**Correlatie en regressie**

**8.1a** Er is sprake van correlatie als er een oorzakelijk (causaal) verband bestaat tussen twee verschijnselen.

**8.1b** Bij nonsenscorrelatie is er sprake van een schijnbaar verband tussen twee verschijnselen zonder dat dit verband causaal is.

**8.1c** In een tijddiagram staat op de X-as de tijd als variabele aangegeven (tijdsperioden of tijdstippen). De tijd is altijd de oorzaak van veranderingen.

**8.1d** - Er kan gewerkt worden met twee y-assen (met daarbij een vaste verhouding tussen de twee waardeverdelingen op de twee tegenoverliggende assen).

- De waarden van de twee reeksen variabelen kunnen ook beide in indexcijfers worden omgerekend en weergegeven. Hierdoor kan dan met een Y-as worden volstaan.

**8.1e** Op de X-as wordt de onafhankelijke variabele (oorzaak) weergegeven.

Op de Y-as wordt de afhankelijke variabele (gevolg) weergegeven.

**8.1f** De regressielijn die door een puntenwolk getrokken kan worden, gaat altijd door het zwaartepunt van die puntenwolk.

De coördinaten van het zwaartepunt worden gevonden door de gemiddelden van respectievelijk de X- en Y-variabelen te berekenen.

**8.1g** De waarde van deze uitbijter moet bij de berekening van de vergelijking van de regressielijn buiten beschouwing gelaten worden.

**8.2a** Het inkomen € 20.000 wordt ingevuld in de lineaire vergelijking:

 p = - 0,0001 x 20.000 + 16

p = 14 (het percentage is 14%).

Bij een inkomen van € 40.000,- geldt na substitutie:

p = - 0,0001 x 40.000 + 16

p = 12 (12%).

**8.2b**

 

**8.2c** Zie figuur: het percentage is 13%.

**8.2d** De correlatie is rechtlijnig negatief.

**8.3** Bereken bij elke vergelijking een aantal punten en trek daarna pas de bijbehorende lijn.

Y = 2X + 3 bijvoorbeeld als X = 0 dan Y = 3 🡪 punt (0 , 3)

 als X = 6 dan Y = 15 🡪 punt (6 , 15)

Y = 0,5X + 7 bijvoorbeeld als X = 0 dan Y = 7 🡪 punt (0 , 7)

 als X = 10 dan Y = 2 🡪 punt (10 , 2)

Y = 3X  6 bijvoorbeeld als X = 2 dan Y = 0 🡪 punt (2 , 0)

 als X = 6 dan Y = 12 🡪 punt (6 , 12).

**

**8.4** Y = aX + b Y = aX + b Y = aX = b

8 = 16a + b 35 = 30a + b 130 = 2500a + b

5 = 4a + b - 5 = 10a + b - 92 = 600a + b

3 = 12a 30 = 20a 38 =1900a

a = 312 = 0,25 a = 30/20 = 1,5 a = 38/1900 = 0,02

invullen: invullen: invullen:

5 = 0,25 × 4 + b 5 = 1,5 × 10 + b 130 = 0,02 × 2500 + b

5 = 1 + b 5 = 15 + b 130 = 50 + b

b = 4 b = - 10 b = 80

Vergelijking: Vergelijking: Vergelijking:

Y = 0,25X + 4 Y = 1,5X - 10 Y = 0,02X + 80.

**8.5a**

 

**8.5b** Positieve correlatie.

**8.5c**



**8.5d** Het zwaartepunt vinden we door van zowel A als van B het gemiddelde uit te rekenen.

Gemiddelde van A: 31.583

Gemiddelde van B: 12.525

Zwaartepunt (31583 , 12525) zie figuur.

**8.5e** We gebruiken hier A en B in plaats van de gebruikelijke X en Y.

Als punten op de lijn nemen we bijvoorbeeld (20.000 , 5000) en (35.000 , 15.000).

B = a A + b

15.000 = 35.000 a + b

 5.000 = 20.000 a + b

10.000 = 15.000 a

a = 10.000/15.000 = 2/3 = 0,67.

Invullen:

5.000 = 20.000 × 0,67 + b

b = 5.000 − 20.000 × 0,67 = −8.400

Conclusie: B = 0,67A − 8.400 (× 1.000).

**8.5f** b = 0,67 × 25.000 − 8.400 = 8.350 (× 1.000).

**8.6** De nieuwe X en Y geven we aan met X1 en Y1. Er geldt dan:

0,001Y1 =0,18 × 0,001X1 + 3

0,001Y1 =0,00018 X1 + 3

Y1= 1.000(0,00018X1 + 3)

Y1= 0,18X1 + 3.000.

Ook andere methodes zijn hierbij mogelijk, zoals m.b.v. het invullen van concrete waarden.

**8.7a** Afhankelijke variabele: onderhoudskosten.

**8.7b**



**8.7c** Zwaartepunt: productie 295 (×1.000 stuks)

 kosten 1.873 (euro).

**8.7d** Er is sprake van een positief en (waarschijnlijk) rechtlijnig verband.

**8.7e** De punten liggen nogal verspreid; het trekken van een regressielijn zal dus wellicht verschillen per uitwerking. De genoemde punten zijn slechts een voorbeeld.

Op de getrokken lijn kiezen we bijvoorbeeld de punten (270, 1.730) en (320, 2.030).

(Ook het zwaartepunt is bruikbaar).

Y = a X + b

2.030 = 320a + b

1.730 = 270a + b

 300 = 50 a

a = 300/50 = 6,0.

Invullen:

1.730 = 270 × 6 + b

b = 1.730 − 270 × 6 = 110

Conclusie**: Y = 6X + 110**

met X is productie in 1.000 stuks

 Y is onderhoudskosten in euro’s.

**8.7f** De regressiecoëfficiënt van 6 geeft aan dat als X met 1 (in 1.000 stuks) toeneemt, Y met 6 toeneemt. Dus als de productie met 1 (× 1.000 stuks) toeneemt, nemen de onderhoudskosten met € 6,- toe. Praktisch gezien betekent dit dat de variabele onderhoudskosten € 0,006 per stuk zijn.

De constante 110 in de vergelijking geeft de vaste onderhoudskosten aan. Deze zijn onafhankelijk van de hoogte van de productie; de vaste onderhoudskosten bedragen dus € 110,-.

**8.7g** Invullen of substitutie in de vergelijking geeft:

Y = 6 × 300 + 110 = 1.910

De onderhoudskosten zullen naar verwachting ongeveer € 1.910,- bedragen.

**8.8a** 0,75 zijn de variabele kosten en 50.000 zijn de constante kosten.

**8.8b**

 

**8.8c** 1,25 is de verkoopprijs van dit product.

**8.8d** Zie TO in de grafiek van 8.8b.

**8.8e** TO = TK

1,25X = 0,75X + 50.000

 0,5X = 50.000

 X = 100.000 stuks.

**8.9a** Als Y = 0, dan C = 40.

Als Y = 100, dan C =120.

Als Y = 200 dan C = 200.

**

**8.9b** S = Y − C

S = Y − (0,8Y + 40)

S = 0,2Y − 40.

**8.9c** Als S= 0, dan Y= − 40.

Als S= 100, dan Y= − 20.

Als S= 200, dan Y= 0.

Zie grafiek.

**8.9d** De correlatie is hier rechtlijnig positief.

**8.10a** Om een duidelijk beeld te krijgen, is het beter een grafiek met twee Y-assen te tekenen.

De gemiddelde grondstofprijs is 21,5 (euro).

De gemiddelde verkoopprijs is 186,25 (euro).

De verhouding is bij benadering: 1 : 8,5.

(Gemakshalve kan ook van 1 : 10 worden uitgegaan.)

**

Uit de figuur blijkt een vertraging van 2 weken.

**8.10b** Rekening houdend met de vertraging van 2 weken hoort de grondstofprijs van week 1 bij de verkoopprijs van week 3, de grondstofprijs van week 2 bij de verkoopprijs van week 4 enz.

Zie grafiek.

**

**8.10d** Van de regressielijn nemen we twee punten (20, 180) en (24, 200),

Y = aX + b

200 = 24a + b

180 = 20a + b -

 20 = 4a

 a = 5

Invullen:

180 = 20 × 5 + b

 b = 80

Vergelijking: **Y = 5 X + 80**

 met X is grondstofprijs in week t

 en Y is verkoopprijs eindproduct in week t+2.

**8.11ab**



**8.11c** 1. De wet van Engel.

2. Naarmate het inkomen toeneemt, daalt relatief het deel van het inkomen dat aan eerste levensbehoeften wordt uitgegeven.

**8.11d**

Inkomen Bestedingen Bestedingen

 (in euro) (in euro) in % van het inkomen

 7.500 800 10,7

10.000 1.000 10,0

12.500 1.190 9,5

15.000 1.300 8,7

20.000 1.620 8,1

25.000 1.860 7,4

30.000 1.980 6,6

40.000 2.040 5,1

50.000 2.070 4,1

60.000 2.080 3,5

**8.11ef**

 

**8.11g** Negatieve (rechtlijnige) correlatie.

**8.11h** We nemen twee punten van de lijn, bijvoorbeeld (10.000, 9,5) en (60.000, 3).

Y = aX + b

9,5 = 10.000a + b

 3 = 60.000a + b -

6,5 = −50.000a

a= 6,5 ÷ −50.000 = − 0,00013.

Invullen:

3 = 60.000 × − 0,00013 + b

b= 10,8.

Vergelijking: **Y= − 0,00013X + 10,8.**

**8.12a**

 

**8.12b** Negatieve correlatie.

**8.12d** We nemen de punten (27,5;16,2) en 75;10,5).

De regressievergelijking is van het type Y = aX + b.

Er moet gelden:

16,2 = 27,5a + b (1)

10,5 = 75a +b (2)

Als we het verschil van (1) en (2) nemen, valt b weg: 5,7 = −47,5a; dus a = $\frac{5,7}{-47,5}$ = −0,12.

Nu weten we: Y = −0,12X + b.

We kunnen nu b berekenen door de coördinaten van één van de twee punten in te vullen:

10,5 = −0,12 × 75 + b ofwel b = 19,5;

de regressievergelijking wordt Y = −0,12X + 19,5.

Let op: dit is een schatting; m.b.v. de kleinste kwadratenmethode kan een beter resultaat worden bereikt.