SOLUTIONS MANUAL

THE AUTHOR

1. Why you should use data science?

—

Solution 1.6-1
All of the choices is the correct answer.

—

Solution 1.6-2
The level of measurement of marital statis is nominal (3.).

—

Solution 1.6-3
The level of measurement of marital status is interval-ratio (2.).

—

Solution 1.6-4
The number of reported robberies in June 2014 is Kalamazoo County is continuous variable (3.).

2. Dataset Descriptive Information

—

Solution 2.5-1

$$(2.)p = \frac{250 - 195}{250} = 0.22 \tag{1}$$

—

Solution 2.5-2

$$(3.)ratio = \frac{195}{250 - 195} = 3.55 \tag{2}$$

—

Solution 2.5-3

$$(2.)rate = \frac{13}{25000} \times 100000 = 52 \tag{3}$$

—

Solution 2.5-4

$$(1.)PC = \frac{(83-89)}{89} \times 100 = -6.7\% \tag{4}$$

—

Solution 2.5-5
(1.) The percentage of nurses who are female is 36.8%. **(2.)** The proportion of orderlies who are males is 0.367. **(3.)** Ratio is 18 females docs to 83 males docs or approximately 1 female doc for every 5 males docs. **(4.)** Percentage of females on the staff is 43.3%.

—

Solution 2.5-6
The measure of the lost hours due to traffic is interval-ratio.

3. Measures of Central Tendency

—

Solution 3.9-1
(3.) 32.167, 32.5, 35

—

Solution 3.9-2
(2.) The three value must be 22.

—

Solution 3.9-3

Mean 472.6
Median 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 3.9-4
(1.) Mode **(2.)** Median **(3.)** Mean **(4.)** Mean **(5.)** Median **(6.)** Mode

—

Solution 3.9-5
Birth Mode="North"; **Legal** Median=2.5; **Expense** Mean=48.5; **Movies** Mean=5.8; **Food** Median=6; **Religion** Mode="Protestant"

4. Measures of Dispersion

—

Solution 4.9-1

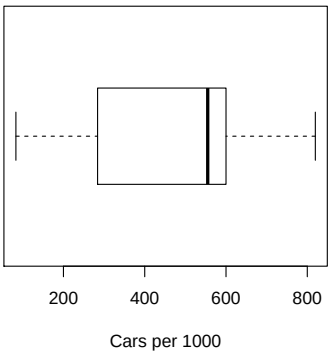

FIGURE 1. Boxplot of Cars per 1000

$$range = (X_{max} - X_{min}) = 42 - 16 = 26$$

—

Solution 4.9-2

$$LB = Q_1 - 1.5(IQR) = 10 - 1.5(20 - 10) = -5$$
$$UB = Q_3 + 1.5(IQR) = 20 + 1.5(20 - 10) = 35$$

There are no outliers.

—

Solution 4.9-3
SD = 1.581

5. Normal Curve

—

Solution 5.8-1

$$SD = \frac{(41 - 32)}{3} = 3 \tag{5}$$

—

Solution 5.8-2

$$z = \frac{173 - 153}{11} = 1.82 \tag{6}$$

—

Solution 5.8-3

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

—

Solution 5.8-4

$$x = \mu + Z\sigma$$
$$x = 155 + 1.28 \times (12)$$
$$x = 170.36$$

6. Inferential Statistics

—

Solution 6.9-1
The population for this sample survey is all adult resident of the United States.

—

Solution 6.9-2
This is an example of stratified random sampling.

—

Solution 6.9-3
This method is called systemic random sampling.

—

Solution 6.9-4
What proportion will exceed 72.0 inches?

$$SE = \frac{SD}{\sqrt{n}} = \frac{2.9}{\sqrt{10}} = 0.9171$$
$$Z = \frac{(X - \mu)}{SE}$$
$$Z = \frac{(72.0 - 69.1)}{0.9171} = 3.1623$$
$$P[Z > 3.1623] = 0.0008$$

—

Solution 6.9-5

What is the chance that this group will average over 220?

$$SE = \frac{SD}{\sqrt{n}} = \frac{28}{\sqrt{40}} = 4.4272$$
$$Z = \frac{(X - \mu)}{SE}$$
$$Z = \frac{(220 - 210)}{4.4272} = 2.2588$$
$$P[Z > 2.2588] = 0.0119$$