THEORETICAL PART:

Definition (Linear Function):

A **linear function** in the variable x is any function that can be written in the form

$$f(x) = mx + b,$$

where m and b are real numbers. If $m \neq 0$, $f(x) = mx + b$ is also called a **first-degree polynomial function**.

Linear Regression:

Goal: for a given number of points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, we would like to find the equation $y = mx + b$ whose graph comes closest to "fitting" the points.

We will use a **Least-Squares Method**:

- Calculate the averages of x- and y-values : $\bar{x} = \dfrac{x_1 + \cdots + x_n}{n}$ and $\bar{y} = \dfrac{y_1 + \cdots + y_n}{n}$.
- Calculate $\Delta x = x - \bar{x}$ and $\Delta y = y - \bar{y}$.
- Calculate
$$\sum \Delta x \Delta y \quad \text{and} \quad \sum (\Delta x)^2$$
- Calculate the slope m and y-intercept b for the linear regression line of best fit:
$$m = \frac{\sum \Delta x \Delta y}{\sum (\Delta x)^2} \quad \text{and} \quad b = \bar{y} - m\bar{x}$$

Important question to ask: given a collection of points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, we want to know if this collection shows a linear dependence of y on x.

The **Pearson correlation coefficient r** is a number that allows us to answer this question objectively.

We compute
$$r = \frac{\sum \Delta x \Delta y}{\sqrt{\sum (\Delta x)^2}\sqrt{\sum (\Delta y)^2}}$$

We have that always $-1 \leq r \leq 1$.

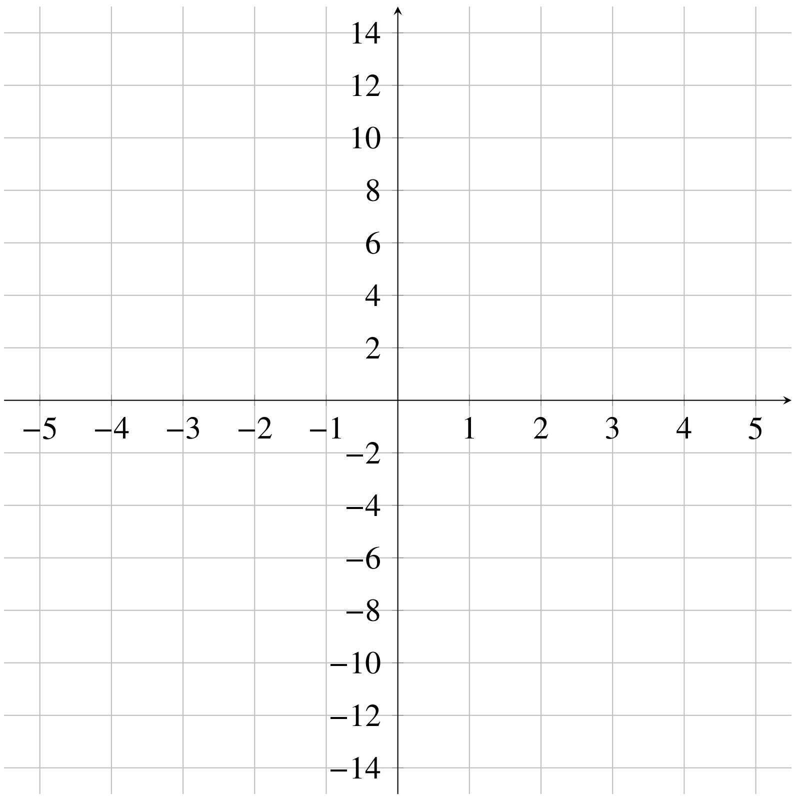
If $r = 0$, then there is no linear dependence of y on x. If $|r| \approx 1$, then there is a strong linear dependence.

PRACTICAL PART:

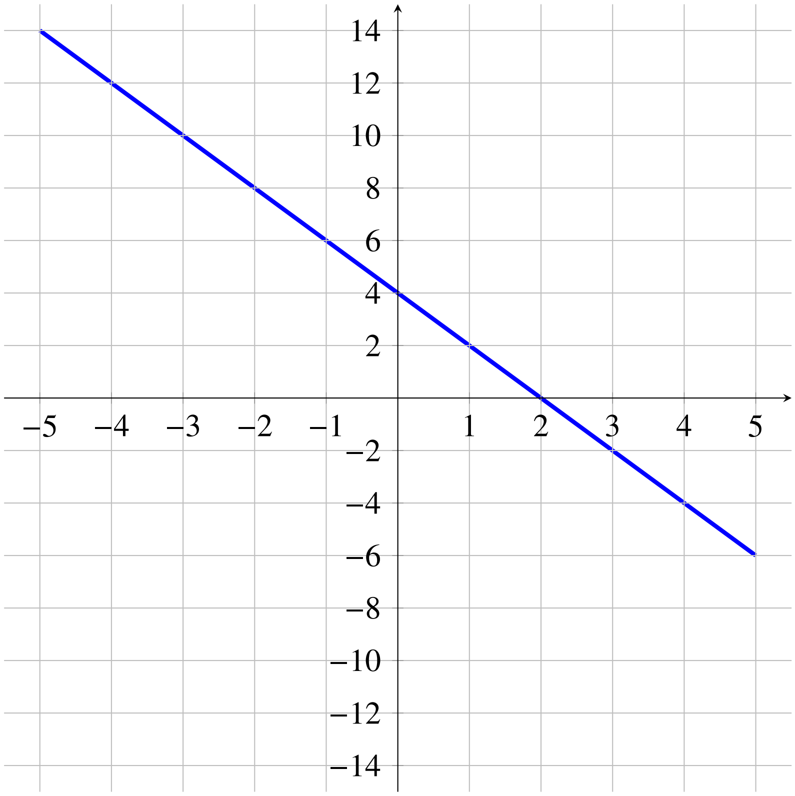
1. Graph the following linear functions:

   (a) $f(x) = 3x + 2$

   (b) $g(x) = 3$



2. Find a formula for the linear function whose graph is given below.



3. Given the collection of points

$$\{(-1, 6), (1, 5), (2, 4), (3, 2), (5, 1)\}$$

   (a) Use linear regression to find and graph the line of best fit.

   (b) Find the Pearson correlation coefficient r.