Question 4

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

1.

Let the number of samples be:

$$n = 11$$

Let the values of X in ascending order be:

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

As well as the corresponding values of Y:

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

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

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{29,848}{1,100} \approx 27.135$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{2,298}{110} \approx 20.891$$

Then, we compute the covariance of X and Y:

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

Hence, we can conclude that Y increases with respect to X.
Answer: Increasing

2.

Assume that Y approximately grows linearly with respect to X:

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

Let the variance of X (denoted 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 calculate a as:

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

Next, we calculate b as:

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

Thus, the linear relationship between Y and X is given by:

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

Answer:

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

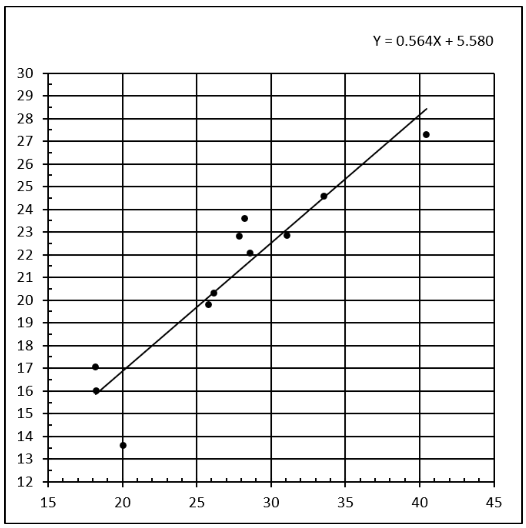

Figure 1: Graph of the relationship between Y and X