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

Loren L. Heun

June 11, 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: 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 2.7-4: 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

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. $\bar{x} = 326.2$
2. $\tilde{x} = 242.5$
3. $sd = 240.7039676$
4. The distribution is Right skewed
5. Decide which measure of center provides the most relevant information about the distribution? Why?

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 5.5-1: $H_0 : \mu_1 = \mu_2$ vs. $H_0 : \mu_1 \neq \mu_2$

Solution 5.5-3: The test statistics is $|-1.66|$ and significance level is 1.8125. Since the test statistic is less than the significance level, fail to reject H_0.

Solution 5.5-5: Standard Normal (z) distribution, $Z = \pm 1.96$.

Solution 5.5-7: $H_0 : P_1 = P_2$ vs. $H_A : P_1 > P_2$
The Normal Distribution

Solution 6.6-1:

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 6.6-2:

1. The percentage of watch batteries last more than 6 months is $P[X > 6] = 0.998067$.
2. What is the life span of a watch battery which lasts longer than 60% of all batteries? $P[X > a] = .60$. Now solve for a = 21.7198761.
3. What proportion of watch batteries last shorter than 2 years or longer than 3 1/2 years (42 months)? $P[X < 24] + P[X > 42] = .5 + 0.0227501 = 0.5227501$

Solution 6.6-3: The standard deviation is $SD = \frac{35-4}{4} = 7.75$

Solution 6.6-4:

1. The probability the value is greater than 6 is $P[X > 6] = 1$
2. The probability the value is less than 12 is $P[X < 12] = 3.3976731 \times 10^{-6}$
3. The probability the value is between 6 and 12 is $P[6 < X < 12] = 0.9999966$
4. 33% is above a, i.e., $P[X > a] = .33; a = 10.8798263$
5. 33% is below b, i.e., $P[X < b] = .33; b = 9.1201737$

Solution 6.6-5:

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 6.6-6:

1. The proportion of adult female heights is below 72 is $P[X < 72] = 0.999683$.
2. 25% of adult females are greater than $P[X > a] = .25$ where a = 65.4187754

Solution 6.6-7:

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

The Binomial Distribution

Solution 7.7-1:

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 $P[1 \leq X \leq 5] = 0.1708984$.

Solution 7.7-2:

1. The probability that at least 2 questions are correct is $P[X \geq 2] = 0.6241904$.
2. The probability that at most 2 questions arecorrect is $P[X \leq 2] = 0.6777995$.
3. The probability that there will be between 1 and 3 questions is $P[1 \leq X \leq 3] = 0.7717519$.

Solution 7.7-3:

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

Solution 7.7-4:

1. The probability of rolling a 6 no more than 3 times is $P[X \leq 3] = 0.9302722$.
2. The probability that no less than 3 times is $P[X \geq 3] = 0.2247732$.

Solution 7.7-5:

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

Solution 7.7-6: The expected value is 1.27 and SD is 1.1125385

Solution 7.7-7: The expected value is 8 and SD is 2.7712813

Sampling Distribution of the Proportion

Solution 8.6-1: $P[X \geq 32] = 0.0760321$

Solution 8.6-2:

1. The estimate of the population proportion is 0.2
2. The standard error of this estimate is 0.04
3. The 95% margin of error is 0.0784
4. The 95% confidence interval is 0.2 ± 0.0784

Solution 8.6-3: The standard error of this estimate is 0.04

Solution 8.6-4:

1. The estimate of the population proportion is $\hat{p} = 0.066$
2. The standard error of this estimate is $SE_{\hat{p}} = 0.0111035$
3. The 95% margin of error is $M_{\hat{p}} = 0.0217629$
4. The 95% confidence interval is 0.066 ± 0.0217629

Solution 8.6-5:

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 8.6-6:

1. The estimate of the population proportion is $\hat{p} = 0.3182725$
2. The standard error of this estimate is $SE_{\hat{p}} = 0.0062747$
3. The margin of error the estimate is $M_{\hat{p}} = 0.0122983$
4. The 95% confidence interval is 0.3182725 ± 0.0122983

Solution 8.6-7:

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

Comparing Two Proportions

Solution 9.5-1:

1. The difference in percentage of drug use between smokers and nonsmokers is 35.0.
2. Calculate a standard error for your estimate in (1).
3. Calculate a 95% confidence interval for the difference in 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 your 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 your 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 9.5-2:

1. Estimate the difference in percentage of Americans who supported the federal tax on cigarettes between smokers and non-smokers.
2. Calculate a standard error for your estimate in (1).
3. Calculate a 95% confidence interval for the difference in percentage of Americans who supported the federal tax on cigarettes between smokers and non-smokers.
4. Estimate the risk ratio of Americans who supported the federal tax on cigarettes between smokers and non-smokers.
5. Estimate the odds ratio of Americans who supported the federal tax on cigarettes between smokers and non-smokers.

Solution 9.5-3: The critical value is [1] 1.984217
Sampling Distribution of the Mean

Solution 3.4-1:

Solution 3.4-1:

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 3.4-1: 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:

Solution 3.4-1: Correlation

Solution 13.3-1: 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 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1:

Solution 3.4-1: 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: 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: