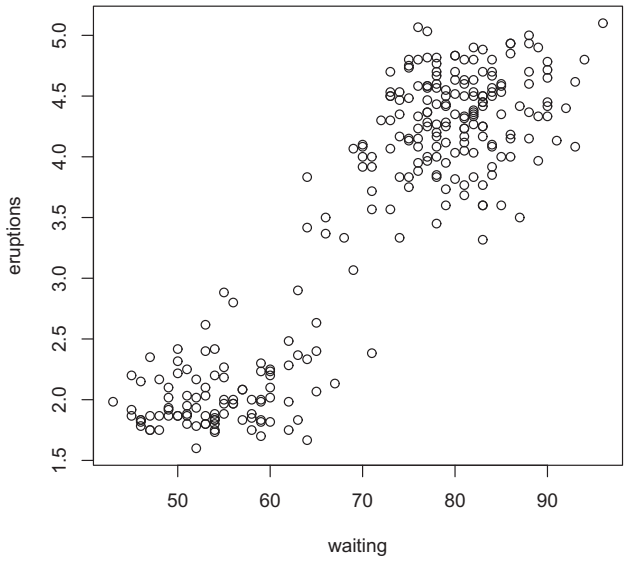
9 | 00000012334

9 | 6

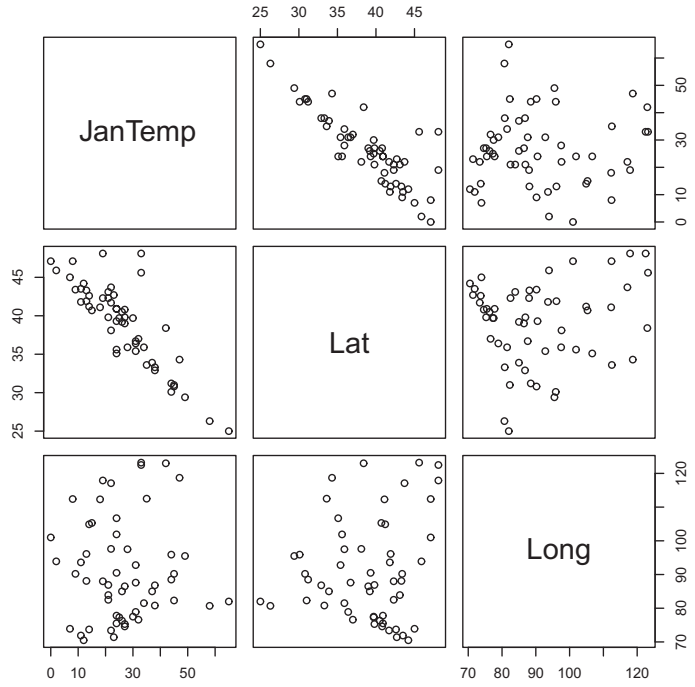
The histogram with title and the colored smooth curve superimposed is shown as



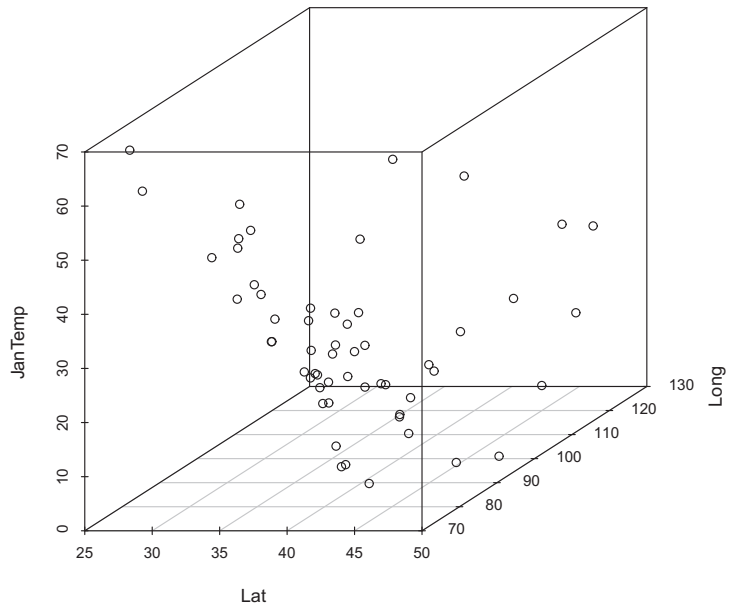
3. The scatterplot is shown below. From the scatter plot, it seems that if the waiting time before eruption is longer, the duration is also longer.



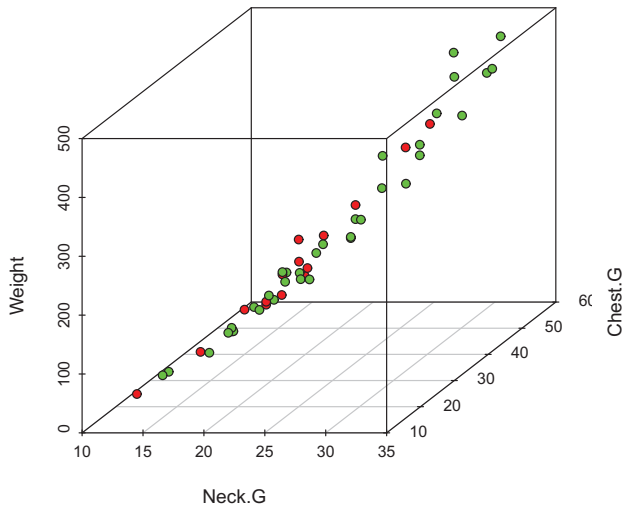
4. (a) The scatterplot matrix is given below. From the figure, it seems that the latitude is a better predictor of the temperature because as the latitude changes, the temperature shows a clear pattern, while there is no pattern as the longitude changes.



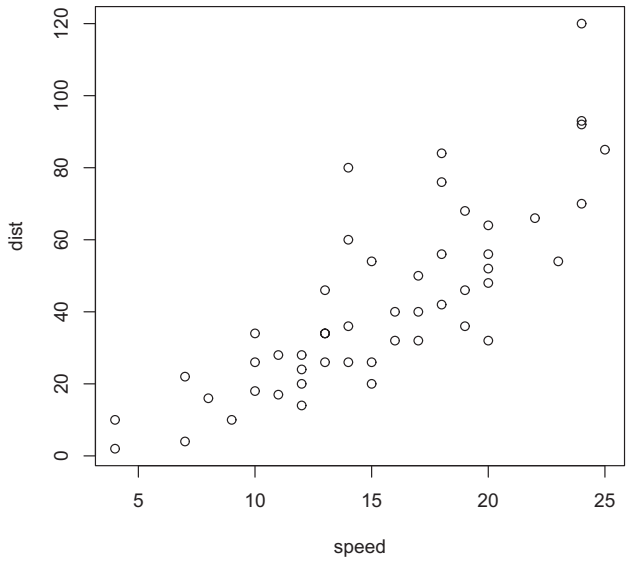
(b) The following figure gives the 3D scatter plot. The 3D scatter plot also shows that the latitude is a better predictor for the temperature.



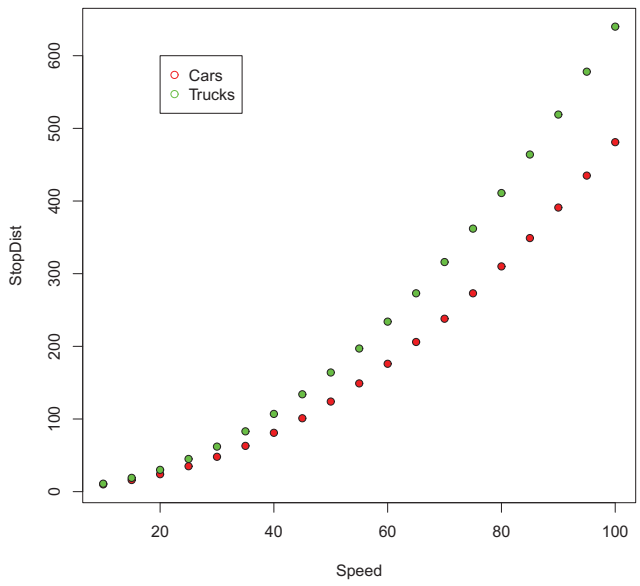
5. The 3D scatterplot is shown below



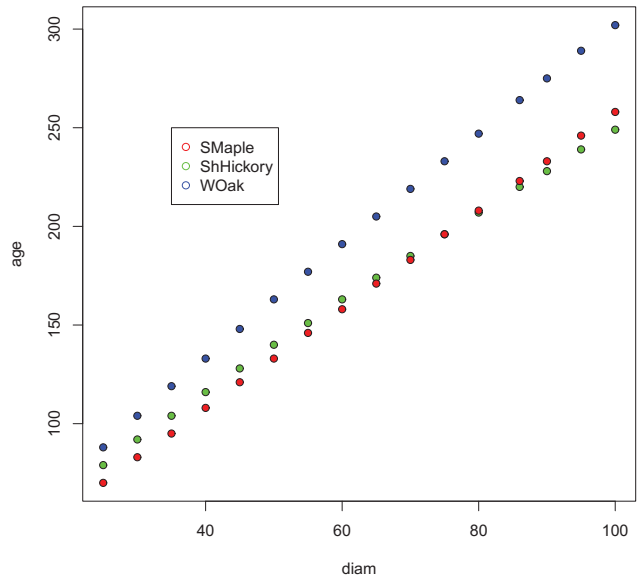
6. The scatterplot is shown below. From the scatterplot, it is clear that in general, if the speed is high, the breaking distance is larger.



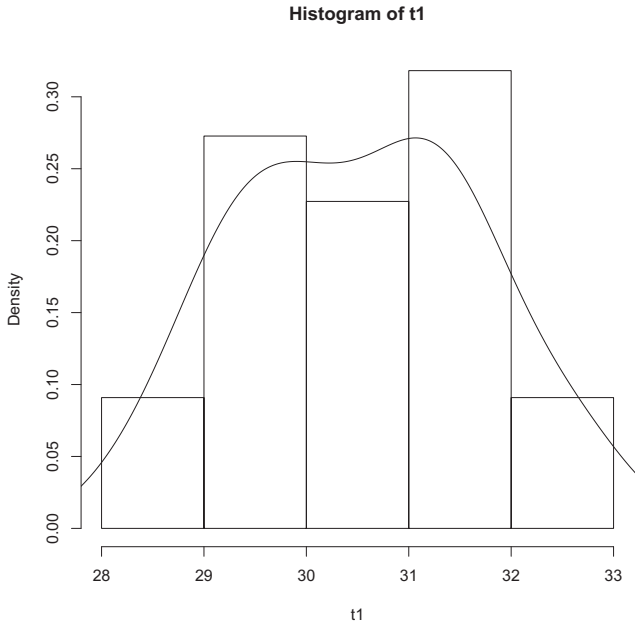
7. The required graph is given below:



8. The resulting graph is given below. The figure shows that for SMaple and WOak, the growing speed in terms of the diameter of the tree is constant, while for ShHickory, when the tree gets older, it grows faster.



9. (a) The basic histogram with smooth curve superimposed:



(b) The stem and leaf plot for the reaction time of Robot 1 is given below. The decimal point is at the |

28 | 4

29 | 0133688

30 | 03388

31 | 0234669

32 | 47

10. The produced basic scatter plot is given below. It seems that the surface conductivity can be used for predicting sediment conductivity.