**Hoofdstuk 6. Regressie Analyse**

Wanneer aan onderzoekselementen twee of meer kenmerken gemeten of waargenomen worden, kunnen we ons afvragen of er tussen deze kenmerken een **verband** staat.

Vb. Verband tussen studie uren en tentamen cijfer.

Als het om twee kwantitatieve kenmerken gaat, zeg x en y, dan kunnen de waarnemingsparen in een punten diagram worden weergegeven.

We spreken van regressie van x op y, indien y de afhankelijke variabele is, en x de onafhankelijke variabele.

In de regressie analyse gaat het om:

1. Het bepalen van een **wiskundige vergelijking** voor het verband tussen de variabelen welke ‘goed’ past bij de waarnemingen in het puntendiagram.
2. Het **schatten** van de waarde van de ene variabele(n) uit de waarden van de andere variabele(n).
3. De **mate** waarin de variabelen met elkaar samenhangen.

We spreken van **lineaire regressie**, indien het verband tussen de variabelen **rechtlijnig** is.

We zullen de methode van de Kleinste Kwadraten (least squares method) gebruiken voor het schatten van het model: y = α + βx.

Schatter voor enkelvoudige regressie: (formules worden in de klas gegeven)

a is een schatter voor α:

b is een schatter voor β:

De voorspellende vergelijking **y = a + bx** wordt de **regressielijn** genoemd.

**De Correlatie Coefficient:**

Indien tussen twee variabelen een linear verband wordt verondersteld, en we geinteresseerd zijn in de mate waarin de twee variabelen met elkaar samenhangen, dan gebruiken we de **correlatie coefficient**.

De corellatie coeffient beweegt zich tussen **-1 en +1** voor een maximale correlatie, en is **0** als er **geen** lineaire samenhang is. (zie bijlage met grafieken van puntendiagrammen)

Om de correlatie coeficient uit te rekenen, gebruiken we de volgende formule: (formule wordt in de klas gegeven)