

INSTRUCTOR'S SOLUTIONS MANUAL

STATISTICS

PRINCIPLES AND METHODS

Eighth Edition

Richard A. Johnson

Gouri K. Bhattacharyya

Richard A. Johnson

PREFACE

This instructor's manual is intended to provide the teacher with solutions to all of the exercises. Many are suitable for digital transfer and projection on a screen to facilitate classroom discussion. We would greatly appreciate receiving your comments, corrections and suggestions for improvements.

K.T. Wu was of tremendous help in checking the solutions and in the preparation of earlier versions of this manuscript.

Richard A. Johnson

Contents

1 INTRODUCTION	1
2 ORGANIZATION AND DESCRIPTION OF DATA	7
3 DESCRIPTIVE STUDY OF BIVARIATE DATA	43
4 PROBABILITY	75
5 PROBABILITY DISTRIBUTIONS	115
6 THE NORMAL DISTRIBUTION	159
7 VARIATION IN REPEATED SAMPLES - SAMPLING DISTRIBUTIONS	201
8 DRAWING INFERENCES FROM LARGE SAMPLES	225
9 SMALL SAMPLE INFERENCES FOR NORMAL POPULATIONS	265
10 COMPARING TWO TREATMENTS	295
11 REGRESSION ANALYSIS -I Simple Linear Regression	337
12 REGRESSION ANALYSIS - II Multiple Linear Regression And Other Topics	369
13 ANALYSIS OF CATEGORICAL DATA	389
14 ANALYSIS OF VARIANCE(ANOVA)	417
15 NONPARAMETRIC INFERENCE	439

Chapter 1

INTRODUCTION

- 1.1 The *statistical population* consists of the entire set of wait/don't wait responses from all men in the United States while the *sample* consists of the responses of the 451 men contacted in the survey.
- 1.2 The sample is the collection of 150 number of passwords memorized by the visitors who agreed to participate. One specification of the statistical population is the collection of the number of passwords memorized by each visitor to the site.
- 1.3 The *statistical population* consists of the answers of all college students when asked their number of close friends. The *sample* is the collection of number of close friends reported by the twenty students who responded.
- 1.4 The *population* consists of the ratings of all tweets. The *sample* is to 4,220 rated tweets from the article.
- 1.5 The *statistical population* is the collection of yes/no answers to the question are you more stressed than last year. The *sample* is the responses of the 40 adults.
- 1.6 The *statistical population* is the rankings of 'favorite types of music' from all students attending a large U.S. university. The *sample* is the rankings from 56 students of a certain large U.S. university
- 1.7 Probably not a random sample of dog ownership among the cities' residents. People who belong to hiking clubs tend to participate in many outdoor activities that are conducive to dog ownership.
- 1.8 (a) Anecdotal. This is the result of an isolated action, sufficient data are not given to support this statement. item[(b)] Sample-based. The sample is the yes/no espresso-based purchase for the 47 drinks

recorded. The orders of a small group selected, hopefully at random, from a large group were used to make this observation.

- (c) Sample-based. The sample is the 200 yes/no answers to thinking about food more than 17 times a day.
- 1.9 (a) This number was recorded from members of a class but one not randomly selected. It is likely recalled because the number is unusually large.
- (b) Anecdotal. Just the opinion of one person.
- (c) Sample-based. The sample is the 4837 yes/no answers to whether they used their cell phone while driving.
- 1.10 Answers will vary. From Table 1, Appendix B, we start reading down starting in row 26 of columns 9 and 10. We obtain

57 19 32 29 21 91 8 75 72 62 58 85 41 3

17 7 60 86 3 15 89 41 93 84 23 87 78 1

Bicycles 1, 3, 7 and 8 are selected. It would be more efficient to use 1,21,41,61 and 81 to represent the first bicycle and so on. This would result in 12, 9, 1 and 11 being selected in the first five numbers.

- 1.11 Answers will vary. It is simpler to select 6 persons who will not go on the bus. Number the students from 1 to 50. In Table 1, we started in row 51 using columns 9 and 10. Reading downward, and ignoring 00 and numbers above 50, we selected students 23, 1, 44, 37, 19, and 26.
- 1.12 The term “on-time” is not well defined. This is not a yes/no question. Presumably, the bus reaches your stop multiple times per day. Moreover, a range of times should be specified for which the bus will be characterized as being on-time. If it was one second late according to your watch, you would likely not characterize it as being late.
- Purpose:* Determine the percentage of time this semester the campus bus reaches your stop within one minute of the scheduled time
- 1.13 The notion of “comfortable” is not well defined. It is different for different people and it will depend on hand size and the position of controls. One improved statement of purpose is:
- Purpose:* Determine if over half the customers prefer a new style mouse to the one they currently use.
- 1.14 *Purpose* Determine the first choice of the campus population as their favorite campus pizza establishment.

- 1.15 *Purpose:* Determine the amount of time it takes those who use the Internet to make hotel reservations in San Francisco.
- 1.16 1.16 One-third of the calls took over 125 minutes to return (11 of 29). Seventeen of 29 took over 90 minutes to return.
- 1.17 At the lab, receptionist and x-ray.
- 1.18 (a) The variable of interest is whether or not a student can identify fake news
- (b) The statistical population is the collection of yes/no responses answers to whether or not the student can identify fake news with one for each college student in the country (or from those colleges included in study).
- (c) The sample is the collection of 2,300 yes/no answers from students included in the survey.
- 1.19 (a) A student at your college is the *unit*.
- (b) The variable of interest is the total monthly entertainment expenses.
- (c) The *statistical population* is the set of data consisting of the total monthly entertainment expenses for each student at your school.
- 1.20 (a) A person living in Chicago is the *unit*.
- (b) The *variable of interest* is the characterization of a Chicago resident as eligible to vote or not.
- (c) The *statistical population* consists of the collection of voter eligibility characterizations of each resident of the city of Chicago.
- 1.21 (a) The *statistical population* consists of the height measurements of male students on campus. The *sample* consists of the height measurements of the members of the basketball team.
- (b) The sample is likely to be non-representative of the population, as basketball players tend to be much taller than the typical student
- (c) There are many ways to choose a sample. One method is to use a table of random digits to select names from a student directory.
- 1.23 No, because a self-selection bias is likely to exist since only people who are interested in this particular exam are likely to answer, and such people perceive such a problem with the values.

- 1.24 The *statistical population* consists of the entire collection of yes/no responses to program purchase from every reader while the *sample* consists of the yes/no responses from those readers who actually sent in the completed form. This is apt to be a very biased sample because persons who have not purchased the program are not as likely to take the time to fill out the form and send it in as readers who have purchased the program
- 1.25 The newspaper is suggesting that the *statistical* is the collection of preferences for each adult in the city while the *sample* is the collection of preferences of the particular persons who sent in their votes. This sample is apt to be nonrepresentative because those persons in the sample are self-selected. Only the few who feel very strongly positive will likely send in a vote.
- 1.26 (a) Sample-based. The sample is the characteristic of lying regularly or not lying regularly reported by the 200 students.
 (b) Anecdotal. Data are not given to support this statement. item[(c)] Sample-based. The sample is the purchase/not purchase data for the 50 persons interviewed.
- 1.27 (a) Anecdotal. No data given. class was not randomly selected.
 (b) Sample-based. The yes/no answer regarding multiple credit cards, for each of the 22 students, is the sample on which the statement is based.
 (c) Sample-based. The yes/no answer regarding destination outside the continental United States, for each of the 55 people at the airport, is the sample on which the statement is based.
- 1.28 The term "too long" is not well defined. By asking a number of people, we may determine that 5 minutes is too long. Further, the time will not be the same for all people. One improved statement of the purpose is
Purpose: Determine if over half the customers take over 5 minutes to get cash during the lunch hour.
- 1.29 Answers vary. First number the boats 0 to 9. In Table 9, we started in row 5 and column 12. Reading downward, we obtain 8 8 9 2 so boats 2, 8, and 9 are selected.
- 1.30 We select 3 of 9 sites using Table 1, Appendix B. Reading across in row 10 starting at column 17, we read one digit at a time and obtain 7 0 6 9. Sites 7, 6 and 9 are the three to visit.

- 1.31 Answers will vary. It is simpler to select 6 persons who will not go on the bus. Number the students from 1 to 50. In Table 1, we started in row 51 using columns 9 and 10. Reading downward, and ignoring 00 and numbers above 50, we selected students 23, 1, 44, 37, 19, and 26.
- 1.32 Answers will vary. We started in row 10 and read down column 9 and then down column 6 from the top ignoring 8 and 9. The 20 pairs of random digits are

4,0	1, 2	4, 2	5, 2	2,1
5,1	3, 2	2,0	7,6	5, 4
0,1	2,5	2,7	3,6	5, 4
2, 4	5,7	3,5	6,7	7, 2

- (a) $4/20 = .20$
- (b) $9/20 = .45$
- (c) $7/20 = .35$
- 1.33 (a) The statement must refer to an average amount per person. Clearly, some persons create much more garbage and others less.
- (b) Most likely from a sample of garbage, the average was 4.44 pounds per person per day. Certainly, the average for the whole population of the United States is unknown.
- (c) You would prefer a nation wide sample. If restricted to households, you could conceivably use random numbers to selected from census listings.
- 1.34 (a) The miniture poodles could never be observed even if the greatly outnumber the Great Danes. Only the big dogs can volunteer to show they were inside the fence.
- (b) Persons who call-in their opinions are self selected because they have strong opinions. This is analogous to the big dogs who are the only volunteers to show they were inside the fence.