

Homework 1 Solutions

Assignment

Chapter 2: 18, 23

Chapter 3: 17, 19, 26, 45

Chapter 4: 8, 16, 40b, 45

Chapter 5: 9ab, 20, 26

Chapter 2

2.18] Flowers. *Who* – 385 species of flowers. *What* – Date of first flowering (in days). *When* – Not specified. *Where* – Southern England. *Why* – The researchers believe that this indicates a warming of the overall climate. *How* – Not specified. *Variables* – Date of first flowering is a quantitative variable. *Concerns* - Hopefully, date of first flowering was measured in days from January 1, or some other convention, to avoid problems with leap years.

2.23] Streams. *Who* – Streams. *What* – Name of stream, substrate of the stream (limestone, shale, or mixed), acidity of the water (measured in pH), temperature (in degrees Celsius), and BCI (unknown units). *When* – Not specified. *Where* – Upstate New York. *Why* – Research is conducted for an Ecology class. *How* – Not specified. *Variables* – There are five variables. Name and substrate of the stream are categorical variables, and acidity, temperature, and BCI are quantitative variables.

Chapter 3

3.17] Oil spills.

The bar chart shows that grounding is the most frequent cause of oil spillage for these 312 spills, and allows the reader to rank the other types as well. If being able to differentiate between these close counts is required, use the bar chart. The pie chart is also acceptable as a display, but it's difficult to tell whether, for example, there is a greater percentage of spills caused by grounding or hull failure. If you want to showcase the causes of oil spills as a fraction of all 312 spills, use the pie chart.

3.19] Global warming.

Perhaps the most obvious error is that the percentages in the pie chart only add up to 92%, when they should, of course, add up to 100%. Furthermore, the three-dimensional perspective view distorts the regions in the graph, violating the area principle. The regions corresponding to No Solid Evidence and Due to Natural Patterns should be roughly the same size, at 20% and 21% of respondents, respectively. However, the angle for the 21% region looks much bigger. Always use simple, two-dimensional graphs.

3.26] Politics.

- a) There are 192 students taking Intro Stats. Of those, 115, or about 59.9%, are male.
- b) There are 192 students taking Intro Stats. Of those, 27, or about 14.1%, consider themselves to be “Conservative”.
- c) There are 115 males taking Intro Stats. Of those, 21, or about 18.3%, consider themselves to be “Conservative”.
- d) There are 192 students taking Intro Stats. Of those, 21, or about 10.9%, are males who consider themselves to be “Conservative”.

3.45] Hospitals.

- a) The marginal totals have been added to the table:

		Discharge delayed		
		Large Hospital	Small Hospital	Total
Procedure	Major surgery	120 of 800	10 of 50	130 of 850
	Minor surgery	10 of 200	20 of 250	30 of 450
	Total	130 of 1000	30 of 300	160 of 1300

- b) Yes. Major surgery patients were delayed 130 of 850 times, or about 15.3% of the time. Minor Surgery patients were delayed 30 of 450 times, or about 6.7% of the time.
- c) Large Hospital had a delay rate of 130 of 1000, or 13%. Small Hospital had a delay rate of 30 of 300, or 10%. The small hospital has the lower overall rate of delayed discharge.
- d) Large Hospital: Major Surgery 15% delayed and Minor Surgery 5% delayed. Small Hospital: Major Surgery 20% delayed and Minor Surgery 8% delayed. Even though small hospital had the lower overall rate of delayed discharge, the large hospital had a lower rate of delayed discharge for each type of surgery.
- e) No. While the overall rate of delayed discharge is lower for the small hospital, the large hospital did better with *both* major surgery and minor surgery.
- f) The small hospital performs a higher percentage of minor surgeries than major surgeries. 250 of 300 surgeries at the small hospital were minor (83%). Only 200 of the large hospital's 1000 surgeries were minor (20%). Minor surgery had a lower delay rate than major surgery (6.7% to 15.3%), so the small hospital's overall rate was artificially inflated. Simply put, it is a mistake to look at the overall percentages. The real truth is found by looking at the rates after the formation is broken down by type of surgery, since the delay rates for each type of surgery are so different. The larger hospital is the better hospital when comparing discharge delay rates.

Chapter 4

4.8] Singers.

a) The distribution of the heights of singers in the chorus is bimodal, with a mode at around 65 inches and another mode around 71 inches. No chorus member has height below 60 inches or above 76 inches.

b) The two modes probably represent the mean heights of the male and female members of the chorus.

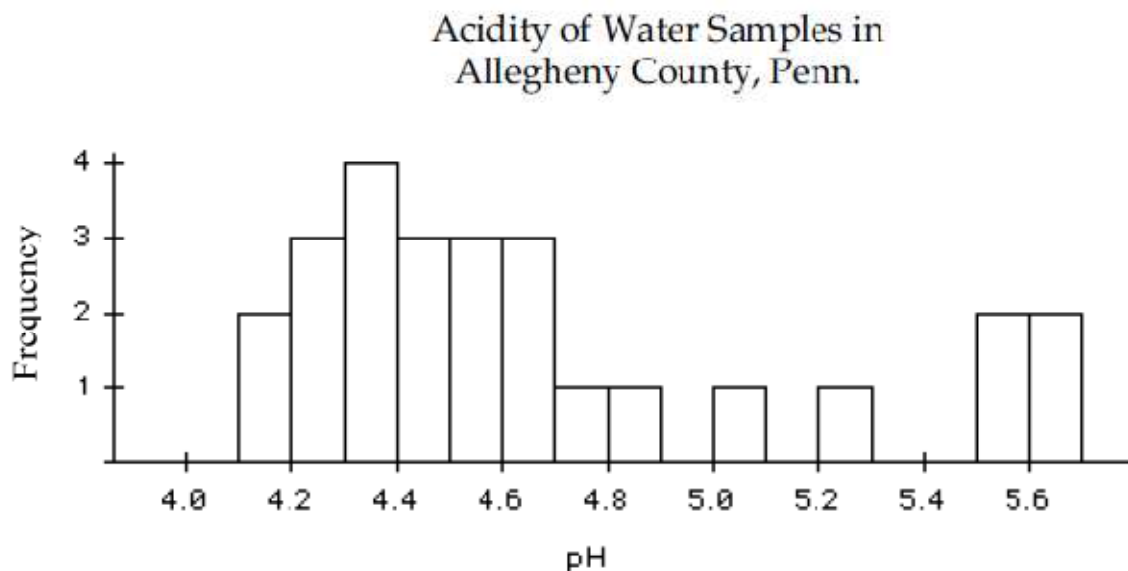
4.16] Paper consumption.

The median and IQR would be used to summarize the distribution of paper consumption, since the distribution is strongly skewed.

4.40b] The distribution of the number of birds spotted by participants in the 1999 Laboratory of Ornithology Christmas Bird Count is skewed right, with a center at around 160 birds. There are several high outliers, with two participants spotting 206 birds and another spotting 228. With the exception of these outliers, most participants saw between 152 and 186 birds.

4.45] Acid rain.

The distribution of the pH readings of water samples in Allegheny County, Penn. is bimodal. A roughly uniform cluster is centered around a pH of 4.4. This cluster ranges from pH of 4.1 to 4.9. Another smaller, tightly packed cluster is centered around a pH of 5.6. Two readings in the middle seem to belong to neither cluster.



Chapter 5

5.9] Still rockin'.

a) The histogram and boxplot of the distribution of “crowd crush” victims’ ages both show that a typical crowd crush victim was approximately 18 - 20 years of age, that the range of ages is 36 years, that there are two outliers, one victim at age 36 - 38 and another victim at age 46 – 48.

b) This histogram shows that there may have been two modes in the distribution of ages of “crowd crush” victims, one at 18 - 20 years of age and another at 22 – 24 years of age. Boxplots, in general, can show symmetry and skewness, but not features of shape like bimodality or uniformity.

5.20] Gas prices.

a) Gas prices have been increasing on average over the three year period, and the spread has been increasing as well. The distribution of prices in 2002 was skewed to the left with several low outliers. Since then, the distribution has been increasingly skewed to the right. There is a high outlier in 2004, although it appears to be pretty close to the upper fence.

b) The distribution of gas prices in 2004 shows the greatest range and the biggest IQR, so the prices varied a great deal.

5.26] Ozone.

a) April had the highest recorded ozone level, approximately 440.

b) February had the largest IQR of ozone level, approximately 50.

c) August had the smallest range of ozone levels, approximately 50.

d) January had a slightly lower median ozone level than June, 340 and 350, respectively, but June’s ozone levels were much more consistent.

e) Generally, ozone levels rose through the winter and were highest in the spring, then fell through the summer and were lowest in the fall. Additionally, ozone levels were very consistent in the summer, became more variable in the fall, were most variable in the winter, and became more consistent through the spring.