Question 4

> The variable X represents the number of cigarettes sold per year (in hundreds per capita), and the variable represents the number of lung cancer deaths per 10,000 inhabitants in 1960. The following data points are observed

Y

1.

Let the number of sample be:

$$n = 11$$

Let the X in ascending order be:

$$X_1 = 18.20, \; X_2 = 18.24, \; ..., \; X_i, \; ..., \; X_n = 40.46, \quad i \in \{1, 2, ..., n\}$$

As well as the corresponding Y:

$$Y_1 = 17.05, \; Y_2 = 15.98, \; ..., \; Y_i, \; ..., \; Y_n = 27.27, \quad i \in \{1, 2, ..., n\}$$

Then we have the mean of X and the mean of Y:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{29\,848}{1\,100} \approx 27.135 \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^{n} Y_i \quad = \frac{2\,298}{110} \approx 20.891 \end{aligned}$$

Then we have the covariance of X and Y:

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y} \approx 23.093 > 0$$

Then we can say that Y is increasing with respect to X.

Answer

- Increasing

2.

Assume that Y approximately grows linearly with respect to X:

$$Y_i \approx aX_i + b, \quad i \in \{1, 2, ..., n\}$$

Let the variance of X (σ_X^2) be:

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 \approx 40.927$$

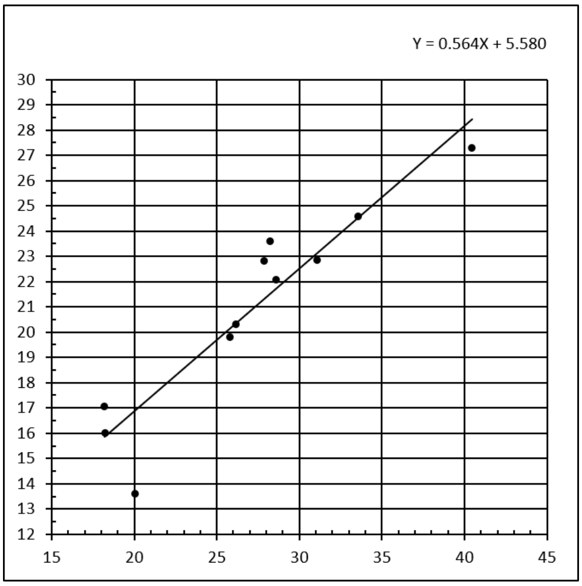
Then we have a:

$$a = \frac{\sigma_{XY}}{\sigma_X^2} \approx 0.564$$

As well as b:

$$b = \bar{Y} - \frac{\sigma_{XY}}{\sigma_X^2} \bar{X} = \bar{Y} - a\bar{X} \approx 5.580$$

Then we have the graph:



Answer

- $Y_i \approx 0.564X_i + 5.580, \quad i \in \{1, 2, ..., 11\}$
-