Question 1

a

$$\#\text{bin} \approx \sqrt{129} \approx 12$$
$$\Rightarrow \#\text{bin} = 12$$
$$\text{bin\_width} = \frac{\text{max} - \text{min}}{\#\text{bin}} = \frac{18.9 - 2.2}{12} \approx 1.4$$
$$\text{bin}_1 = \{a, a \in [2.2, 3.6)\}$$
$$V_{\text{bin}_1} = 7$$
$$f_{\text{bin}_1} = 7 \div 1.4 = 5$$
$$\text{bin}_2 = \{a, a \in [3.6, 5.0)\}$$
$$V_{\text{bin}_2} = 13$$
$$f_{\text{bin}_2} = 13 \div 1.4 \approx 9.29$$
$$\text{bin}_3 = \{a, a \in [5.0, 6.4)\}$$
$$V_{\text{bin}_3} = 28$$
$$f_{\text{bin}_3} = 28 \div 1.4 = 20$$
$$\text{bin}_4 = \{a, a \in [6.4, 7.8)\}$$
$$V_{\text{bin}_4} = 30$$
$$f_{\text{bin}_4} = 30 \div 1.4 \approx 21.43$$
$$\text{bin}_5 = \{a, a \in [7.8, 9.2)\}$$
$$V_{\text{bin}_5} = 11$$
$$f_{\text{bin}_5} = 11 \div 1.4 \approx 7.86$$
$$\text{bin}_6 = \{a, a \in [9.2, 10.6)\}$$
$$V_{\text{bin}_6} = 20$$
$$f_{\text{bin}_6} = 20 \div 1.4 \approx 14.29$$
$$\text{bin}_7 = \{a, a \in [10.6, 12.0)\}$$
$$V_{\text{bin}_7} = 10$$
$$f_{\text{bin}_7} = 10 \div 1.4 \approx 7.14$$
$$\text{bin}_8 = \{a, a \in [12.0, 13.4)\}$$
$$V_{\text{bin}_8} = 2$$
$$f_{\text{bin}_8} = 2 \div 1.4 = \approx 1.43$$
$$\text{bin}_9 = \{a, a \in [13.4, 14.8)\}$$
$$V_{\text{bin}_9} = 3$$
$$f_{\text{bin}_9} = 3 \div 1.4 \approx 2.14$$

$\text{bin}_{10} = \{a, a \in [14.8, 16.2)\}$

$V_{\text{bin}_{10}} = 4$

$f_{\text{bin}_{10}} = 4 \div 1.4 \approx 2.86$
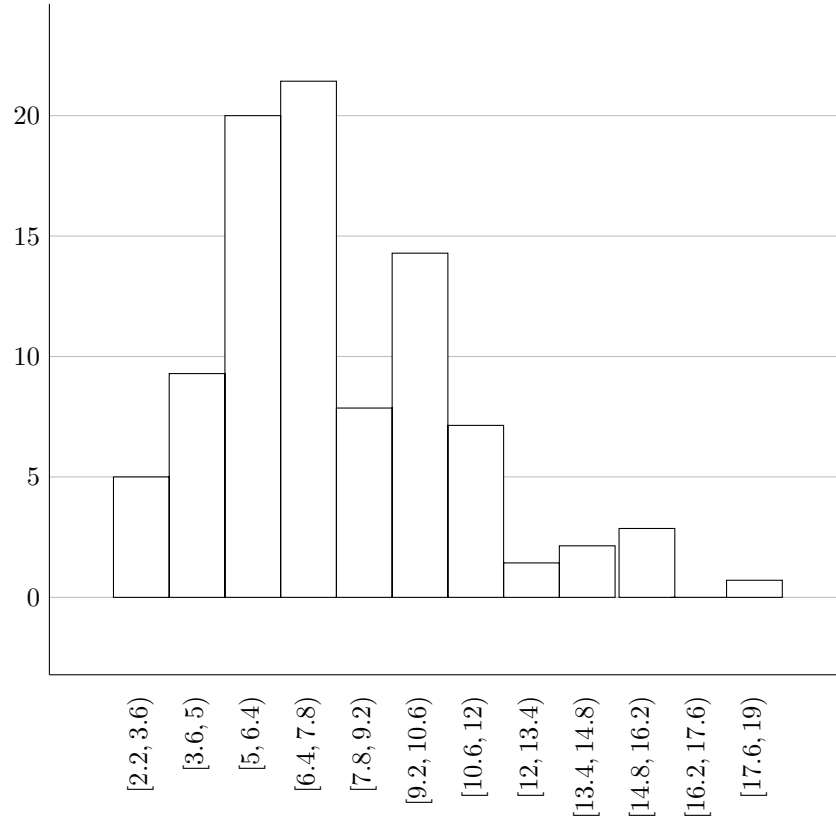
$\text{bin}_{11} = \{a, a \in [16.2, 17.6)\}$

$V_{\text{bin}_{11}} = 0$

$f_{\text{bin}_{11}} = 0 \div 1.4 = 0$

$\text{bin}_{12} = \{a, a \in [17.6, 19.0)\}$

$V_{\text{bin}_{12}} = 1$

$f_{\text{bin}_{12}} = 1 \div 1.4 \approx 0.71$

b

The typical value should be median since the date values are not symmetrical and have outliers.Med $= 7$. So the typical value is 7.

c

The display is spread out because there is a wide range of values and there are data points.

d

$$\text{Med} = 7$$

$$\text{Mean} = \frac{\sum(x_i)}{129} \approx 7.7$$

$$\Rightarrow \text{Med} < \text{Mean}$$

$$\Rightarrow \text{Right-skewed}$$

e

$Q_1 = (5.6 + 5.6) \div 2 = 5.6$
$Q_3 = (8.8 + 9.0) \div 2 = 8.9$
$\text{IQR} = Q_3 - Q_1 = 3.3$
$Q_1 - 1.5\text{IQR} = 5.6 - 4.95 = 0.65$
$Q_3 + 1.5\text{IQR} = 8.9 + 4.95 = 13.85$
$\text{Outliers} = \{x, x \in (-\infty, 0.65) \cup (13.85, \infty)\}$
$\Rightarrow \text{Outliers} = \{14.3, 14.6, 15, 15, 15.3, 15.5, 18.9\}$
There is no outliers smaller than the median, but are outliers larger than the median

Question 2

There exists negative value for the x-axis, meaning that there are people speeding up during 35km-40km.The time differences range from -50 to 750 sec. The most frequent classes are 50 to 150,meaning that most runners are 50 to 150 seconds slower. The diagram is skew to the right, so the mean is greater than the median.

Question 3

a

$$\bar{x} = \frac{\sum x_i}{n} = \frac{116.4 + 115.9 + 114.6 + 115.2 + 115.8}{5} = 115.58$$
$$x_1 - \bar{x} = 116.4 - 115.58 = 0.82$$
$$x_2 - \bar{x} = 115.9 - 115.58 = 0.32$$
$$x_3 - \bar{x} = 114.6 - 115.58 = -0.98$$
$$x_4 - \bar{x} = 115.2 - 115.58 = -0.38$$
$$x_5 - \bar{x} = 115.8 - 115.58 = 0.22$$

b

$$s^2 \frac{0.82^2 + 0.32^2 + 0.98^2 + 0.38^2 + 0.22^2}{4} = 0.482$$
$$s = \sqrt{s^2} \approx 0.694$$

c

$$S_{xx} = \sum x_i{}^2 - \frac{\left(\sum x_i\right)^2}{n}$$
$$= 66795.61 - \frac{577.9^2}{5}$$
$$= 1.928$$
$$s^2 = \frac{S_{xx}}{4} = 0.482$$

d

$$x_1' = 116.4 - 100 = 16.4$$

$$x_2' = 115.9 - 100 = 15.9$$

$$x_3' = 114.6 - 100 = 14.6$$

$$x_4' = 115.2 - 100 = 15.2$$

$$x_5' = 115.8 - 100 = 15.8$$

$$\overline{x'} = 15.58$$

$$x_1' - \overline{x'} = 16.4 - 15.58 = 0.82$$

$$x_2' - \overline{x'} = 15.9 - 15.58 = 0.32$$

$$x_3' - \overline{x'} = 14.6 - 15.58 = -0.98$$

$$x_4' - \overline{x'} = 15.2 - 15.58 = -0.38$$

$$x_5 - \overline{x'} = 15.8 - 15.58 = 0.22$$

$$s'^2 = \frac{0.82^2 + 0.32^2 + 0.98^2 + 0.38^2 + 0.22^2}{4} = 0.482$$

$\Rightarrow$Variance stays the same when the whole data set is added or subtracted a certain value

Question 4

The two sets of data have similar medians. Machine 2 has a wider range and a larger IQR than machine 1. Machine 2 ranges from 85 to 115 roughly,with IQR about 20, but machine 1 has much smaller range and IQR. There is an outlier in machine 1 but no outlier in machine 2. Also, it is easy to see that the data of machine 1 is closer to the

In conclusion, I think machine 1 has a stabler performance than machine 2.

Question 5

$$\text{min} = 55.3$$
$$\text{max} = 83$$
$$\text{Median } = 64.7$$
$$\text{Mean } = 64.89$$
$$Q_1 = 57.8$$
$$Q_3 = 70.4$$
$$IQR = Q_3 - Q_1 = 12.6$$
$$\text{Range } = \text{max} - \text{min} = 27.7$$
$$\text{Median } + 1.5 IQR = 83.6 > \text{max}$$
$$\text{Median } - 1.5 IQR = 45.8 < \text{min}$$

According to the date, mean is greater than median, which means that the histogram should be right-skewed. The maximum is smaller than Median+1.5 IQR, and the minimum is bigger than Median-1.5 IQR, this means that there is no outlier in this set of data.

Question 6

a

$$\text{min} = 15$$
$$\text{max} = 829$$
$$\text{Median} = \frac{195 + 196}{2} = 195.5$$
$$\text{Mean} = \frac{\sum x_i}{n} = \frac{22198}{100} = 221.98$$
$$Q_1 = \frac{124 + 131}{2} = 127.5$$
$$Q_3 = \frac{282 + 284}{2} = 283$$
$$\text{Range} = 829 - 15 = 814$$
$$IQR = Q_3 - Q_1 = 283 - 127.5 = 155.5$$

b

$$s^2 = \frac{\sum (x - \overline{x})^2}{n - 1} = \frac{2070524}{99} \approx 20914.384$$
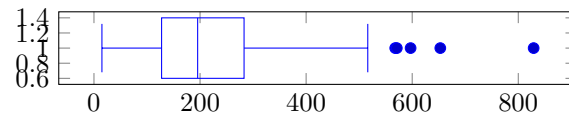$$s = \sqrt{s^2} \approx 144.618$$

c

$$Q_1 - 1.5 IQR = 127.5 - 1.5 \times 155.5 = -105.75$$
$$Q_3 + 1.5 IQR = 283 + 1.5 \times 155.5 = 516.25$$
$$\Rightarrow x \text{ is an outlier } \Leftrightarrow x \in (-\infty, -105.75) \cup (516.25, \infty)$$

d

e

The distribution is right skewed, can be seen from the fact that the mean is bigger than the median. There are five outliers in this set of data. Those big numbers are the cause of the right-shift of the mean. And because of those outliers, median is the better way of representing the center instead of mean. So the data ranges at [15,497] and center at 195.5, excluding the outliers. It's IQR is 155.5, and its standard deviation is 144.618.

Collaborators

Frank Zhu

Jeffery Shu

Sam Sun