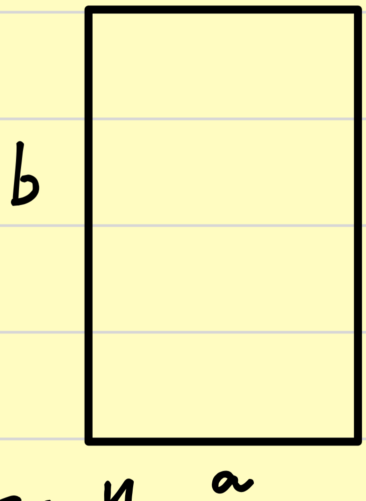
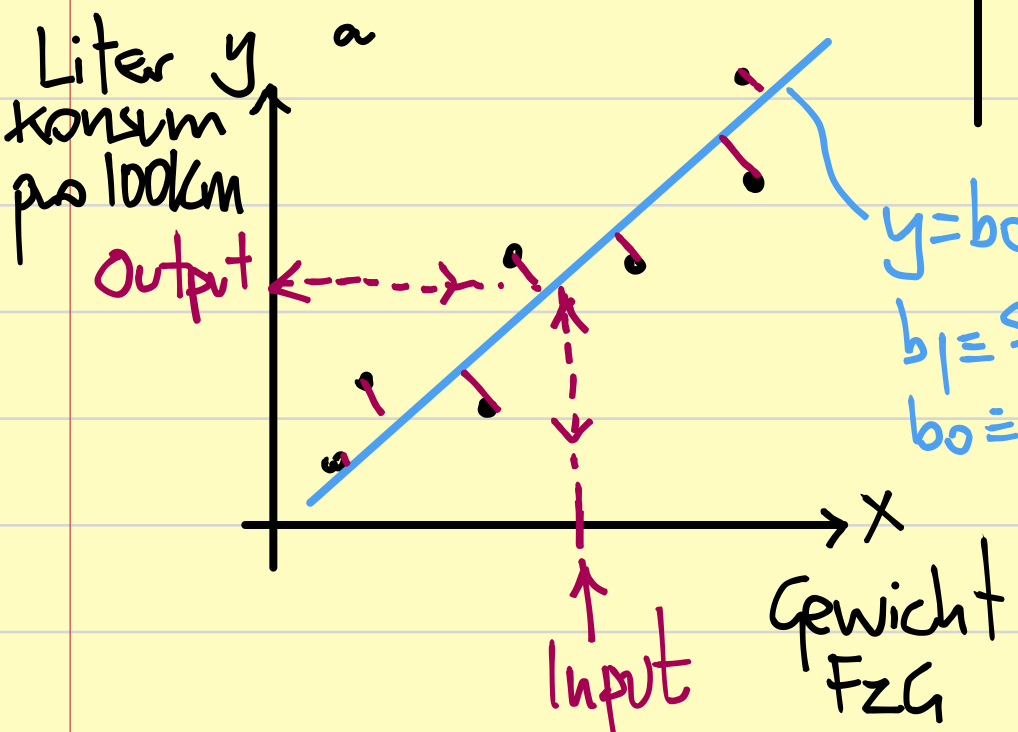
1. LINEARE REGRESSION



Fläche = a · b

Wir streben die Ermittlung einer (Prognose) Linie $y = b_0 + b_1 x$, dessen Abstand zu den Daten minimal ist.



$y = b_0 + b_1 x$

b_1 = Steigung

b_0 = wenn $x = 0$, $y = b_0$

Schritte der Linearen Regression (Prognose)

1. Datenermittlung

	x	y
1.	3	6'5
2.	4	8'5
3.	6	13
4.	3	3'5

2. Mittelwert der Variabeln (Schwerpunkt des Datasets)

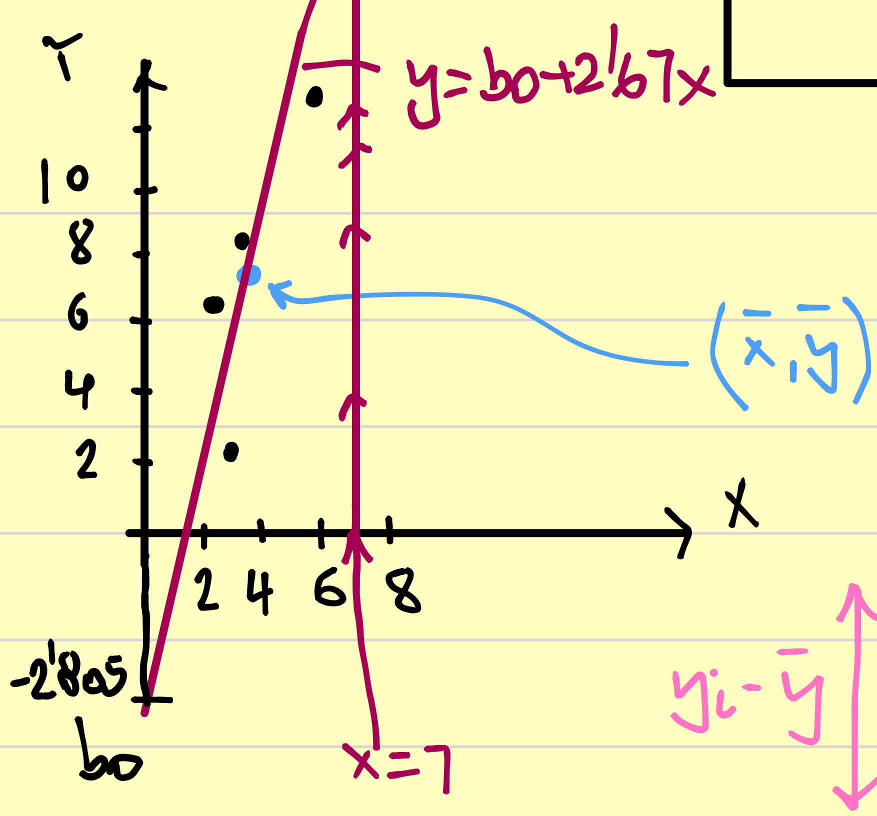
DIE LINEARE REGRESSION DURCH DEN SCHWERPUNKT GEHT:

$$\bar{y} = b_0 + b_1 \bar{x}$$

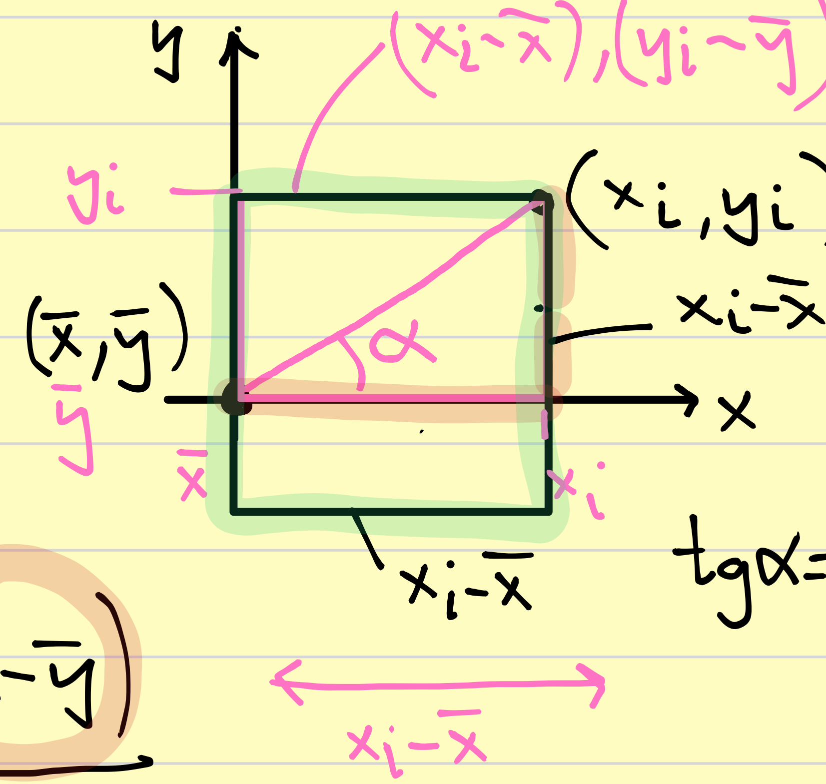
$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{3+4+6+3}{4} = 4$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i = \frac{6'5+8'5+13+3'5}{4} = 7'875$$

15'885



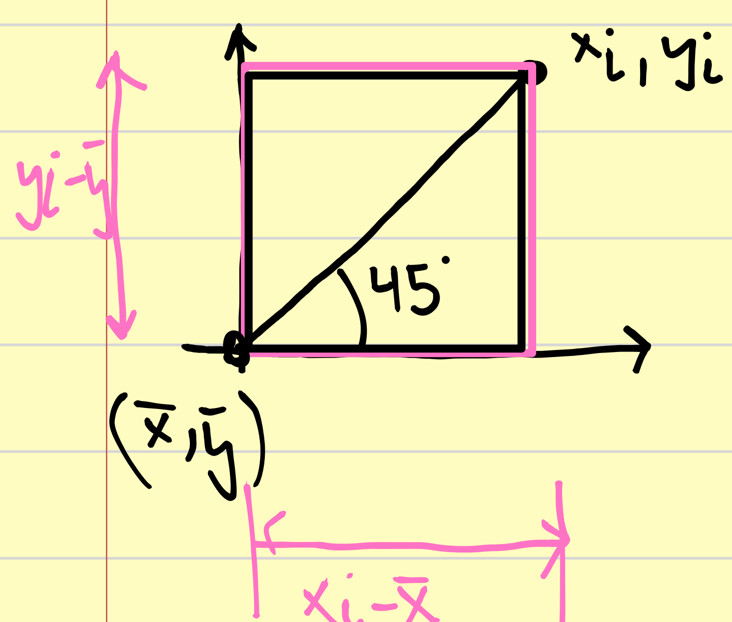
3. NEIGUNG DER LINEAREN REGRESSIONSLINIE



$y_i-\bar{y}$

$$tg\alpha=\frac{y_i-\bar{y}}{x_i-\bar{x}}$$

$$b_1=\frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$



$\square=(x_i-\bar{x})(y_i-\bar{y})$

$\square=(x_i-\bar{x})(x_i-\bar{x})$

$$b_1=\frac{(3-4)(6'5-7'875)+(4-4)(8'5-7'875)+(6-4)(13-7'875)+(3-4)(3'5-7'875)}{(3-4)^2+(4-4)^2+(6-4)^2+(3-4)^2}=2'67$$

$y=b_0+2'67\cdot x$

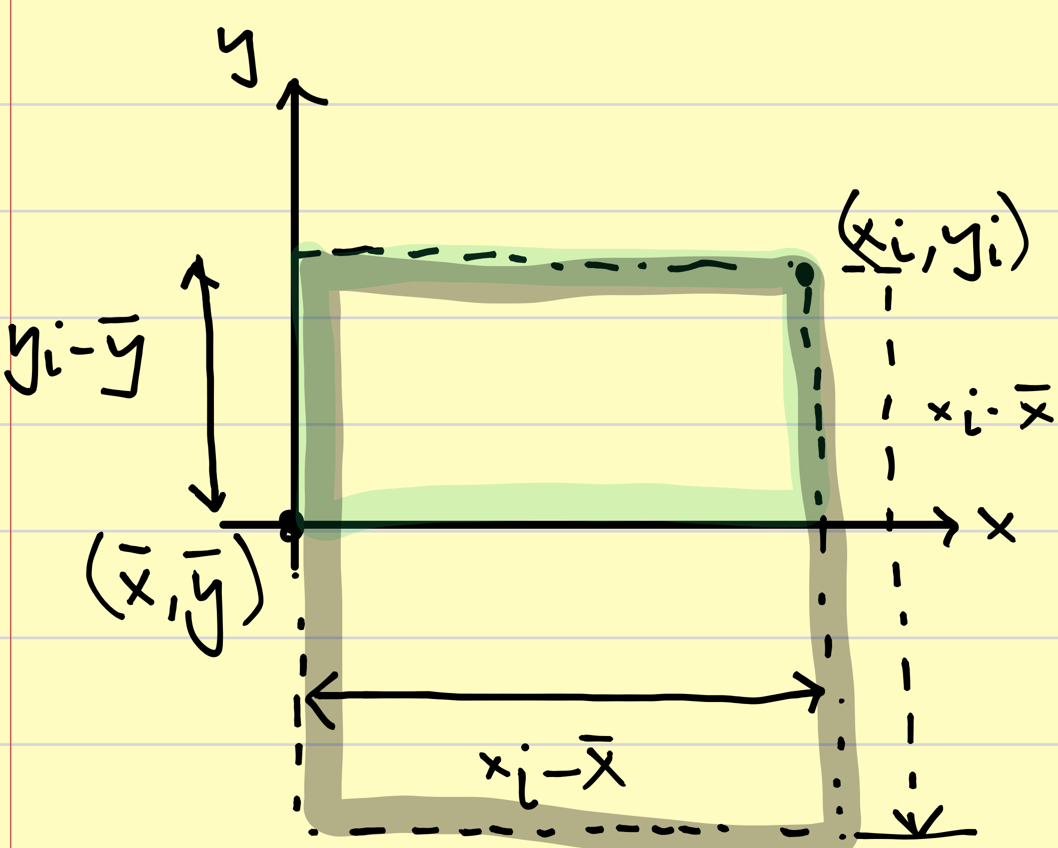
4. ERMITTLUNG VON b_0:

$$\bar{y} = b_0 + b_1 \bar{x} \rightarrow 7'875 = b_0 + 2'67 \cdot 4 \rightarrow b_0 = -2'805$$

$$\boxed{y = -2'805 + 2'67 \cdot x}$$ Lineare Regression

- Was ist der Wert von y bei $x = 7$?

$$y = -2'805 + 2'67 \cdot 7 = 15'885$$



$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = b_1$$

2. LOGISTISCHE REGRESSION

1. Ausgangspunkt Lineare Regression

sie wird verwendet, um den Zusammenhang zw. einer abhängigen Variablen (z.B. Konsum) und einer oder mehreren unabhängigen Variablen (z.B. Gewicht) zu beschreiben.

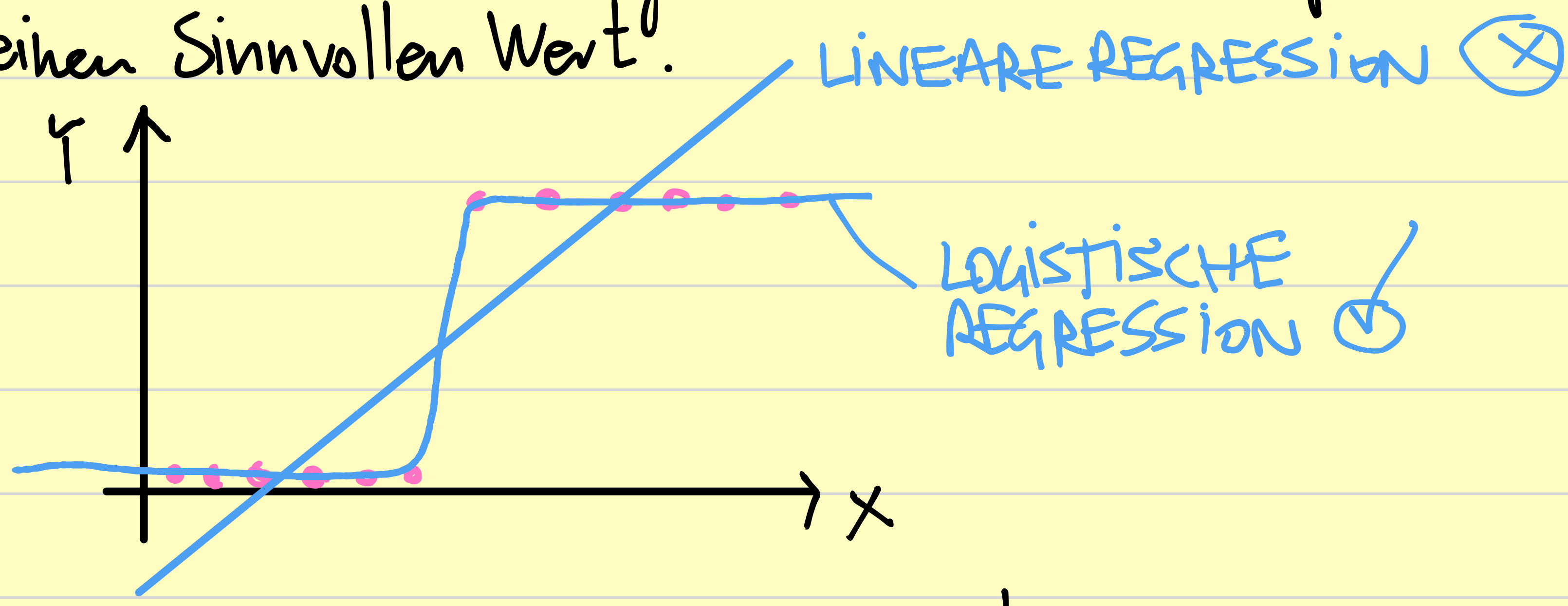
Die lineare Regression hat die allgemeine Form:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$$

2. Grenzen der linearen Regression.

Ein Problem tritt auf, wenn unsere Zielvariable nur zwei Werte annehmen kann.

In diesen Fällen liefert die lineare Regression keinen sinnvollen Wert.



3. Von linearen zur logistischen Regression.

Um Wahrscheinlichkeiten zw. 0 & 1 zu erhalten, nutzen wir eine spezielle Funktion, die s.g. LOGISTISCHE FUNKTION (SIGMOID FUNKTION).

$$\boxed{\sigma(z) = \frac{1}{1+e^{-z}} \qquad z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n}$$

- $b_0, b_1, b_2, \ldots, b_n$ sind Koeffizienten.
- $X_1, X_2, \ldots, X_n$ sind unabhängige Einflußfaktoren

Die Interpretation der Koeffizienten:

- Positiver Koeffizient ($b_i > 0$): Je größer der Faktor, desto höher die Wahrscheinlichkeit.
- Negativer Koeffizient ($b_i < 0$): Je größer der Faktor, desto geringer die Wahrscheinlichkeit.