Solutions

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:

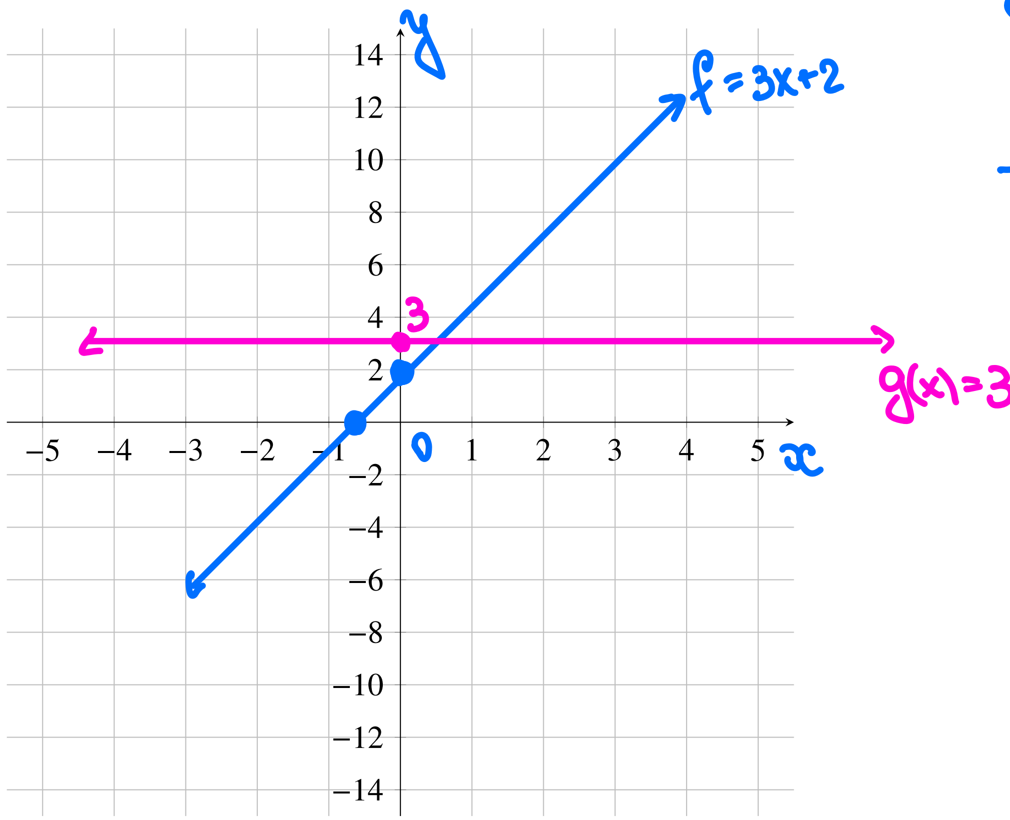
1. Graph the following linear functions:

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

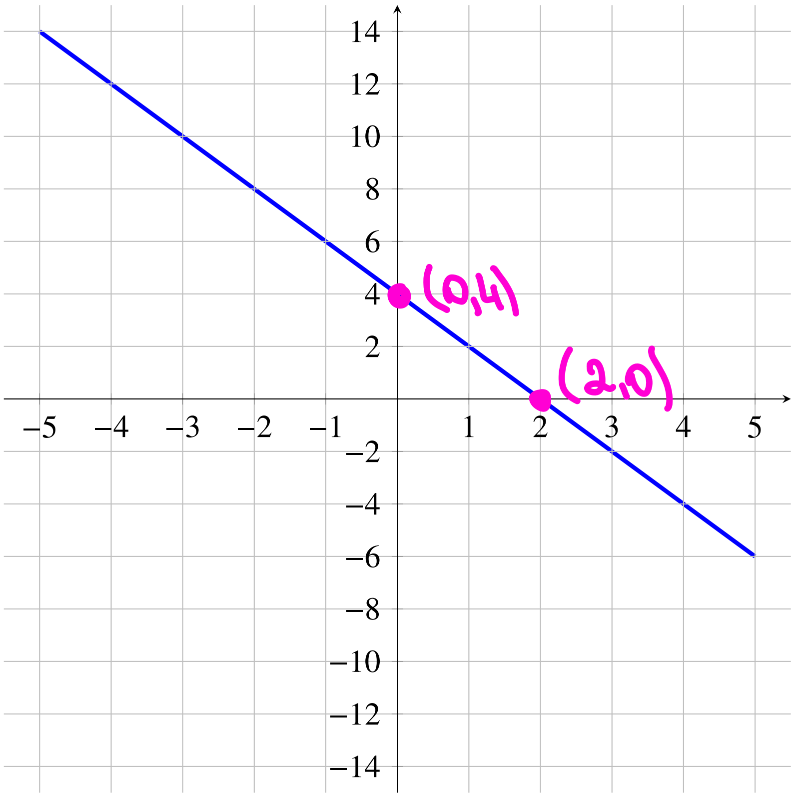
   (b) $g(x) = 3$

(a) $y = 3x + 2$

x	y
0	2
$-\frac{2}{3}$	0



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



$y = mx + b$

$A(2,0)$
$B(0,4)$

$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{4-0}{0-2} = -2$

$y = -2x + b$

$0 = -4 + b \Rightarrow b = 4$

$\boxed{y = -2x + 4}$

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.

(a) $\bar{x} = \dfrac{\cancel{-1}+\cancel{1}+2+3+5}{5} = \dfrac{10}{5} = 2$

$\bar{y} = \dfrac{6+5+4+2+1}{5} = \dfrac{18}{5}$

$\Delta x = x - \bar{x}$	$\Delta y = y - \bar{y}$	
$-1-2=-3$	$6-\frac{18}{5}$	$12/5$
-1	$5-\frac{18}{5}$	$7/5$
0	$4-\frac{18}{5}$	$2/5$
1	$2-\frac{18}{5}$	$-8/5$
3	$1-\frac{18}{5}$	$-13/5$

$$m = \frac{\sum \Delta x \Delta y}{\sum (\Delta x)^2}$$

$$A = \sum \Delta x \Delta y = -3\cdot\frac{12}{5} + (-1)\cdot\frac{7}{5} + 0\cdot\frac{2}{5} + 1\cdot\left(-\frac{8}{5}\right) + 3\cdot\left(-\frac{13}{5}\right)$$

$$B = \sum (\Delta x)^2 = 9+1+0+1+9$$

$$m = \frac{A}{B}$$

$$b = \frac{18}{5} - \frac{A}{B}\cdot 2$$

$$\boxed{y = \frac{A}{B}x + \left(\frac{18}{5} - \frac{A}{B}\cdot 2\right)}$$

$$C = \sum (\Delta y)^2 = \left(\frac{12}{5}\right)^2 + \left(\frac{7}{5}\right)^2 + \left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2 + \left(\frac{13}{5}\right)^2$$

$$r = \frac{A}{B \cdot C}$$

- if $r = 0$ $\Rightarrow$ there is no linear dependence of y on x
- if $|r| \approx 1$ $\Rightarrow$ there is a strong linear dependence