# Class Exercise

1. **Marathon Finish Times**
   A local marathon recorded the finish times (in minutes) for 40 amateur runners. The data is sorted in ascending order below:

   | | | | | | | | | | |
   |---|---|---|---|---|---|---|---|---|---|
   | 195 | 202 | 210 | 215 | 215 | 218 | 220 | 225 | 230 | 232 |
   | 235 | 240 | 245 | 248 | 250 | 252 | 255 | 258 | 260 | 265 |
   | 270 | 272 | 275 | 278 | 280 | 285 | 290 | 295 | 300 | 305 |
   | 310 | 315 | 320 | 330 | 345 | 350 | 360 | 375 | 390 | 410 |

   a. **Calculate the IQR:**
   $Q_1$ is at location $L = 41 \times 0.25 = 10.25$. Thus $Q_1 = 232 + 0.25 \times (235 - 232) = 232.75$.
   $Q_3$ is at location $L = 41 \times 0.75 = 30.75$. Thus $Q_3 = 305 + 0.75 \times (310 - 305) = 308.75$
   $IQR = Q_3 - Q_1 = 308.75 - 232.75 = 76 min$
   **Interpretation:** The middle 50% of runners finished within a range of 76 minutes of each other.

   b. $P_{90}$**:** Position $L = 41 \times 0.90 = 36.9$. Thus $P_{90} = 350 + 0.9 \times (360 - 350) = 359$

   c. $P_{15}$**:** Position $L = 41 \times 0.15 = 6.15$. Thus $P_{15} = 218 + 0.15 \times (220 - 218) = 218.3$

2. **Coffee Consumption**
   A survey was conducted at a local office asking employees how many cups of coffee they consumed over a 5-day work week. The results are presented in the frequency table below ($n = 80$):

   | Cups | Freq | Cum. Freq |
   |---|---|---|
   | 0 | 5 | 5 |
   | 2 | 14 | 19 |
   | 4 | 22 | 41 |
   | 6 | 18 | 59 |
   | 8 | 12 | 71 |
   | 10 | 7 | 78 |
   | 12 | 2 | 80 |

   a. **Bottom** 10% ($P_{10}$)**:** Location $0.10 \times 81 = 8.1$. Value is 2 cups. **Top** 10% ($P_{90}$)**:** Location $0.90 \times 81 = 72.9$. Value is 10 cups.

   b. **Quartiles:** $Q_1$ (Location $0.25 \times 81 = 20.25$): 4 cups. $Q_2$ (Location $0.5 \times 81 = 40.5$): 4 cups. $Q_3$ (Location $0.75 \times 81 = 60.75$): 8 cups.

   c. **IQR:** $Q_3 - Q_1 = 8 - 4 = 4$ cups. The middle 50% of the coffee drinkers drink between 4 and 8 cups a work week.

3. **Percentile Definition**

   You have a dataset of $n = 24$ data points. Which integer percentiles are not defined for your dataset?

   Using the formula $L = \frac{(n+1)P}{100}$, we have that the first data point (location $L = 1$) is at percentile

   $$P = \frac{100 \cdot 1}{n+1} = \frac{100}{25} = 4$$

   Thus percentiles $P_1, P_2, P_3$ are not defined in this dataset. Similarly we have that the last (24th) data point is at percentile

   $$P = \frac{100 \cdot 24}{n+1} = \frac{2400}{25} = 96$$

   Hence, percentiles $P_{97}, P_{98}, P_{99}$ are not defined either.

4. **Laptop Battery Life Comparison**

   **Model A** $(n = 20)$: $Q_1 = 7.1 + 0.25(7.2 - 7.1) = 7.125$, $Med = 7.7 + 0.5(7.8 - 7.7) = 7.75$, $Q_3 = 8.4 + 0.75(8.5 - 8.4) = 8.475$.
   **Model B** $(n = 15)$: $Q_1 = 6.1$, $Med = 6.6$, $Q_3 = 7.2$.

   a. **Model A Outliers:** $IQR = 8.475 - 7.125 = 1.35$.
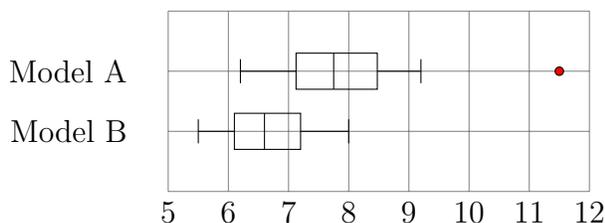      Lower fence: $= 7.125 - 1.5 \times 1.35 = 5.1$: No outliers on the left.
      Upper fence $= 8.475 + 1.5(1.35) = 10.5$: **Outlier identified: 11.5**

   b. **Model B Outliers:** $IQR = 7.2 - 6.1 = 1.1$.
      Lower fence: $= 6.1 - 1.5 \times 1.1 = 4.45$: No outliers on the left.
      Upper fence $= 7.2 + 1.5 \times 1.1 = 8.85$; No Outliers on the right.

   c. **Box-and-Whisker Plots:**

   

   d. **Comparison:** Model B has a smaller IQR (1.1 vs 1.35), suggesting it is slightly more consistent in its range, though both are similar. However, Model A generally lasts longer, as its median (7.75) is higher than Model B's entire IQR.