

Statistics 101 – Homework 2

Solution

Assignment:

1. Do the following problems from the text, *Intro Stats*, 3rd Edition. If you have an earlier edition of the text, check with someone who has the 3rd Edition to make sure you do the correct problems.

- Chapter 4 – problems 5, 8, 10, 13, 22, 30, 43 and 45

If you have questions about the solutions to these problems, contact your instructor.

2. The birth weights (in grams) of 18 baby girls born at a Brisbane, Australia hospital are given below.

3,837 3,334 2,208 2,576 3,208 3,746 3,523 3,430 3,480
3,116 3,428 2,184 2,383 3,500 3,866 3,542 3,278 1,745

- a) Answer the questions Who? and What? for these data.

Who? Newborn baby girls.

What? Birth weight (grams).

- b) Make a stem-and-leaf display of the baby girls' birth weights.

```
1 |  
1*|745  
2 |184, 208, 383  
2*|576  
3 |116, 208, 278, 334, 428, 430, 480  
3*|500, 523, 542, 746, 837, 866
```

- c) Using the stem-and-leaf display, describe the distribution of birth weights for baby girls. Make sure you mention the shape, center, spread and any outliers or other interesting characteristics of the distribution.

There appear to be two mounds. A larger one in the 3,000 to 4,000 gram range and a smaller one in the low 2,000 gram range. One could also say there is a skew to the left (toward lower values). The center is in the lower 3,000 grams range and weights are spread between 1,700 g and 3,900 g. There are no apparent outliers.

- d) Calculate the sample mean birth weight.

$$\text{sample mean} = \bar{y} = \frac{56384}{18} = 3,132.4 \text{ grams}$$

- e) Calculate the sample median birth weight. Why is the sample median birth weight greater than the sample mean birth weight?

$$\text{sample median} = \frac{(3,334 + 3,428)}{2} = 3,381.0 \text{ grams}$$

The sample median is greater than the sample mean because the data are skewed to the left (toward lower values) and the sample mean is pulled in the direction of the skew.

- f) Calculate the five number summary for these data.

minimum: 1,745 g
Q_L: 2,576 g
median: 3,381 g
Q_U: 3,523 g
maximum: 3,866 g

- g) Given the value $\sum (y - \bar{y})^2 = 6,781,240.5$, calculate the sample standard deviation for these data.

$$s = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}} = \sqrt{\frac{6,781,240.5}{17}} = \sqrt{398,896.5} = 631.58 \text{ grams}$$

- h) Which do you think provides a more informative summary of these data; a five number summary or the sample mean and sample standard deviation? Briefly explain your answer.

The five number summary gives a more informative summary because the data are skewed to the left and this is shown in the five number summary. The mean and standard deviation are better for symmetric distributions.

There were also 26 baby boys born at this Brisbane, Australia hospital. On the next page is JMP output for the distribution of baby boy birth weights.

- i) Describe the distribution of birth weights for baby boys. Make sure you mention the shape, center, spread and any outliers or other interesting characteristics of the distribution.

The shape is fairly symmetric and mounded in the middle, although some might detect a slight skew to the left. The center is around 3,400 grams and the weights are spread from 2,100 g to 4,200 g.

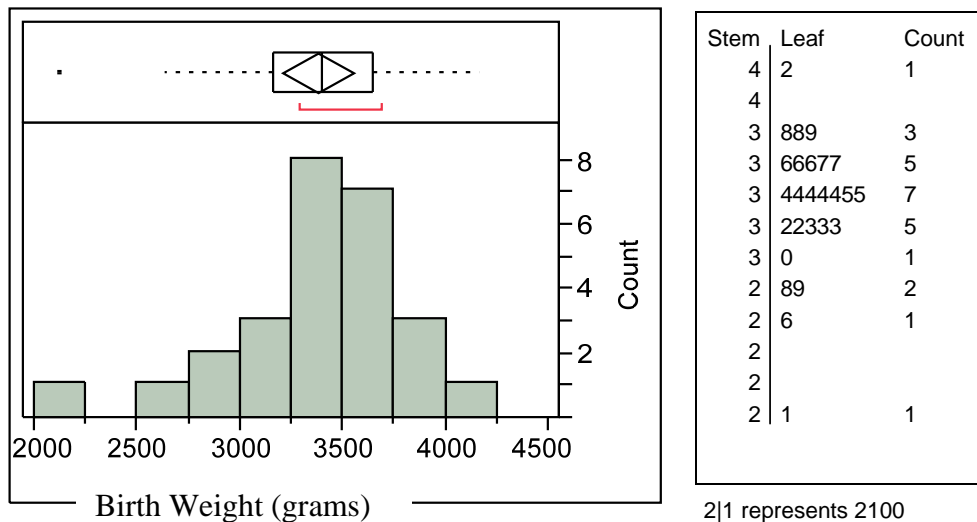
- j) Give the value of the sample mean birth weight for boys. Give the value of the sample median birth weight for boys. Explain why the sample mean is less than the sample median.

sample mean weight = 3,375.3 grams
sample median weight = 3,404.0 grams

- k) Give the value of the sample standard deviation for boys.

sample standard deviation = 428.05 grams

Male Birth Weight



Five Number Summary		Summary Statistics	
100.0%	maximum	4162.0	Mean 3375.3077
75.0%	quartile	3645.0	Std Dev 428.04605
50.0%	median	3404.0	N 26
25.0%	quartile	3162.0	
0.0%	minimum	2121.0	

3. **(JMP assignment)** A sample of 57 cars is selected from the 2004 model year. Data on company, model, horsepower, city mpg, highway mpg and weight are recorded. The data are available on the course web page both as a JMP (.jmp). Follow the directions in the JMP Guide to download the file and create output. Make sure to print the output and turn it in with your assignment. Use the output in answering the following questions.
- Answer the question Who?

Cars from the 2004 model year.

- b. Answer the question What? For each What? indicate whether the variable is categorical/qualitative or numerical/quantitative.

Company – categorical
Model – categorical
Horsepower – numerical
City mpg – numerical
Highway mpg – numerical
Weight – numerical

- c. Looking at the histogram describe the shape of the distribution of horsepower. What might this shape indicate about horsepower?

The shape is bi-modal. There are two mounds one around 150 horsepower and the other between 200 and 225 horsepower. These two mounds may indicate two different types of cars in the sample; big high horsepower cars and smaller lower horsepower cars.

- d. Give the five number summary for horsepower.

Minimum: 103 horsepower
Q_L: 150 horsepower
median: 182 horsepower
Q_U: 210 horsepower
maximum: 275 horsepower

- e. Calculate the sample range and sample IQR for these data.

Sample range = $275 - 103 = 172$ horsepower
Sample IQR = $210 - 150 = 60$ horsepower

- f. Looking at the box plot describe the shape of the distribution of horsepower.

The box plot indicates a fairly symmetric shape. The median (and mean) are in the middle of the box. If you folded the box plot on the median line, the right side would match up pretty well with the left side.

- g. Looking at the box plot, are there any apparent outliers? If so, what are their values?

According to the box plot there are no apparent outliers because there are no values noted separately in the plot.

h. Give the sample mean and sample standard deviation for these data.

sample mean = 184.2 horsepower
sample standard deviation = 41.27 horsepower

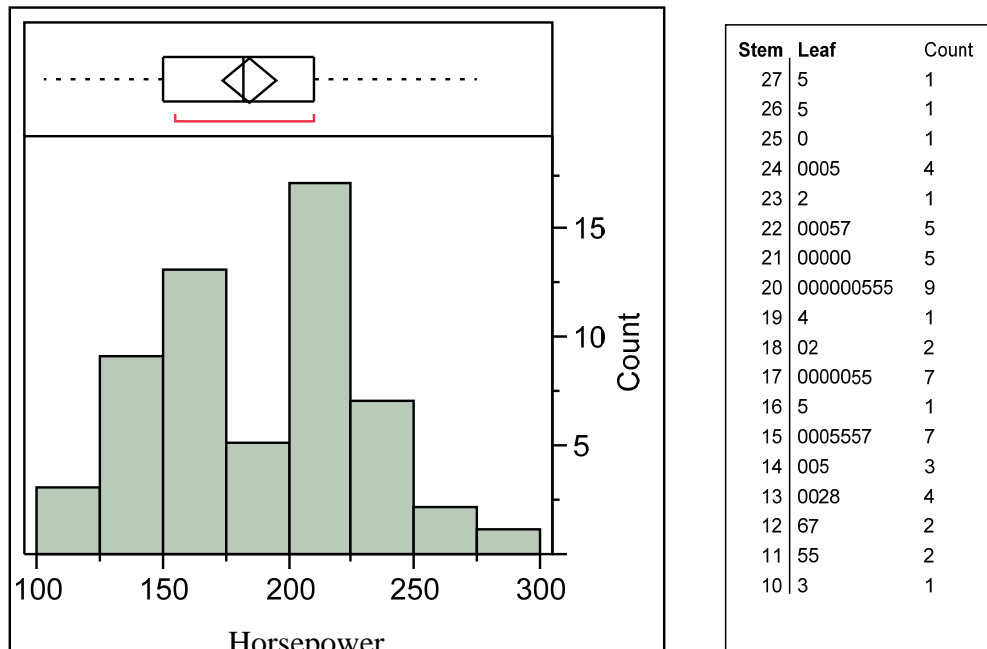
i. Did JMP split the stems in the stem and leaf plot?

No, there is only one line of numbers for each stem.

j. Which do you think is the most informative display of the data, the box plot, the histogram or the stem and leaf plot? Explain briefly.

The stem and leaf plot is the most informative. You can see the multiple mounds in the horsepower values. You also retain the actual value of the horsepower for each individual car. The histogram groups the horsepower values into groups (bins) and the box plot only shows the five number summary values. The stem and leaf plot is also more informative because give this plot you can construct the other plots. Given either the histogram or the box plot you cannot construct any of the other plots accurately.

JMP Output



10|3 represents 103

Five Number Summary		Sample Statistics		
100.0%	maximum	275	Mean	184.17544
75.0%	quartile	210	Std Dev	41.267906
50.0%	median	182	N	57
25.0%	quartile	150		
0.0%	minimum	103		