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 ($n_\female$) be:

$$n_\sigma = 120$$
$$n_\female = 80$$

Then we have the number of individuals of total:

$$n = n_\sigma + n_\female = 200$$

Let the number of classes be:

$$K = 4$$

Then let the relative frequency of males ($p_{\sigma k}$) and the relative frequency of females ($p_{\female k}$) be:

$$p_{\sigma 1} = 10\%, \quad p_{\sigma 2} = 10\%, \ ..., \ p_{\sigma k}, \ ..., p_{\sigma K} = 50\%, \quad k \in \{1, 2, ..., K\}$$
$$p_{\female 1} = 20\%, \quad p_{\female 2} = 20\%, \ ..., \ p_{\female k}, \ ..., p_{\female K} = 30\%, \quad k \in \{1, 2, ..., K\}$$

Then we have the absolute frequency of modalities of males ($N_{\sigma k}$) and of females ($N_{\female k}$):

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

As well as the absolute frequency of modalities of total:

$$N_k = N_{\sigma k} + N_{\varphi k}, \quad k \in \{1, 2, ..., K\}$$

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:

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

Answer

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

2.

Let the relative frequency of total be:

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

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

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

Answer

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

3.

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

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

Answer

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

4.

Let the classes be:

$$z_1 = y/o \in \{0, 1, ..., 19\}$$
$$z_2 = y/o \in \{20, 21, ..., 29\}$$
$$...$$
$$z_k$$
$$...$$
$$z_K = y/o \in \{50, 51, ..., 89\}, \quad k \in \{1, 2, ..., K\}$$

Let the position of the median of total be:

$$p_m = 50\%$$

We know that:

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

Then we know that z_3 includes the median

Let the lower boundary of z_3 be:

$$L_{z_3} = \frac{29 + 30}{2} = 29.5$$

And let the size of z_3 be:

$$w_{z_3} = \frac{(49 + 50) - (29 + 30)}{2} = 20$$

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

$$m = L_{z_3} + \left(\frac{p_m - \sum_{k=1}^{2} p_k}{p_3}\right) w_{z_3} = \frac{265}{6} = 44.167$$

Answer

- $m = \frac{265}{6} \approx 44.167$