Solutions to the exercises, specified in the Stat 1600 ed. 2017-2018

Loren L. Heun

August 8, 2018

Knowledge and Data

Solution 1.6-1: 1. What might be wrong about these headlines? a. A study proclaims: “Slightly overweight people live longer than thin people.” “Slightly overweight people live longer than thin people.” It is a fallacy of correlation equals causation. It statement implies that being slightly overweight causes people to live longer when compared to thin people. We would want to dig deeper into the study to see what confounding variables may be present: did we take into account health of the individuals? Perhaps we have some individuals in poor health leading them to be thin and shortening their lives.

Solution 1.6-3: No randomization; the product was given to several people; a skin specialist should make the determination as to the effectiveness of the product; the company should not mass produce this anti-wrinkle cream.
Data Presentation

Solution 2.7-1:

1. numeric, ratio
2. categorical, nominal
3. numeric, ratio
4. categorical, nominal
5. numeric, interval
6. categorical, ordinal
7. categorical, ordinal
8. categorical, ordinal
9. numeric: ratio
10. categorical, ordinal
11. categorical, nominal
12. categorical, ordinal
13. numeric, interval

Solution 2.7-3: Compute relative frequencies:

Interval	Frequency	Relative Frequency
$(20, 25]$	4	4
$(25, 30]$	11	11
$(30, 35]$	23	23
$(35, 40]$	31	31
$(40, 45]$	15	15
$(45, 50]$	10	10
$(50, 55]$	6	6

Solution 2.7-4: 2.5-7.1: Left-skewed
2.5-7.2: Symmetric
2.5-7.3: Left-skewed
2.5-7.4: Right-skewed
2.5-7.5: Symmetric

Solution 2.7-5: The level of measurement of marital status is nominal.

Solution 2.7-6: The level of measurement of marital status is category - nominal.

Solution 2.7-7: The number of reported robberies in June 2014 is Kalamazoo County is a continuous variable.

Solution 2.7-8: Numeric - interval

Solution 2.7-9: Numeric - ratio

Solution 2.7-10: all adult residents of the U.S.

Solution 2.7-11: birth weight.

Solution 2.7-12: Variable: Homicide rate
Level of Measurement: Interval-ratio
Type: Continuous
Application: Descriptive (two variables)

Solution 2.7-13: Variable: party, gender, opinion
Level of Measurement: nominal, nominal, ordinal
Type: discrete, discrete, discrete
Application: inferential, NA, NA

Solution 2.7-14: All of the choices is the correct answer.
Location and Spread

Solution 3.4-1: Compute the mean, median, ... for carbon monoxide:

1. mean = 12.5
2. median = 13.0
3. Trimmed mean = 12.7
4. SD = 4.74
5. Mean = 0.0125 and standard deviation 0.00474 (in grams)

Solution 3.4-3:

1. The range is from 13 to 59
2. the mean is 46.7
3. The median is 49.5
4. There is no mode since there are no duplicates.
5. Removing the smallest value changes the mean to 50.4. The new value is closer to median because 13 is an outlier.

Solution 3.4-1:

Solution 3.4-5:

1. $\bar{x} = 271.6$
2. $\tilde{x} = 265.5$
3. $sd = 154.3727092$
4. The distribution is Right skewed
5. The median provided the most relevant information because to the extreme outlier.

Solution 3.4-6: Compute the mean, median, and mode for the weekly grocery budget mean = 32.167, median = 32.5, mode = 35

Solution 3.4-7: The third value must be 22 so that the average is 20 for all three quizzes.

Solution 3.4-8: The mean is 472.59875 The median is 555 There is no mode since there are no duplicates.
1. The greater value is the median (555). **2.** There is a negative skewness in the middle half the of the dataset. (Refer to Figure 3.1.)
3. When we compare the median to the mean, we find the median (555) is greater the mean (472.6). Therefore, we can say the the dataset is left skewed.

Solution 3.4-9: **(1.)** Mode **(2.)** Median **(3.)** Mean **(4.)** Mean **(5.)** Median **(6.)** Mode

Solution 3.4-10: **Birth** Mode="North"; **Expense** Mean=48.5; **Movies** Mean=5.8; **Food** Median=6; **Religion** Mode="Protestant"

Solution 3.4-11: The mean for 2005 is 55.4. The median for 2005 is 54.5. The mean for 2015 is 57.1. The median for 2015 is 57.

Solution 3.4-12: The mean is [1] 31.8
The median is [1] 35

Solution 3.4-13: The mean is 28.72, and the median is 30

Solution 3.4-14: The pretest mean is 9.3333333
The pretest median is 10
The posttest mean is 12.9333333
The pretest median is 12
Threats to Valid Comparisons

Solution 4.4-1: $H_0 : \mu = 6.2$ vs. $H_A : \mu \neq 6.2$

Solution 4.4-3: From the distribution of t, using row df = 24 and column 0.025, CV(t) = 2.064.
Study Designs

Solution 5.5-2: Randomized controlled experiments are necessary to establish the cause of such a disease firmly. For instance, we might see if potassium deficiency causes the disease, we may give our treatment group a potassium supplement every day. However, if the illness is rare, then it is impractical for us to use a randomized controlled experiment. It would also be unethical to be influencing or causing disease in people potentially. Additionally, we would have no guarantee that any subjects would even develop the disease if it is sporadic. Thus, observational studies may be the most appropriate choice unless scientific advances in, say, genetics are far enough along to make a more practical kind of randomized controlled experiment.

Solution 5.5-4: A randomized controlled experiment is the "gold standard of scientific knowledge" so that it would be the best way to study this situation. We would study four groups of patients: group one would get a standard treatment for an illness, group two would get a placebo without being told it's a placebo, group 3 would get a placebo and told it's a placebo, and group 4 would get no treatment at all. This method would, interestingly, give controls both from the standard and no-treatment groups. The groups for primary comparisons would be the two groups that get the placebos and are either informed or not.

Solution 5.5-6: Yes, there are potential confounders. For instance, it might be that players feel better rested at the start of a game, so they play harder and get injured more frequently because of this possibility. The organization could investigate the frequency of injuries that occur during later kickoffs in the game or to conduct small-scale randomized controlled experiments that change kickoff rules before recommending substantial sweeping changes.
The Normal Distribution

Solution 6.7-1:

1. The proportion of cell phone users are on their phones between 1 hour and 3 hours per day is 0.6057221
2. Just to be safe, suppose you decide to be in the 5th percentile of cell phone users in terms of monthly usage. How much time can you spend on your phone per day? You decide to spend no more than 35.4396606 minutes on your phone per day.

Solution 6.7-3: The standard deviation is $SD = \frac{35-4}{4} = 7.75$

Solution 6.7-5:

1. The probability that the stock price is between 39.88 and 46.01 is $P[39.88 < X < 46.01] = 0.9395927$.
2. The probability that the stock price is above 40 is $P[X > 40] = 0.9327388$.
3. The probability that the stock price is below 40 si $P[x < 40] = 0.0672612$.

Solution 6.7-7:

1. The area to the left of 0.0 is.5000
2. The area to the left of 0.2 is.5793
3. The area to the left of 0.25 is.5987
4. The area to the left of 2.25 is.9878

Solution 6.7-8:

$$SD = \frac{(41-32)}{3} = 3 \tag{1}$$

Solution 6.7-9:

$$z = \frac{173-153}{11} = 1.82 \tag{2}$$

Solution 6.7-10: $z = \frac{143-155}{12} = -1$
$P[z < -1] = 15.87$

Solution 6.7-11: $x = \mu + Z\sigma$
$x = 155 + 1.28 \times (12)$
$x = 170.36$

Solution 6.7-12: $x = \mu + Z\sigma$
$x = 155 - 0.6745 \times (12)$
$x = 146.91$

Solution 6.7-13: The value of the 15^{th} percentile is 4.9585388.
There are six countries: Congo (4.1), Indonesia (3.1), Saudi Arabia (3.2), Venezuela (3.6), India (4.0), Malaysia (4.0), and Thailand (4.6).

Solution 6.7-14: The probability of having a mean less than 7.3 ($P[\bar{x} < 7.3]$) is 0.3782144.

Solution 6.7-15: The probability of having a mean greater than 9.3 ($P[\bar{x} > 9.3]$) is 0.6217856.

Solution 6.7-16: The probability of having a mean between 7.3 and 9.3 ($P[7.3 < \bar{x} < 9.3]$) is 0.2435711.
The Binomial Distribution

Solution 7.8-1:

1. The mean and standard deviation are 5 and 1.5811388, respectively.
2. The probability that there are more than 5 successes is $P[X > 5] = 0.3769531$.
3. The probability that there are fewer than 5 successes is $P[X < 5] = 0.3769531$.
4. The probability that there are between 1 and 3 successes is $P[1 \le X \le 5] = 0.1708984$.

Solution 7.8-3:

1. The probability that at least 100 people are Apple users is $P[X \ge 100] = 0.3070581$.
2. The probability that at most 100 people are Apple users is $P[X \le 100] = 0.7302684$.
3. The probability that between 80 and 120 people are Apple users is $P[80 \le X \le 120] = 0.9572505$.

Solution 7.8-5:

1. The probability that he gets at least 7 hits is $P[X \ge 7] = 0.0048184$.
2. The probability that he gets at most 1 hit is $P[X \le 1] = 0.1461307$.
3. The probability that he gets between 4 and 6 hits is $P[4 \le X \le 6] = 0.3245785$.

Solution 7.8-7: The expected value is 8 and SD is 2.7712813

Solution 7.8-8:

1. $P[x \ge 33] = 0.2352$
2. $P[x \le 25] = 0.0970$

Sampling Distribution of the Proportion

Solution 8.7-1: $P[X \ge 32] = 0.0760321$

Solution 8.7-3: The standard error of this estimate is 0.04

Solution 8.7-5:

1. The estimate of the population proportion is $\hat{p} = 0.2583333$
2. The standard error of this estimate is $SE_{\hat{p}} = 0.039958$
3. The 95% confidence interval is 0.2583333 ± 0.0783177

Solution 8.7-7:

1. The estimate of the population proportion is $\hat{p} = 0.3$
2. The standard error of this estimate is $SE_{\hat{p}} = 0.010247$
3. The margin of error the estimate is $M_{\hat{p}} = 0.020084$
4. The 95% confidence interval is 0.3 ± 0.020084
5. The 95% confidence interval is 0.3 ± 0.020084

Solution 8.7-8: The standard error of this estimate is 97

Comparing Two Proportions

Solution 9.5-1:

1. The difference in the percentage of drug use between smokers and nonsmokers is 35.0.
2. Calculate a standard error for our estimate in (1).
3. Calculate a 95% confidence interval for the difference in the percentage of drug use between smokers and nonsmokers.
4. Estimate the risk ratio of drug use between smokers and nonsmokers.
5. Calculate a standard error for the natural log of our estimate in 4.
6. Calculate a 95% confidence interval for the risk ratio of drug use between smokers
7. Estimate the odds ratio of drug use between smokers and nonsmokers.
8. Calculate a standard error for the natural log of our estimate in 7.
9. Calculate a 95% confidence interval for the odds ratio of drug use between smokers and nonsmokers.
10. Interpret the above confidence intervals in parts 3, 6, and 9. Which are significant, and which are not? Why or why not?

Solution 9.5-3: The critical value is 1.984217

Solution 9.5-5:

1. Conclude that there is a difference since zero is not within the interval that the difference in the proportion of men and women who think that it is wrong for married people to have an affair.
2. SE = 0.0243
3. RR = 1.0581245
4. OR = 1.5752837

Sampling Distribution of the Mean

Solution 3.4-1:

Solution 10.7-2:

1. The standard error of this estimate is $SE = \frac{2}{\sqrt{16}} = 0.5$
2. The 95% margin of error is $ME = 1.96 \times SE = 1.96 \times 0.5 = 0.98$
3. The 95% CI is $\bar{x} \pm ME = 10.5 \pm 0.98 = (9.52, 11.48)$
4. $$\begin{aligned} P[\bar{x} > \mu] &= P[z > z^*] \\ P[z > z^*] &= P\Big[z > \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\Big] \\ &= P\Big[\frac{z > 9-10}{2/\sqrt{16}}\Big] \\ &= P\Big[z > \frac{-1}{0.5}\Big] \\ &= P[z > -2] \end{aligned} =$$ $1 - 0.025 = 0.975$
5. $$\begin{aligned} P[\bar{x} < \mu] &= P[z < z^*] \\ P[z < z^*] &= P\Big[z < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\Big] \\ &= P\Big[z < \frac{9-10}{2/\sqrt{16}}\Big] \\ &= P\Big[z < \frac{-1}{0.5}\Big] \\ &= P[z < -2] \end{aligned} =$$

0.025

6.

$$
\begin{aligned}
P[8<\bar{x}<10] &= P[z^{*}<z<z^{**}] \\
P[z<z^{*}] &= P\Big[z<\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\Big] \\
&= P\Big[z<\frac{8-10}{2/\sqrt{16}}\Big] \\
&= P\Big[z<\frac{-2}{0.5}\Big] \\
&= P[z<-4] \\
P[z<z^{*}] &= P\Big[z<\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\Big] \\
&= P\Big[z<\frac{10-10}{2/\sqrt{16}}\Big] \\
&= P\Big[z<\frac{-2}{0.5}\Big] \\
&= P[z<0]P[z^{*}<z<z^{**}] \\
P[-4<z<0] & \\
&= P[z<0]-P[z<-4]\approx 0.5-0 \\
&= 0.5
\end{aligned}
$$

7. 33% of sample means are above what value?
8. 33% of sample means are below what value?

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1: Comparing Two Means

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 11.5-7:

1. $\bar{X}_d = 0.4$
2. $SE_d = 0.4$
3. $95\% CI \to (-0.4525798, 1.2525798)$
4. Does the data support the claim that the new material gives superior wear? No, since zero is contained in the 95% CI.

Categorical Variables: Association or Independence

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

$=$

Solution 3.4-1:

Solution 3.4-1: Correlation

Solution 13.4-1: 13.5-1.1: $r = 0.9527$
13.5-1.2: same, $r = 0.9527$
13.5-1.3: same, $r = 0.9527$
13.5-1.4: same, $r = 0.9527$
13.5-1.5: same but the sign changed, $r = -0.9527$

Solution 13.4-2: $r = 1$ because the points (x, y) all fall perfectly along a line tilted upward.

Solution 3.4-1:

Solution 3.4-1:

Solution 13.4-5: $-1 < r < 0$. We know r is negative because the line of best fit is tilted downward. We know r is not -1 nor 0 because the points are clustered, but do not perfectly fall on, the line of best fit.

Solution 3.4-1:

Solution 13.4-7: Since $r = \frac{\sum Z_x Z_y}{(n-1)}$, and we know that all the products of Z scores are negative, we know that r must be -1r¡0. Since $r = \frac{\sum Z_x Z_y}{(n-1)}$, and we know that all the products of Z scores are negative, we know that r must be $-1 \leq r < 0$.
Linear Regression

Solution 14.7-1:

1. Calculate the regression line for predicting Y from X.
2. Draw the scatterplot with an overlaid regression line.
3. Add 5 to Y, so the new values are 5, 8, 15, 6, 20. Calculate the new regression line.
4. Multiply Y by 5, so the new values are 0, 15, 50, 5, 75. Calculate the new regression line.

Solution 3.4-1:

Solution 3.4-1:

Solution 14.7-4: $BAC = -0.012 + 0.017(9) = 0.14$

Solution 14.7-5: A moderately strong positive straight-line relationship between number of beers and BAC.

Solution 14.7-6: The correllation coefficient is $r = 0.875 = \sqrt{0.765}$ where r^2 is the coefficient of determination.

Solution 14.7-7: The independent variable is HEIGHT and the dependent variable is rincom06.

Solution 14.7-8: $H_0 : \beta = 0" vs.\mathrm{H}_1 : \beta \neq 0$

Solution 14.7-9: The correlation coefficient between the respondent's height and income is 0.0819404

Solution 3.4-1:

[width=6cm]figure/f3$_1 - 1.pdf$

Figure 1: Boxplot of Cars per 1000

Solution 14.7-11: For a sample of 40 subjects, we tested the relationship between height and income. We found a weak positive relationship using the correlation coefficient (0.2862524). We then looked at the slope: as *height* increases by one inch, *income* increases by 0.388985 with a p-value of

Solution 14.7-12: The independent variable is income, age, and absingle and the dependent variable is happiness.

Solution 14.7-13: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
H_1 : at least one slope is not equal zero.
where β_1 is the slope for Happiness and Income, β_2 is the slope for Happiness and Age and β_3 is the slope for Happiness and single.

Solution 14.7-14: Read the results from the computer generated output.

Solution 14.7-15: The results are in the table.

Solution 14.7-16: The results are in the table.

Solution 14.7-17: The independent variable is income, age, and absingle and the dependent variable is happiness.

Solution 14.7-18: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
H_1 : at least one slope is not equal zero.
where β_1 is the slope for Happiness and Income, β_2 is the slope for Happiness and Age and β_3 is the slope for Happiness and single.

Solution 14.7-19: Read the results from the computer generated output.

Solution 14.7-20: The results are in the table.

Solution 14.7-21: The results are in the table. Workshops

Solution 15.-1:

1. Construct a relative frequency table for total student majors (column above).
2. The highest percentage of students fall under what major?
3. What percentage of students are Art majors?
4. What percentage of students' majors fall in both Sociology and Psychology?
5. How does the percentage of students who are communications majors in Class A compare to the percentage of communication majors overall for total students?

Solution 15.-2:

1. Using the column 'RF' above, construct a relative frequency table for student majors in the Web Class only.
2. The highest percentage of students fall under what major?
3. What percentage of students are Music majors?
4. What percentage of students are found in the majors of Psychology, Social Work and Sociology (combined)?
5. How does the percentage of students who are education majors in the Web class compare to the percentage of education majors overall for total students? Which is higher?

Solution 15.-3:

1. Using the column 'RF' above, construct a relative frequency table for student majors in the Web Class only.
2. The highest percentage of students fall under what major?
3. What percentage of students are Music majors?
4. What percentage of students are found in the majors of Psychology, Social Work and Sociology (combined)?
5. How does the percentage of students who are education majors in the Web class compare to the percentage of education majors overall for total students? Which is higher?

Solution 15.-4:

1. Fill in the frequency and relative frequency table above
2. Create a bar chart and pie chart for the above data
3. What percentage of students got a grade of 'A'?
4. Looking at the bar chart in part a, identify the shape of the data (symmetric, right-skewed, left-skewed)?

Solution 15.-5:

1. Fill in the frequency and relative frequency above.
2. Complete a stem and leaf plot with the stem representing the 10's place and using 3-9.
3. Draw a histogram for the test scores using the intervals in the relative frequency table above.
4. What percentage of students had scores of 59 or less?

Solution 15.-6:

1. Fill in the frequency and relative frequency above.
2. Complete a stem and leaf plot with the stem representing the 10's place and using 0-2.
3. Draw a histogram for the absences using the intervals in the relative frequency table above.
4. What percentage of students had number of absences less than 20?

Solution 15.-7:

1. Fill in the frequency and relative frequency above.
2. Complete a stem and leaf plot with the stem representing the 10's place and using 0-2.

3. Draw a histogram for the absences using the intervals in the relative frequency table above.
4. What percentage of students had number of absences less than 20?

Solution 15.-8:

1. Fill in the frequency and relative frequency above.
2. Complete a stem and leaf plot with the stem representing the 10's place and using 0-2.
3. Draw a histogram for the absences using the intervals in the relative frequency table above.
4. What percentage of students had number of absences less than 20?

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1: Workshops

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1: