EX 1.4.1: Given the following sample of fuel efficiencies of 6-cylinder vehicles (in miles/gallon):

$$x: \ 21.0,\ 15.0,\ 21.0,\ 21.4,\ 18.1,\ 19.2,\ 17.8,\ 19.7,\ 13.0,\ 35.0$$

a) Compute the sample range, x_R.

First, sort sample $\quad x: \ 13.0,\ 15.0,\ 17.8,\ 18.1,\ 19.2,\ 19.7,\ 21.0,\ 21.0,\ 21.4,\ 35.0$

Then, $x_R = x_{(n)} - x_{(1)} =$ (Largest Data Point) − (Smallest Data Point) $= 35.0 - 13.0 =$ $\boxed{22 \text{ miles/gallon}}$

b) Compute the sample variance, s^2.

Use easier formula since $\overline{x}$ is not known:

$$S_{xx} = \sum_{k=1}^{n} x_k^2 - \frac{1}{n}\left(\sum_{k=1}^{n} x_k\right)^2 = 4360.14 - \left(\frac{1}{10}\right)(201.2)^2 = 311.996 \implies s^2 = \frac{S_{xx}}{n-1} = \frac{311.996}{9} \approx \boxed{34.6662 \text{ (miles/gallon)}^2}$$

How to compute sum $\sum_{k=1}^{n} x_k^2$ via calculator: `13.0^2+15.0^2+17.8^2+18.1^2+19.2^2+19.7^2+21.0^2+21.0^2+21.4^2+35.0^2` `=`

c) Compute the sample standard deviation, s.

$s = \sqrt{s^2} = \sqrt{34.6662} \approx \boxed{5.8878 \text{ miles/gallon}}$

d) Compute the interquartile range, x_{IQR}.

$x_{Q1} = x_{0.25} = x_{(\lceil 0.25n \rceil)} = x_{(\lceil 0.25(10) \rceil)} = x_{(\lceil 2.5 \rceil)} = x_{(3)} = (3^{rd}$ data point in sorted sample$) = 17.8$

$x_{Q3} = x_{0.75} = x_{(\lceil 0.75n \rceil)} = x_{(\lceil 0.75(10) \rceil)} = x_{(\lceil 7.5 \rceil)} = x_{(8)} = (8^{th}$ data point in sorted sample$) = 21.0$

$\therefore \quad x_{IQR} = x_{Q3} - x_{Q1} = 21.0 - 17.8 =$ $\boxed{3.2 \text{ miles/gallon}}$

e) Compute the interhinge range, x_{IHR}.

(Left Half of x) : $\quad 13.0,\ 15.0,\ 17.8,\ 18.1,\ 19.2 \quad \implies x_{LH} =$ (Median of left half of x) $= 17.8$

(Right Half of x) : $\quad 19.7,\ 21.0,\ 21.0,\ 21.4,\ 35.0 \quad \implies x_{UH} =$ (Median of right half of x) $= 21.0$

$\therefore \quad x_{IHR} = x_{UH} - x_{LH} = 21.0 - 17.8 =$ $\boxed{3.2 \text{ miles/gallon}}$

f) Identify, if any, mild & extreme outliers in the sample.

Lower Extreme Outlier Boundary: $x_{LH} - 3.0x_{IHR} = 17.8 - (3.0)(3.2) = 8.2$

Lower Mild Outlier Boundary: $x_{LH} - 1.5x_{IHR} = 17.8 - (1.5)(3.2) = 13.0$

Upper Mild Outlier Boundary: $x_{UH} + 1.5x_{IHR} = 21.0 + (1.5)(3.2) = 25.8$

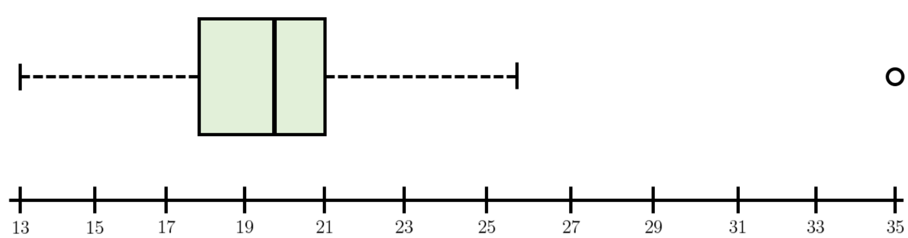
Upper Extreme Outlier Boundary: $x_{UH} + 3.0x_{IHR} = 21.0 + (3.0)(3.2) = 30.6$

$\therefore$ Mild Outliers lie in interval $[8.2, 13.0) \cup (25.8, 30.6]$ and Extreme Outliers lie in interval $(-\infty, 8.2) \cup (30.6, \infty)$

$\therefore$ Visual inspection of the sample implies that **35.0** is the only outlier and it's an **extreme outlier.**

g) Construct the horizontal boxplot for the sample. (Use hinges, not quartiles!)

First, compute median as it is needed for boxplot: $\widetilde{x} = \frac{19.2+19.7}{2} = 19.45$



h) Use the boxplot to describe the skewness of the sample.

The median is slightly closer to the upper hinge, therefore the middle 50% of the sample is slightly negatively skewed.

The two dashed lines between hinges and outlier boundaries are equal in length, so the data is overall nearly symmetric.