# INSTRUCTOR'S SOLUTIONS MANUAL

PAMELA OMER

*Western New England University*

# ESSENTIAL STATISTICS: EXPLORING THE WORLD THROUGH DATA

THIRD EDITION

Robert Gould

*University of California, Los Angeles*

Rebecca Wong

*West Valley College*

Colleen Ryan

*Moorpark Community College*





Reproduced by Pearson from electronic files supplied by the author.



ISBN-13: 978-0-13-576024-6
ISBN-10: 0-13-576024-0

CONTENTS

## Chapter 6: Modeling Random Events: The Normal and Binomial Models



## Chapter 7: Survey Sampling and Inference



## Chapter 8: Hypothesis Testing for Population Proportions



## Chapter 9: Inferring Population Means



## Chapter 10: Analyzing Categorical Variables and Interpreting Research

# Chapter 1: Introduction to Data

## Section 1.2: Classifying and Storing Data

1.1 There are eight variables: “Female”, “Commute Distance”, “Hair Color”, “Ring Size”, “Height”, “Number of Aunts”, “College Units Acquired”, and “Living Situation”.

1.2 There are eleven observations.

1.3 a. Living situation is categorical.

b. Commute distance is numerical.

c. Number of aunts is numerical.

1.4 a. Ring size is numerical.

b. Hair color is categorical.

c. Height is numerical.

1.5 Answers will vary but could include such things as number of friends on Facebook or foot length. *Don’t copy these answers*.

1.6 Answers will vary but could include such things as class standing (“Freshman”, “Sophomore”, “Junior”, or “Senior”) or favorite color. *Don’t copy these answers*.

1.7 0 = male, 1 = female. The sum represents the total number females in the data set.

1.8 There would be seven 1’s and four 0’s.

1.9 Female is categorical with two categories. The 1’s represent females, and the 0’s represent males. If you added the numbers, you would get the number of females, so it makes sense here.

1.10 a. Freshman

| |
|---|
| 0 |
| 1 |
| 1 |
| 0 |
| 1 |
| 1 |
| 0 |
| 1 |
| 1 |
| 0 |
| 0 |

b. numerical

c. categorical

1.11 a. The data is stacked.

b. 1 = male, 0 = female.

c.

| Male | Female |
|---|---|
| 1916 | 9802 |
| 183 | 153 |
| 836 | 1221 |
| 95 | |
| 512 | |

1.12 a. The data is unstacked.

b. Labels for columns will vary.

| Gender | Age |
|---|---|
| 1 | 29 |
| 1 | 23 |
| 1 | 30 |
| 1 | 32 |
| 1 | 25 |
| 0 | 24 |
| 0 | 24 |
| 0 | 32 |
| 0 | 35 |
| 0 | 23 |

c. Gender is categorical; Age is numerical

1.13 a. Stacked and coded:

| Calories | Sweet |
|---|---|
| 90 | 1 |
| 310 | 1 |
| 500 | 1 |
| 500 | 1 |
| 600 | 1 |
| 90 | 1 |
| 150 | 0 |
| 600 | 0 |
| 500 | 0 |
| 550 | 0 |

The second column could be labeled "Salty" with the 1's being 0's and the 0's being 1's.

b. Unstacked:

| Sweet | Salty |
|---|---|
| 90 | 150 |
| 310 | 600 |
| 500 | 500 |
| 500 | 550 |
| 600 | |
| 90 | |

1.14 a. Stacked and coded:

| Cost | Male |
|---|---|
| 10 | 1 |
| 15 | 1 |
| 15 | 1 |
| 25 | 1 |
| 12 | 1 |
| 8 | 0 |
| 30 | 0 |
| 15 | 0 |
| 15 | 0 |

The second column could be labeled "Female" with the 1's being 0's and the 0's being 1's.

b. Unstacked:

| Male | Female |
|---|---|
| 10 | 8 |
| 15 | 30 |
| 15 | 15 |
| 25 | 15 |
| 12 | |

## Section 1.3: Investigating Data

1.15 Yes. Use College Units Acquired and Living Situation.

1.16 Yes. Use Female and Height.

1.17 No. Data on number of hours of study per week are not included in the table.

1.18 Yes. Use Ring Size and Height.

1.19 a. Yes. Use Date.

b. No. data on temperature are not included in the table.

c. Yes. Use Fatal and Species of Shark.

d. Yes. Use Location.

1.20 Use Time and Activity.

## Section 1.4: Organizing Categorical Data

1.21 a. 33/40 = 82.5%

b. 32/45 = 71.1%

c. 33/65 = 50.8%

d. 82.5% of 250 = 206 men

1.22 a. 4/27 = 14.8%

b. 14/27 = 51.9%

c. 4/18 = 22.2%

d. 14.8% of 600 = 89 men

1.23 a. 15/38 = 39.5% of the class were male.

b. 0.641(234) = 149.994, so 150 men are in the class.

c. $0.40(x) = 20$, so 20/0.40 = 50 total students in the class.

1.24 a. 0.35(346) = 121 male nurses.

b. 66/178 = 37.1% female engineers.

c. $0.65(x) = 169$ so 169/0.65 = 260 lawyers in the firm.

1.25 The frequency of women 6, the proportion is 6/11, and the percentage is 54.5%.

1.26 The frequency is 8, the proportion is 8/11, and the percentage is 72.7%.

1.27 a. and b.

| | **Men** | **Women** | **Total** |
|---|---|---|---|
| Dorm | 3 | 4 | 7 |
| Commuter | 2 | 2 | 4 |
| Total | 5 | 6 | 11 |

c. 4/6 = 66.7%

d. 4/7 = 57.1%

e. 7/11 = 63.6%

f. 66.7% of 70 = 47

1.28 a. and b.

| | **Men** | **Women** | **Total** |
|---|---|---|---|
| Brown | 3 | 5 | 8 |
| Black | 2 | 0 | 2 |
| Blonde | 0 | 1 | 1 |
| Total | 5 | 6 | 11 |

c. 5/6 = 83.3%

d. 5/8 = 62.5%

e. 8/11 = 72.7%

f. 83.3% of 60 = 50

1.29 1.26(x) = 160328 so 160328/1.26 = 127,244 personal care aids in 2014

1.30 .1295(x) = 3480000 so 3480000/.1295 = $26,872,587.87 total candy sales

1.31

| State | Prison | Rank Prison | Population | Population (thousands) | Prison per 1000 | Rank Rate |
|---|---|---|---|---|---|---|
| California | 136,088 | 1 | 39,144,818 | 39145 | 3.48 | 4 |
| New York | 52518 | 2 | 19,795,791 | 19796 | 2.65 | 5 |
| Illinois | 48278 | 3 | 12,859,995 | 12860 | 3.75 | 3 |
| Louisiana | 30030 | 4 | 4,670,724 | 4671 | 6.43 | 1 |
| Mississippi | 18793 | 5 | 2,992,333 | 2992 | 6.28 | 2 |

California has the highest prison population. Louisiana has the highest rate of imprisonment.

The two answers are different because the state populations are different.

1.32 a. Miami: 4,919,000/2891 = 1701 Detroit: 3,903,000/3267 = 1195

Atlanta: 3,500,000/5083 = 689 Seattle: 2,712,000/1768 = 1534

Baltimore: 2,076,000/1768 = 1174

Ranks: 1- Miami, 2- Seattle, 3- Detroit, 4- Baltimore, 5- Atlanta

b. Atlanta

c. Miami

1.33

| Year | %Uncovered |
|---|---|
| 1990 | $\frac{34,719}{249,778} = 13.9\%$ |
| 2000 | $\frac{36,586}{279,282} = 13.1\%$ |
| 2015 | $\frac{29758}{316574} = 9.4\%$ |

The percentage of uninsured people have been declining.

1.34

| Year | % Subscribers |
|---|---|
| 2012 | $\frac{103.6}{114.7} = 90.3\%$ |
| 2013 | $\frac{103.3}{114.1} = 90.5\%$ |
| 2014 | $\frac{103.7}{115.7} = 89.6\%$ |
| 2015 | $\frac{100.2}{116.5} = 86.0\%$ |
| 2016 | $\frac{97.8}{116.4} = 84.0\%$ |

The percentage of cable subscribers rose slightly between 2012 and 2013 but has declined each year since then.

1.35

| Year | %Older Population |
|---|---|
| 2020 | $\frac{54.8}{334} = 16.4\%$ |
| 2030 | $\frac{70.0}{358} = 19.6\%$ |
| 2040 | $\frac{81.2}{380} = 21.4\%$ |
| 2050 | $\frac{88.5}{400} = 22.1\%$ |

The percentage of older population is projected to increase.

1.36

| Year | D/M % |
|---|---|
| 2000 | $\frac{4.0}{8.2} = 48.8\%$ |
| 2005 | $\frac{3.6}{7.6} = 47.4\%$ |
| 2010 | $\frac{3.6}{6.8} = 52.9\%$ |
| 2014 | $\frac{3.2}{6.9} = 46.4\%$ |

The rate has fluctuating over this period, decreasing, then increasing, and then decreasing again.

1.37 We don't know the percentage of female students in the two classes. The larger number of women at 8a.m. may just result from a larger number of students at 8 a.m., which may be because the class can accommodate more students because perhaps it is in a large lecture hall.

1.38 No, we need to know the population of each city so we can compare the rates.

## Section 1.5 Collecting Data to Understand Causality

1.39 Observational study.

1.40 Controlled experiment.

1.41 Controlled experiment.

1.42 Controlled experiment.

1.43 Controlled experiment.

1.44 Observational study.

1.45 Anecdotal evidence are stories about individual cases. No cause-and effect conclusions can be drawn from anecdotal evidence.

1.46 These testimonials are anecdotal evidence. There is no control group and no comparison. No cause-and-effect conclusions can be drawn from anecdotal evidence.

1.47 This was an observational study, and from it you cannot conclude that the tutoring raises the grades. Possible confounders (answers may vary): 1. It may be the more highly motivated who attend the tutoring, and this motivation is what causes the grades to go up. 2. It could be that those with more time attend the tutoring, and it is the increased time studying that causes the grades to go up.

1.48 a. If the doctor decides on the treatment, you could have bias.

b. To remove this bias, randomly assign the patients to the different treatments.

c. If the doctor knows which treatment a patient had, that might influence his opinion about the effectiveness of the treatment.

d. To remove that bias, make the experiment double-blind. The talk-therapy-only patients should get a placebo, and no patients should know whether they have a placebo or antidepressant. In addition, the doctor should not know who took the antidepressants and who did not.

1.49 a. The sample size of this study is not large (40). The study was a controlled experiment and used random assignment. It was not double-blind since researchers new what group each participant was in.

b. The sample size of the study was small, so we should not conclude that physical activity while learning caused higher performance.

1.50 This is an observational study because researchers did not determine who received PCV7 and who did not. You cannot conclude causation from an observational study. We must assume that it is possible that there were confounding factors (such as other advances in medicine) that had a good effect on the rate of pneumonia.

1.51 a. Controlled experiment. Researchers used random assignment of subjects to treatment or control groups.

b. Yes. The experiment had a large sample size, was controlled, randomized, and double-blind; and used a placebo.

1.52 a. Observational study. There was no random assignment to treatment/control groups. The subjects kept a food diary and had their blood drawn.

b. We cannot make a cause-and-effect conclusion since this was an observational study.

1.53 No, this was not a controlled experiment. There was no random assignment to treatment/control groups and no use of a placebo.

1.54 No. There was no control group and no comparison. From observation of 12 children it is not possible to come to a conclusion that the vaccine causes autism. It may simply be that autism is usually noticed at the same age the vaccine is given.

1.55 a. Intervention remission: 11/33 = 33.3%; Control remission: 3 /34 = 8.8%

b. Controlled experiment. There was random assignment to treatment/control groups.

c. While this study did use random assignment to treatment/control groups, the sample size was fairly small (67 total) and there was no blinding in the experimental design. The difference in remission may indicate that the diet approach is promising and further research in this area is needed.

1.56 Ask whether there was random assignment to groups. Without random assignment there could be bias, and we cannot infer causation.

1.57 No. This is an observational study.

1.58 This is likely a conclusion from observational studies since it would not be ethical to randomly assign a subject to a group that drank large quantities of sugary drinks. Since this was likely based on observational studies, we cannot conclude drinking sugary beverages causes lower brain volume.

## Chapter Review Exercises

1.59 a. 61/98 = 62.2%

b. 37/82 = 45.1%

c. Yes, this was a controlled experiment with random assignment. The difference in percentage of homes adopting smoking restrictions indicates the intervention may have been effective.

1.60 No. Cause-and-effect conclusions cannot be drawn from observational studies.

1.61 a. Gender (categorical) and whether students had received a speeding ticket (categorical)

b.

| | **Male** | **Female** |
|---|---|---|
| Yes | 6 | 5 |
| No | 4 | 10 |

c. Men: 6/10=60%; Women: 5/15 = 33.3%; a greater percentage of men reported receiving a speeding ticket.

1.62 a. Gender (categorical) and whether students had driven over 100 mph (categorical).

b.

| | **Male** | **Female** |
|---|---|---|
| Yes | 6 | 5 |
| No | 3 | 10 |

c. Men: 6/9 = 66.7%; Women: 5/15 = 33.3%; a greater percentage of men reported driving over 100 mph.

1.63 Answers will vary. *Students should not copy the words they see in these answers.* Randomly divide the group in half, using a coin flip for each woman: Heads she gets the vitamin D, and tails she gets the placebo (or vice versa). Make sure that neither the women themselves nor any of the people who come in contact with them know whether they got the treatment or the placebo ("double-blind"). Over a given length of time (such as three years), note which women had broken bones and which did not. Compare the percentage of women with broken bones in the vitamin D group with the percentage of women with broken bones in the placebo group.

1.64 Answers will vary. *Students should not copy the words they see here.* Randomly divide the group in half, using a coin flip for each person: Heads they get Coumadin, and tails they get aspirin (or vice versa). Make sure that neither the subjects nor any of the people who come in contact with them know which treatment they received ("double-blind"). Over a given length of time (such as three years), note which people had second strokes and which did not. Compare the percentage of people with second strokes in the Coumadin group with the percentage of people with second strokes in the aspirin group. There is no need for a placebo because we are comparing two treatments. However, it would be acceptable to have three groups, one of which received a placebo.

1.65 a. The treatment variable is mindful yoga participation. The response variable is alcohol use.

b. Controlled experiment (random assignment to treatment/control groups).

c. No, since the sample size was fairly small; however, the difference in outcomes for treatment/control groups may indicate that further research into the use of mindful yoga may be warranted.

1.66 a. The treatment variable was neurofeedback; the response variable is ADHD symptoms.

b. Controlled experiment (random assignment to treatment/control groups).

c. No because there were no significant differences in outcomes between any of the groups.

1.67 No. There was no control group and no random assignment to treatment or control groups.

1.68 a. Long course antibiotics: 39/238 = 16.4%; short course antibiotics: 77/229 = 33.6%.

The longer course recipients did better.

b.

| | **10 days** | **5 days** |
|---|---|---|
| Failure | 39 | 77 |
| Success | 199 | 152 |

c. Controlled experiment (random assignment to treatment/control groups).

d. Yes. This was a controlled, randomized experiment with a large sample size.

1.69 a. LD: 8% tumors; LL: 28% tumors A greater percentage of the 24 hours of light developed tumors.

b. A controlled experiment. You can tell by the random assignment.

c. Yes, we can conclude cause and effect because it was a controlled experiment, and random assignment will balance out potential confounding variables.

1.70 a. 43/53, or about 81.1%, of the males who were assigned to Scared Straight we rearrested. 37/55, or 67.3%, of those receiving no treatment were rearrested. So the group from Scared Straight had a higher arrest rate.

b. No, Scared Straight does not cause a lower arrest rate because the arrest rate was higher.

# Chapter 2: Picturing Variation with Graphs

## Section 2.1: Visualizing Variation in Numerical Data and Section 2.2: Summarizing Important Features of a Numerical Distribution

2.1 a. 4 people had resting pulse rates more than 100.

b. $\frac{4}{125} = 3.2\%$ of the people had resting pulse rates of more than 100.

2.2 a. 8 people have glucose readings above 120 mg/dl.

b. $8 = 6.1\%$ of these people have glucose readings above 120 mg/dl.

2.3 New vertical axis labels: $\frac{1}{25} = 0.04$, $\frac{2}{25} = 0.08$, $\frac{3}{25} = 0.12$, $\frac{4}{25} = 0.16$, $\frac{5}{25} = 0.20$

2.4 a. The bin width is 100.

b. The histogram is bimodal because two bins have a much higher relative frequency than the others.

c. About 19% (combine 6% and 13%). Due to the scale on the graph, any answer between 18% to 20% is acceptable.

2.5 Yes, since only about 7% of the pulse rates were higher than 90 bpm. Conclusion might vary, but students must mention that 7% of pulse rates were higher than 90 bpm.

2.6 No, because on roughly half of the days the post office served more than 250 customers, so 250 would not be unusual.

2.7 a. Both cereals have similar center values (about 110 calories). The spread of the dotplots differ.

b. Cereal from manufacturer K tend to have more variation.

2.8 a. Both distributions have more than one mode. The center for the coins from the United States is much larger than the center for other countries. The spreads are similar.

b. Coins in the United States tend to weigh more, as we conclude because the center of the distribution is higher for the United States coins.

2.9 Roughly bell shaped. The lower bound is 0, the mean will be a number probably below 9, but a few students might have slept quite a bit (up to 12 hours?) which creates a right-skew.

2.10 Roughly right-skewed (most students with no tickets, very few with many tickets).

2.11 It would be bimodal because the men and women tend to have different heights and therefore different arm spans.

2.12 It might be bimodal because private colleges and public colleges tend to differ in amount of tuition.

2.13 About 75 beats per minute.

2.14 About 500 Calories.

2.15 The BMI for both groups are right skewed. For the men it is maybe bimodal (hard to tell). The typical values for the men and women are similar although the value for the men appears just a little bit larger than the typical value for the women. The women's values are more spread out.

2.16 a. Both distributions are right skewed. They have similar typical values.

b. The men's distribution is more spread out and has a greater percentage of values that are considered high. So, the women's levels are somewhat better.

2.17 a. The distribution is multimodal with modes at 12 years (high school), 14 years (junior college), 16 years (bachelor's degree), and 18 years (possible master's degree). It is also left-skewed with numbers as low as 0.

b. Estimate: 300 + 50 + 100 + 40 + 50, or about 500 to 600, had 16 or more years.

c. Between $\frac{500}{2018}$, or about 25%, and $\frac{600}{2018}$, or about 30%, have a bachelor's degree or higher. This is very similar to the 27% given.

2.18 a. The distribution is right-skewed.

b. About 2 or 3.

c. Between 80 and 100.

d. $\frac{80}{2000} = 4\%$ or $\frac{100}{2000} = 5\%$

2.19 Ford typically has higher monthly costs (the center is near 250 dollars compared with 225 for BMW) and more variation in monthly costs.

2.20 Both makes have similar typical mpg (around 23 mpg). BMW has more variation in mpg (more horizontal spread in the data).

2.21 1. The assessed values of homes would tend to be lower with a few higher values: This is histogram B.

2. The number of bedrooms in the houses would be slightly skewed right: This is histogram A.

3. The height of house (in stories) for a region would be that allows up to 3 stories would be histogram C.

2.22 1. The consumption of coffee by a person would be skewed right with many people who do not drink coffee and a few who drink a lot: This is histogram A.

2. The maximum speed driven in a car would be roughly symmetrical with a few students who drive very fast: This is histogram C.

3. The number of times a college student had breakfast would skew left with students who rarely eat breakfast: This is histogram B.

2.23 1. The heights of students would be bimodal and roughly symmetrical: This is histogram B.

2. The number of hours of sleep would be unimodal and roughly symmetrical, with any outliers more likely being fewer hours of sleep: This is histogram A.

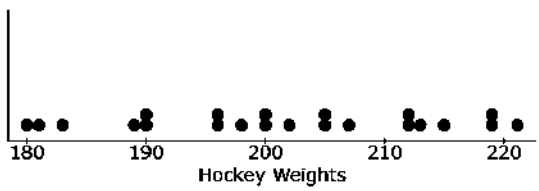
3. The number of accidents would be left skewed, with most student being involved in no or a few accidents: This is histogram C.

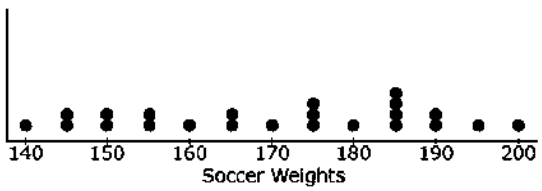
2.24 1. The SAT scores would be unimodal and roughly symmetrical: This is histogram C.

2. The weights of men and women would be bimodal and roughly symmetrical, but with more variation that SAT scores: This is histogram A.

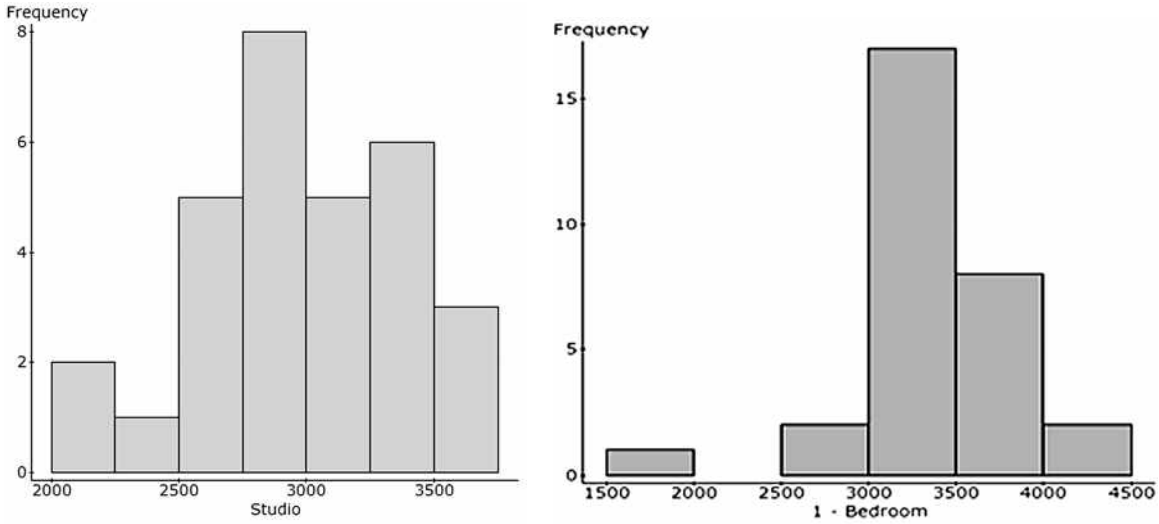
3. The ages of students would be left skewed, with most student being younger: This is histogram B.

2.25 Students should display a pair of dotplots or histograms. One graph for Hockey and one for Soccer. The hockey team tends to be heavier than the soccer team (the typical hockey player weighs about 202 pounds while the typical soccer player weighs about 170 pounds). The soccer team has more variation in weights than the hockey team because there is more horizontal spread in the data. Statistical Question (answers may vary): Are hockey players heavier than soccer players? Which type of athlete has the most variability in weight?



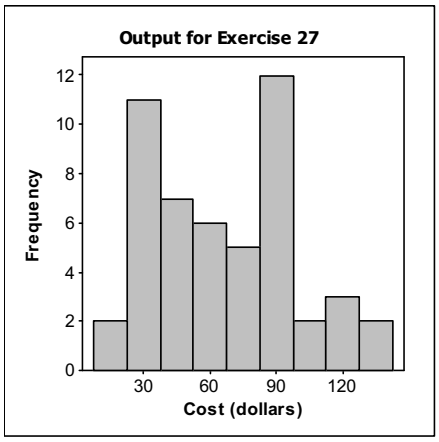



2.26 (Answers may vary). Which type of apartment tends to cost more? See histograms.




Studio apartments tend to be less expensive and have more variation in price than do one-bedrooms.

2.27 See histogram. The shape will depend on the binning used. The histogram is bimodal with modes at about \$30 and about \$90.




2.28 See histogram. The shape will depend on the binning used. The 800 score could be an outlier or not, and the graph could appear left-skewed or not.