

Question 1

A sample is composed of 120 males and 80 females. The following table shows their age in years with the percentage distribution by gender

1.

Let the number of sample of males (n_{σ}) and the number of sample of females (n_{φ}) be:

$$n_{\sigma} = 120$$

$$n_{\varphi} = 80$$

Then we have the number of individuals of total:

$$n = n_{\sigma} + n_{\varphi} = 200$$

Let the number of classes be:

$$K = 4$$

Then let the relative frequency of males (p_{σ_k}) and the relative frequency of females (p_{φ_k}) be:

$$p_{\sigma_1} = 10\%, p_{\sigma_2} = 10\%, \dots, p_{\sigma_k}, \dots, p_{\sigma_K} = 50\%, \quad k \in \{1, 2, \dots, K\}$$

$$p_{\varphi_1} = 20\%, p_{\varphi_2} = 20\%, \dots, p_{\varphi_k}, \dots, p_{\varphi_K} = 30\%, \quad k \in \{1, 2, \dots, K\}$$

Then we have the absolute frequency of modalities of males (N_{σ_k}) and of females (N_{φ_k}):

$$N_{\sigma_k} = n_{\sigma} p_{\sigma_k}, \quad k \in \{1, 2, \dots, K\}$$

$$N_{\varphi_k} = n_{\varphi} p_{\varphi_k}, \quad k \in \{1, 2, \dots, K\}$$

As well as the absolute frequency of modalities of total:

$$N_k = N_{\sigma_k} + N_{\varphi_k}, \quad k \in \{1, 2, \dots, K\}$$

Let the year of old of the individual be denoted by y/o

Then we have the number of people that are younger than 20 years old:

$$N_{\#\{y/o \in \{0,1,\dots,19\}\}} = N_1 = 28$$

Answer

- $N_{\#\{y/o \in \{0,1,\dots,19\}\}} = 28$

2.

Let the relative frequency of total be:

$$p_k = \frac{N_k}{n}, \quad k \in \{1, 2, \dots, K\}$$

Then we have the percentage of individuals that are 50 years old or older:

$$p_{\frac{\#\{y/o \in \{50,51,\dots,89\}\}}{\#\{y/o \in \{0,1,\dots,89\}\}}} = p_4 = 42\%$$

Answer

- $p_{\frac{\#\{y/o \in \{50,51,\dots,89\}\}}{\#\{y/o \in \{0,1,\dots,89\}\}}} = 42\%$

3.

Let the number of males that are 30 years old or older be:

$$N_{\sigma_{\#\{y/o \in \{30,31,\dots,89\}\}}} = \sum_{k=3}^K N_{\sigma_k} = 96$$

Answer

- $N_{\sigma_{\#\{y/o \in \{30,31,\dots,89\}\}}} = 96$

4.

Let the classes be:

$$z_1 = y/o \in \{0, 1, \dots, 19\}$$

$$z_2 = y/o \in \{20, 21, \dots, 29\}$$

...

$$z_k$$

...

$$z_K = y/o \in \{50, 51, \dots, 89\}, \quad k \in \{1, 2, \dots, K\}$$

Let the position of the median of total be:

$$p_m = 50\%$$

We know that:

$$\sum_{k=1}^2 N_k = 28\% < p_m < 58\% = \sum_{k=1}^3 N_k$$

Then we know that z_3 includes the median

Assume that the data distributes linearly in z_3 , then we have the median:

$$m = 30 + (49 - 30) \left(\frac{p_m - \sum_{k=1}^2 N_k}{p_3} \right) = \frac{659}{15} \approx 43.933$$

Answer

- $m = \frac{659}{15} \approx 43.933$
