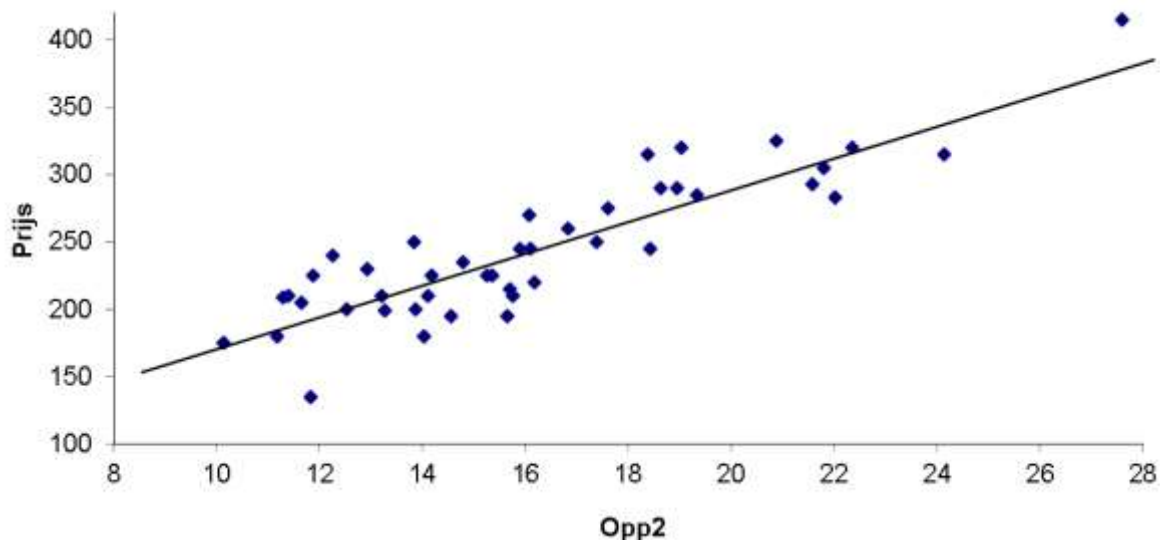


Extra uitleg bij lineaire regressie.

In de vorige paragraaf hebben we gesteld dat de correlatie het lineaire verband meet tussen 2 grootheden, m.a.w. de mate waarin de punten van de puntenwolk op een rechte liggen. Als de correlatie groot is (dicht bij 1 of -1) liggen de punten praktisch op een rechte of in een smalle strook parallel met een rechte zoals in onderstaand voorbeeld.

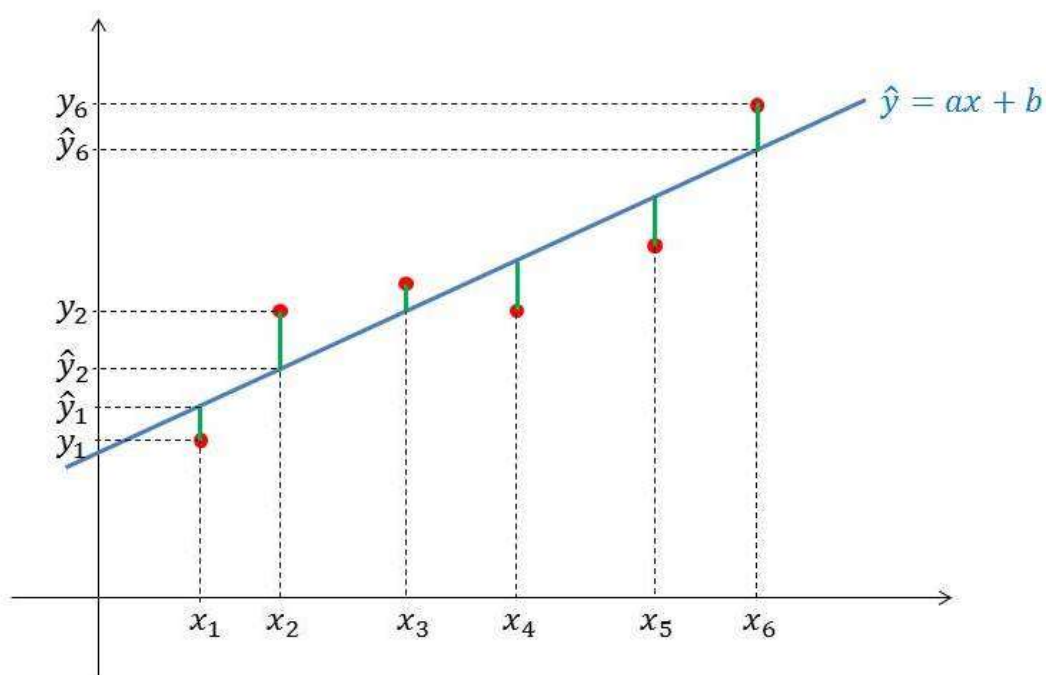


In lineaire regressie gaan we nu de vergelijking van deze rechte opstellen.

Deze is zeker van de vorm: $\hat{y} = ax + b$.

De waargenomen punten zijn (x_i, y_i) , de punten op de rechte met dezelfde x-waarden zijn dan (x_i, \hat{y}_i) met natuurlijk $\hat{y}_i = ax_i + b$.

Om de regressierechte te bepalen (m.a.w. a en b) stellen we als eis dat de som van de kwadratische afwijkingen (of afstanden) tussen de waargenomen waarden y_i en de waarden \hat{y}_i op de rechte zo klein mogelijk is. (kleinstekwadratenmethode)



Of dus:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - ax_i - b)^2$$

zo klein mogelijk.

In cursus heb je als oplossing gevonden:

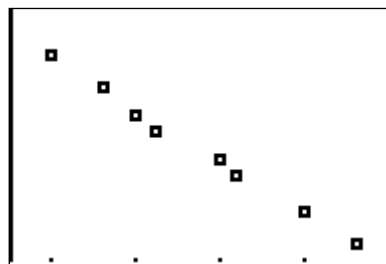
$$a = \frac{\text{cov}(x, y)}{\text{var}(x)} = \frac{\sum xy - \frac{1}{n} \sum x \sum y}{\sum x^2 - \frac{1}{n} (\sum x)^2}$$

$$b = \bar{y} - a\bar{x}$$

- Bekijken we terug het voorbeeld:

p	3.00	5.20	6.60	4.00	3.58	4.23	5.00	6.00
q	710	468	332	593	648	561	503	394

Op het screenshot zagen we al dat de punten bijna op een rechte lagen, welke rechte is dit nu...



We hadden al gevonden: $\text{cov}(p, q) = -155.55589$ en we hadden dit opgeslagen onder C Hiermee is:

$$a = \frac{\text{cov}(p, q)}{\text{var}(p)} = \text{ALPHA C} \div \text{VAR S 5: Statistics XY 3: } S_x^2 = -103.75049$$

Sla deze waarde op (je hebt ze nodig om b te berekenen): STO ALPHA A

En:

$$b = \bar{q} - a\bar{p} = \text{VAR S 5: Statistics XY 5: } \bar{y} - \text{ALPHA A} \times \text{VAR S 5: Statistics XY 2: } \bar{x} = 1013.882004$$

De vergelijking van de regressierechte is dus (afgerond tot op 4 cijfers na de komma):

$$\hat{q} = -103.7505p + 1013.8820$$

Je kan deze vergelijking ook rechtstreeks vinden met je RM via:

STAT CALC 4:LinReg(ax+b)

Geef de gepaste lijsten in en druk ENTER, de coëfficiënten a en b verschijnen.