First Week

Q1.1

Let the number of classes be:

$$K = 15$$

Then we have the number of individuals:

$$n = \sum_{k=1}^{K} N_k = 2 + 6 + ... + 1 = 9\,465$$

Answer

- $n = 9465$

Q1.2

Let the individual be:

$$x_i, \ \ i \in \{1, 2, ..., n\}$$

Let the relative frequency be:

$$p_k = \frac{N_k}{n}$$

Then we have the mean of the sample:

$$\overline{x} = \sum_{k=1}^{K} p_k Z_k = \frac{1}{n} \sum_{k=1}^{K} N_k Z_k = \frac{1\,035\,467}{9\,465} \approx 109.400$$

Then we have the variance of the sample:

$$\sigma^2 = \sum_{k=1}^{K} p_k (Z_k - \overline{x})^2 = \frac{1}{n} \sum_{k=1}^{K} N_k (Z_k - \overline{x})^2 = \frac{1\,748\,957}{9\,465} \approx 184.782$$

Answer

- $\overline{x} = \frac{1\,035\,467}{9\,465} \approx 109.400$
- $\sigma^2 = \frac{1\,748\,957}{9\,465} \approx 184.782$

Q1.3

The individual becomes:

$$y_i = 28.349\, x_i, \ \ i \in \{1, 2, ..., n\}$$

Then the class becomes:

$$Z_{yk} = 28.349\, Z_k, \ \ k \in \{1, 2, ..., K\}$$

Then the mean becomes:

$$\begin{aligned} \overline{y} &= \frac{1}{n} \sum_{k=1}^{K} N_k Z_{yk} = 28.349\, (\frac{1}{n} \sum_{k=1}^{K} N_k Z_k) \\ &= 28.349\, \overline{x} = \frac{29\,354\,453\,983}{9\,465\,000} \approx 3\,101.369 \end{aligned}$$

Then the variance becomes:

$$\begin{aligned} \sigma_y^2 &= \frac{1}{n} \sum_{k=1}^{K} N_k (Z_{yk} - \overline{y})^2 = 28.349^2\, (\frac{1}{n} \sum_{k=1}^{K} N_k (Z_k - \overline{x})^2) \\ &= 28.349^2 \sigma^2 = \frac{140\,557\,692\,831}{946\,500} \approx 148\,502.581 \end{aligned}$$

Answer

- $\overline{y} = 28.349\, \overline{x} = \frac{29\,354\,453\,983}{9\,465\,000} \approx 3\,101.369$
- $\sigma_y^2 = 28.349^2 \sigma^2 = \frac{140\,557\,692\,831}{946\,500} \approx 148\,502.581$

Q1.4

The index of the median:

$$i_m = \frac{n+1}{2} = 4\,733$$

Then we know the classes surrounding the median:

$$\sum_{k=1}^{6} N_k = 3\,049 < i_m < 5\,289 = \sum_{k=1}^{7} N_k$$

Then we have the proportions of the distances of the median to the classes surrounding:

$$\begin{aligned} p_{[6,i_{c_m}]} &= \frac{i_m - \sum_{k=1}^{6} N_k}{N_7} = \frac{1\,684}{2\,240} \approx 0.752 \\ p_{[i_{c_m},7]} &= \frac{\sum_{k=1}^{7} N_k - i_m}{N_7} = \frac{556}{2\,240} \approx 0.248 \end{aligned}$$

Then we have a possible median:

$$m_p = p_{[6,i_{c_m}]} Z_6 + p_{[i_{c_m},7]} Z_7 = \frac{226\,208}{2\,240} \approx 100.986$$

As well as the second quartile:

$$Q_{2_p} = m_p = \frac{226\,208}{2\,240} \approx 100.986$$

The index of the first quartile (Q_1) and the index of the third quartile (Q_3):

$$\begin{aligned} i_{Q_1} &= \frac{i_m + \frac{1}{2}}{2} &&= 2\,366.75 \\ i_{Q_3} &= \frac{i_m + \frac{1}{2} + n}{2} &&= 7\,099.25 \end{aligned}$$

Then we know the classes surrounding the quartiles:

$$\begin{aligned} \sum_{k=1}^{5} N_k = 1\,320 < & i_{Q_1} < 3\,049 = \sum_{k=1}^{6} N_k \\ \sum_{k=1}^{7} N_k = 5\,289 < & i_{Q_3} < 7\,296 = \sum_{k=1}^{8} N_k \end{aligned}$$

Then we have the proportions of the distances of the quartiles to the classes surrounding:

$$p_{[5,i_{c_{Q_1}}]} = \frac{i_{Q_1} - \sum_{k=1}^{5} N_k}{N_6} = \frac{104\,675}{172\,900} \approx 0.605$$
$$p_{[i_{c_{Q_1}},6]} = \frac{\sum_{k=1}^{6} N_k - i_{Q_1}}{N_6} = \frac{68\,225}{172\,900} \approx 0.395$$

$$p_{[7,i_{c_{Q_3}}]} = \frac{i_{Q_3} - \sum_{k=1}^{7} N_k}{N_8} = \frac{181\,025}{200\,700} \approx 0.902$$
$$p_{[i_{c_{Q_3}},8]} = \frac{\sum_{k=1}^{8} N_k - i_{Q_3}}{N_8} = \frac{19\,675}{200\,700} \approx 0.098$$

Then we have a possible first quartile and a possible third quartile:

$$Q_{1_p} = p_{[5,i_{c_{Q_1}}]} Z_5 + p_{[i_{c_{Q_1}},6]} Z_6 = \frac{16\,279\,700}{172\,900} \approx 94.157$$

$$Q_{3_p} = p_{[7,i_{c_{Q_3}}]} Z_7 + p_{[i_{c_{Q_3}},8]} Z_8 = \frac{21\,632\,300}{200\,700} \approx 107.784$$

Answer

- $m_p = \frac{226\,208}{2\,240} \approx 100.986$
- $Q_{1_p} = \frac{16\,279\,700}{172\,900} \approx 94.157$
- $Q_{2_p} = \frac{226\,208}{2\,240} \approx 100.986$
- $Q_{3_p} = \frac{21\,632\,300}{200\,700} \approx 107.784$