

## Homework 1 Solutions

Chapter 1: # 10; Chapter 2: # 20

Chapter 2: # 40a, 54

Chapter 2: #81, 85

### Assigned Friday, September 25

- 1.10** a. This statement was almost surely based on an observational study. Most likely, the researchers observed who was breast-fed and who was not, and did not randomly assign mothers to breast-feed or not. It would not be ethical to randomly assign this treatment.
- b. The better headline is “Link Found Between Breast-feeding and School Performance” because it does not imply that there is a cause-and-effect conclusion, while the other headline does. Because the study was observational, a cause-and-effect conclusion cannot be made. It is likely that children who were breast-fed as infants have other (confounding) factors in their lives that differ from children who were not, and that may influence school performance. For example, perhaps they are more likely to have mothers who don’t work, perhaps they are more likely to be first-born children, and so on. We can say that a link was found, but cannot say that breast-feeding *leads* to better performance.
- 2.20** a. A unit is a person. Dominant hand is a categorical variable and IQ is a quantitative variable. Explanatory variable is dominant hand and response variable is IQ.
- b. A unit is a married couple. Eventual divorce status and pet ownership are both categorical variables. Explanatory variable is pet ownership and response variable is eventual divorce status.
- c. A unit is a college student. GPA and hours of study each week are both quantitative variables. Explanatory variable is hours of study and response variable is GPA.
- d. A unit is a tax-paying individual in the United States. Tax bracket is an ordinal variable and percentage donated to charities is a quantitative variable. Explanatory variable is tax bracket and response variable is percentage donated to charities.

### Assigned Monday, September 28

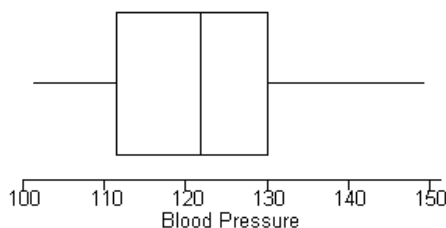
- 2.40** a. You should use 2 stem values for each 10s digit, one for leaves of 0-4 and one for leaves of 5-9. Using only 1 stem value would give just 5 values, and using 5 would give 20 stem values. The ideal is to have between 6 and 15 stem values.

Figure for Exercise 2.40a

```
|5| 6
|6| 2
|6| 8
|7|
|7| 55
|8| 02344
|8| 77
|9| 13
|9| 5
```

- 2.54**
- a. Remember to put the values into numerical order before finding the middle. The ordered values are 102, 110, 118, 122, 123, 127, 132, 150 and  $n = 8$ , so the median is the average of the 4<sup>th</sup> and 5<sup>th</sup> values:  $(122+123)/2 = 122.5$ .
  - b. The quartiles are the medians of the lower 4 numbers and the upper 4 numbers, so  $Q_1 = 114$  and  $Q_3 = 129.5$ .
  - c.  $IQR = Q_3 - Q_1 = 129.5 - 114 = 15.5$ .
  - d.  $1.5 \times IQR = 23.25$ . A number will be considered an outlier if it is either below  $114 - 23.25 = 90.75$  or above  $129.5 + 23.25 = 152.75$ . No values fit this criterion, so there are no outliers.
  - e.

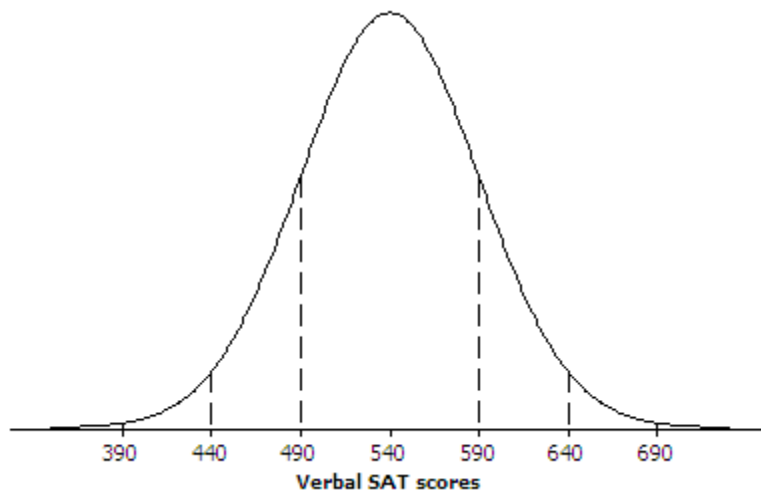
Figure for Exercise 2.54e



### Assigned Wednesday, September 30

- 2.81** The interval is  $-6.1$  to  $22.7$  hours. The interval includes negative values, which are impossible times. Thus, the interval based on an assumption of a bell-shaped curve would not reflect reality. A substantial proportion of a normal curve should lie between one and two standard deviations below the mean, with more values between two and three standard deviations below the mean. That is clearly impossible here.
- 2.85**
- a. About 68% fall in the interval  $540 \pm 50$ , which is 490 to 590.  
About 95% fall in the interval  $540 \pm (2)(50)$ , which is 440 to 640.  
About 99.7% fall in the interval  $540 \pm (3)(50)$ , which is 390 to 690.

Figure for Exercise 2.85a



- b.  $s^2 = 50^2 = 2500$  (Variance = squared standard deviation.)