

## Uitwerkingen

In de uitgewerkte voorbeelden worden vanwege de leesbaarheid afgeronde tussenresultaten gepresenteerd. De eindresultaten zijn echter altijd berekend zonder tussentijds afronden.

## Hoofdstuk 4

In hoofdstuk 4 staat in de tekst op bladzijde 68 dat rechts in vergelijking (4.28) voor  $F$  de ‘tweezijdige  $F$ -waarde uit tabel 5 van bijlage 1’ moet worden ingevuld. Dit moet zijn de ‘eenzijdige  $F$ -waarde uit tabel 4 van bijlage 1’.

Deze correctie heeft geen effecten op de tekst in hoofdstuk 4 maar wel in hoofdstuk 6.

Zie hiervoor de informatie bij hoofdstuk 6.

### Antwoord 4.1

Correctie: In de tabel in de opgave moet in de reeks voor  $x_H$  ‘0639’ worden vervangen door ‘0,639’

Hartley test voor  $k = 2$  niveaus met  $n = 10$  herhalingen.

$$x_L: s^2 = 3,61 \cdot 10^{-4}$$

$$x_H: s^2 = 2,80 \cdot 10^{-4}$$

$$F_{\max} = \frac{s_{\max}^2}{s_{\min}^2} = \frac{3,61 \cdot 10^{-4}}{2,80 \cdot 10^{-4}} = 1,29$$

De berekende  $F_{\max} = 1,29$ ; De kritische  $F_{\text{Hartley}}$  uit tabel 4.1 is 4,03 voor  $k = 2$  en  $f = n - 1 = 10 - 1 = 9$ .  $F_{\max} < F_{\text{Hartley}}$ . De varianties zijn constant te beschouwen.

### Antwoord 4.2

i	$x_i$	$y_i$	$x_i^2$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(y_i - \bar{y})^2$	$\hat{y}_i$	$(\hat{y}_i - \bar{y})^2$	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1	0,005	0,075	2,50E-05	1,00E-04	1,78E-03	3,15E-02	0,0872	0,02736	-0,0122	1,488E-04
2	0,010	0,182	1,00E-04	2,50E-05	3,53E-04	4,98E-03	0,1699	0,00684	0,0121	1,464E-04
3	0,015	0,248	2,25E-04	0,00E+00	0,00E+00	2,12E-05	0,2526	0,00000	-0,0046	2,116E-05
4	0,020	0,357	4,00E-04	2,50E-05	5,22E-04	1,09E-02	0,3353	0,00684	0,0217	4,709E-04
5	0,025	0,401	6,25E-04	1,00E-04	1,48E-03	2,20E-02	0,4180	0,02736	-0,0170	2,890E-04
som	0,0750	1,2630	1,38E-03	2,50E-04	4,14E-03	6,95E-02		0,06839		1,076E-03
gem	0,0150	0,2526		SS <sub>xx</sub>	SS <sub>xy</sub>	SS <sub>yy</sub> = SS <sub>T</sub>		SS <sub>Reg</sub>		SS <sub>Res</sub>

met  $n = 5$ ;  $p = 2$ ;  $t_{(0,05;3)} = 3,182$

SS<sub>T</sub>, SS<sub>Reg</sub>, en SS<sub>Res</sub>, zie tabel.

De helling  $b_1$  wordt berekend met (4.15):

$$b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{4,14 \cdot 10^{-3}}{2,50 \cdot 10^{-4}} = 16,54$$

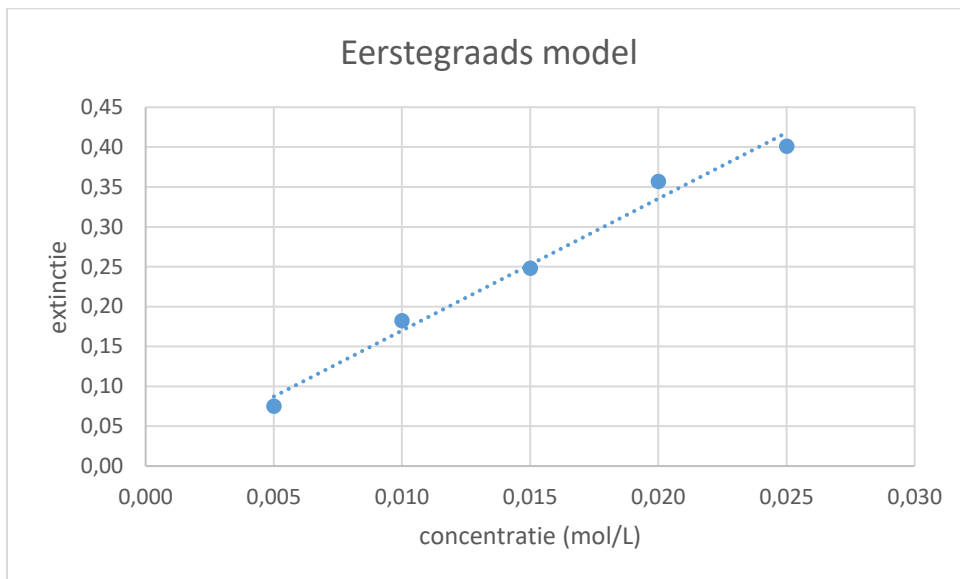
De as-afsnede  $b_0$  wordt berekend met (4.16):

$$b_0 = \bar{y} - b_1 \bar{x} = 0,2526 - 16,54 \cdot 0,0150 = 0,0045$$

De vergelijking voor de eerstegraads kalibratielijn is,

$$y = b_0 + b_1 x = 0,0045 + 16,54 \cdot x$$

De kalibratielijn is in de grafiek hierna getekend.



De standaarddeviatie van de residuen, berekend m.b.v. (4.20) is,

$$s_r = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n-p)}} = \sqrt{\frac{S_{yy}}{(n-p)}} = \sqrt{\frac{6,95 \cdot 10^{-2}}{(5-2)}} = 1,89 \cdot 10^{-2}$$

De standaarddeviaties van de regressieparameters  $b_0$  en  $b_1$  kunnen worden berekend met (4.24) en (4.25):

$$s_{b_0} = s_r \cdot \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} = s_r \cdot \sqrt{\frac{\sum x_i^2}{n \cdot SS_{xx}}} = 1,89 \cdot 10^{-2} \cdot \sqrt{\frac{1,38 \cdot 10^{-3}}{5 \cdot 2,50 \cdot 10^{-4}}} = 1,99 \cdot 10^{-2}$$

$$s_{b_1} = \frac{s_r}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{s_r}{\sqrt{SS_{xx}}} = \frac{1,89 \cdot 10^{-2}}{\sqrt{2,50 \cdot 10^{-4}}} = 1,20$$

Het bijbehorende 95%-betrouwbaarheidsinterval  $BI(b)$  voor beide parameters kan worden berekend met (4.26) en  $t = 3,182$ :

$$BI(b_0) = b_0 \pm t_{(0,05;n-p)} s_{b_0} = 0,0045 \pm 3,182 \cdot 1,99 \cdot 10^{-2} = 0,0045 \pm 0,0632$$

$$BI(b_1) = b_1 \pm t_{(0,05;n-p)} s_{b_1} = 16,54 \pm 3,182 \cdot 1,20 = 16,54 \pm 3,81$$

Merk op dat nul ingesloten ligt in het betrouwbaarheidsinterval van  $b_0$ .

De geschatte concentratie voor  $y = 0,20$  is met (4.29):

$$\hat{x}_0 = \frac{y - b_0}{b_1} = \frac{0,20 - 0,0045}{16,54} = 1,18 \cdot 10^{-2} \text{ mol/L}$$

Het betrouwbaarheidsinterval voor de geschatte concentratie bij  $y = 0,20$  is met (4.31)

$$\begin{aligned} \text{BI}(\hat{x}_0) &= \hat{x}_0 \pm \frac{t_{(0,05;n-p)} s_r}{b_1} \cdot \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b_1^2 \sum (x_i - \bar{x})^2}} = \\ &1,18 \cdot 10^{-2} \pm \frac{3,182 \cdot 1,89 \cdot 10^{-2}}{16,54} \cdot \sqrt{\frac{1}{1} + \frac{1}{5} + \frac{(0,20 - 0,2526)^2}{16,54^2 \cdot 2,50 \cdot 10^{-4}}} = \\ &= 1,18 \cdot 10^{-2} \pm 4,06 \cdot 10^{-3} \end{aligned}$$

De regressie voor het eerstegraads model kan ook worden uitgevoerd met Excel:

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9922
R-kwadraat	0,9845
Aangepaste kleinste kwadraat	0,9793
Standaardfout	0,0189
Waarnemingen	5

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	1	0,068393	0,068393	190,63	0,000822
Storing	3	0,001076	0,000359		
Totaal	4	0,069469			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,0045	0,0199	0,2265	0,8354	-0,0587	0,0677
Variabele X 1	16,5400	1,1979	13,8070	0,0008	12,7276	20,3524

Merk op dat  $b_0$  niet significant is omdat  $p(b_0) > 0,05$ .

Deze uitvoer bevat de resultaten voor  $s_r$ ,  $SS_{\text{Reg}}$ ,  $SS_{\text{Res}}$ ,  $SS_{\text{T}}$ ,  $b_0$ ,  $b_1$ ,  $s(b_0)$  en  $s(b_1)$  en de 95% betrouwbaarheidsintervallen van  $b_0$  en  $b_1$ . Deze gegevens kunnen worden gebruikt voor de berekening van de geschatte concentratie voor  $y = 0,20$  en voor het bijbehorende betrouwbaarheidsinterval zoals hiervoor is beschreven.

### Antwoord 4.3

De regressie voor het tweedegraads model, uitgevoerd met Excel, levert het volgende resultaat:

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9958
R-kwadraat	0,9916
Aangepaste kleinste kwadraat	0,9832
Standaardfout	0,0171
Waarnemingen	5

Variantieanalyse

	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	2	0,068885	0,034442	117,9076	0,0084
Storing	2	0,000584	0,000292		
Totaal	4	0,069469			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	-0,0370	0,0367	-1,0094	0,4191	-0,1947	0,1207
x	23,6543	5,5870	4,2338	0,0515	-0,3846	47,6932
x <sup>2</sup>	-237,1429	182,7142	-1,2979	0,3238	-1023,2987	549,0130

Hieruit kunnen de volgende waarden worden afgelezen:

- de modelparameters:  $b_0 = -0,0370$ ;  $b_1 = 23,65$ ;  $b_2 = -237$
- de standaarddeviatie van de residuen  $s_r = 0,0171$ .
- $SS_T = 0,069469$  ( $f = 4$ );  $SS_{Reg} = 0,068885$  ( $f = 2$ );  $SS_{Res} = 5,84 \cdot 10^{-4}$  ( $f = 2$ )

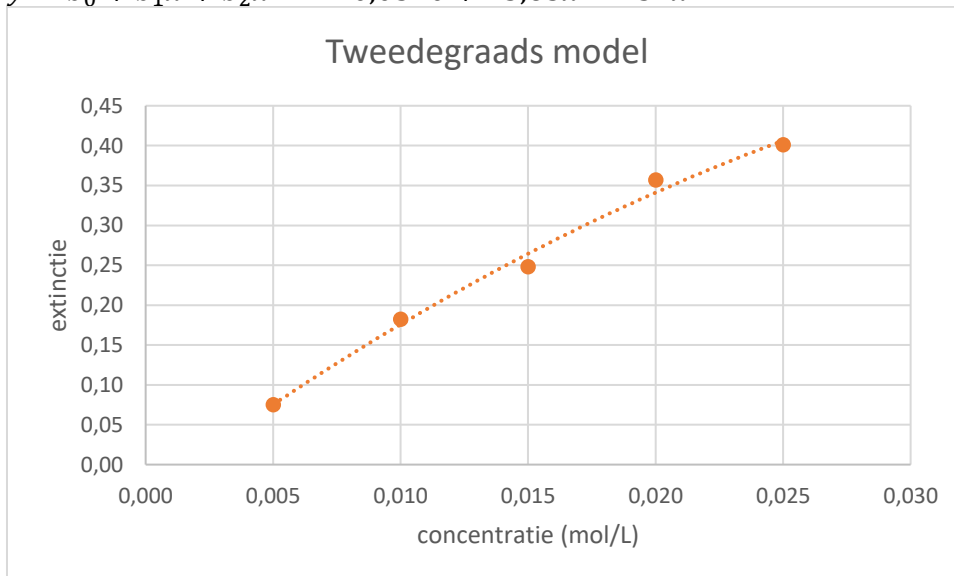
De uiterste grenzen van de 95%-betrouwbaarheidsintervallen van de modelparameters staan in de Excel regressie-uitvoer. Hieruit kan de helft van het betrouwbaarheidsinterval (die bij de antwoorden staat vermeld) worden berekend door het verschil te nemen van de hoogste waarde van het 95% betrouwbaarheidsinterval en de waarde van de parameter.

De resultaten zijn:  $b_0 = -0,0370 \pm 0,1577$ ;  $b_1 = 23,65 \pm 24,03$ ;  $b_2 = -237 \pm 786$ .

Merk op dat voor dit model zowel  $b_0$  als  $b_2$  niet significant zijn omdat  $p(b_0) > 0,05$  en  $p(b_2) > 0,05$ .

De vergelijking voor de tweedegraads kalibratielijns is,

$$y = b_0 + b_1x + b_2x^2 = -0,0370 + 23,65x - 237x^2$$



De geschatte concentratie voor  $y = 0,20$  kan worden berekend met (4.40) met een plusteken vóór het wortelteken:

$$\hat{x}_0 = \frac{-b_1 + \sqrt{b_1^2 - 4b_2(b_0 - y_0)}}{2b_2} = \frac{-23,65 + \sqrt{23,65^2 - 4 \cdot (-237) \cdot (-0,0370 - 0,20)}}{2 \cdot (-237)} = 1,72 \cdot 10^{-2} \text{ mol/L}$$

### Test1: Test op lack of fit

De test of het tweedegraads model significant beter is dan het eerstegraadsmodel kan niet worden uitgevoerd met een test op lack of fit omdat er geen replica's zijn. De overige testen uit paragraaf 4.7 kunnen wel worden uitgevoerd.

### Test2: Vergelijking van de residuvarianties (paragraaf 4.7.2) met (4.73):

$$F = \frac{s_{r1}^2}{s_{r2}^2} = \frac{MS_{\text{Res},1\text{egr}}}{MS_{\text{Res},2\text{egr}}} = \frac{0,000359}{0,000292} = 1,23$$

De gemiddelde kwadratensommen staan in de Excel uitvoer voor beide modellen.

De eenzijdige kritische waarde is  $F_{(0,05;3;2)} = 19,16$ .

De test op lineariteit slaagt omdat de variantie van het eerstegraadsmodel niet significant groter is dan die van het tweedegraadsmodel,  $F < F_{(0,05;3;2)}$ . Het tweedegraadsmodel is dus niet significant beter dan het eerstegraadsmodel.

### Test3: Test van Mandel (paragraaf 4.7.3) met (4.76):

$$F = \frac{SS_{\text{Res},1\text{egr}} - SS_{\text{Res},2\text{egr}}}{SS_{\text{Res},2\text{egr}}/(n-3)} = \frac{0,001076 - 0,000584}{0,000584/(5-3)} = 1,68$$

De kwadratensommen staan in de Excel uitvoer voor beide modellen.

De eenzijdige kritische waarde is  $F_{(0,05;1;2)} = 18,51$ .

De test op lineariteit slaagt omdat de verschilvariantie in de teller niet significant groter is dan de variantie van het tweedegraadsmodel,  $F < F_{(0,05;1;2)}$ . Het tweedegraadsmodel is dus niet significant beter dan het eerstegraadsmodel.

### Test4: Test op significantie van de $b_2$ -coëfficiënt (paragraaf 4.7.4):

De regressiecoëfficiënt  $b_2$  is niet significant omdat:

- $p(b_2) = 0,32$ . Dit is groter dan 0,05;
- het 95%-betrouwbaarheidsinterval van  $b_2$  nul omvat:  $BI(b_2) = [-1023; 549]$ ;
- de berekende  $t$ -waarde  $|t| = 1,2979$  (zie in de kolom T-statistische gegevens van het tweedegraads model). Deze is kleiner dan de kritische  $t$ -waarde  $t_{(0,05;2)} = 4,303$ .

Het tweedegraadsmodel is dus volgens deze test niet significant beter dan het eerstegraadsmodel.

### **Antwoord 4.4**

Bij dit design is het betrouwbaarheidsinterval verkleind door:

- de kalibratie uit te voeren met een groter aantal standaarden vanwege de duplo's voor  $i = 1$  en 5 ( $n$  groter en  $t$  kleiner);
- duplo's van standaarden te meten aan de uiteinden van het concentratiebereik. Dit vergroot de term  $\sum(x_i - \bar{x})^2$  in (4.31) voor het betrouwbaarheidsinterval van de geschatte concentratie.

i	j	$x_i$	$y_{ij}$	$x_i^2$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \cdot (y_{ij} - \bar{y})$	$(y_{ij} - \bar{y})^2$	$\hat{y}_i$	$(\hat{y}_i - \bar{y})^2$	$\bar{y}_i$
1	1	0,005	0,215	2,50E-05	1,00E-04	2,31E-03	0,0534	0,2307	4,642E-02	0,2065
1	2	0,005	0,198	2,50E-05	1,00E-04	2,48E-03	0,0616	0,2307	4,642E-02	0,2065
2	1	0,010	0,380	1,00E-04	2,50E-05	3,31E-04	0,0044	0,3384	1,160E-02	0,3800
3	1	0,015	0,472	2,25E-04	0,00E+00	0,00E+00	0,0007	0,4461	3,081E-33	0,4720
4	1	0,020	0,571	4,00E-04	2,50E-05	6,24E-04	0,0156	0,5539	1,160E-02	0,5710
5	1	0,025	0,645	6,25E-04	1,00E-04	1,99E-03	0,0395	0,6616	4,642E-02	0,6435
5	2	0,025	0,642	6,25E-04	1,00E-04	1,96E-03	0,0384	0,6616	4,642E-02	0,6435
som		0,1050	3,1230	2,025E-03	4,50E-04	9,70E-03	0,2135		2,089E-01	
gem		0,0150	0,4461		$SS_{xx}$	$SS_{xy}$	$SS_{yy} = SS_T$		$SS_{Reg}$	

Vervolg tabel

i	$r_i^2 = (y_{ij} - \hat{y}_i)^2$	$(\hat{y}_i - \bar{y}_i)^2$	$(y_{ij} - \bar{y}_i)^2$
1	2,464E-04	5,856E-04	7,225E-05
1	1,069E-03	5,856E-04	7,225E-05
2	1,729E-03	1,729E-03	0,000E+00
3	6,686E-04	6,686E-04	0,000E+00
4	2,936E-04	2,936E-04	0,000E+00
5	2,751E-04	3,272E-04	2,250E-06
5	3,837E-04	3,272E-04	2,250E-06
som	4,665E-03	4,516E-03	1,490E-04
gem	$SS_{Res}$	$SS_{Lof}$	$SS_{Pe}$

met  $n = 7$ ;  $k = 5$ ;  $p = 2$ ;  $t_{(0,05;5)} = 2,571$

$SS_T$ ,  $SS_{Reg}$ ,  $SS_{Res}$ ,  $SS_{Lof}$  en  $SS_{Pe}$ , zie voorgaande tabel.

De helling  $b_1$  wordt berekend met (4.15):

$$b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{9,70 \cdot 10^{-3}}{4,50 \cdot 10^{-4}} = 21,5444$$

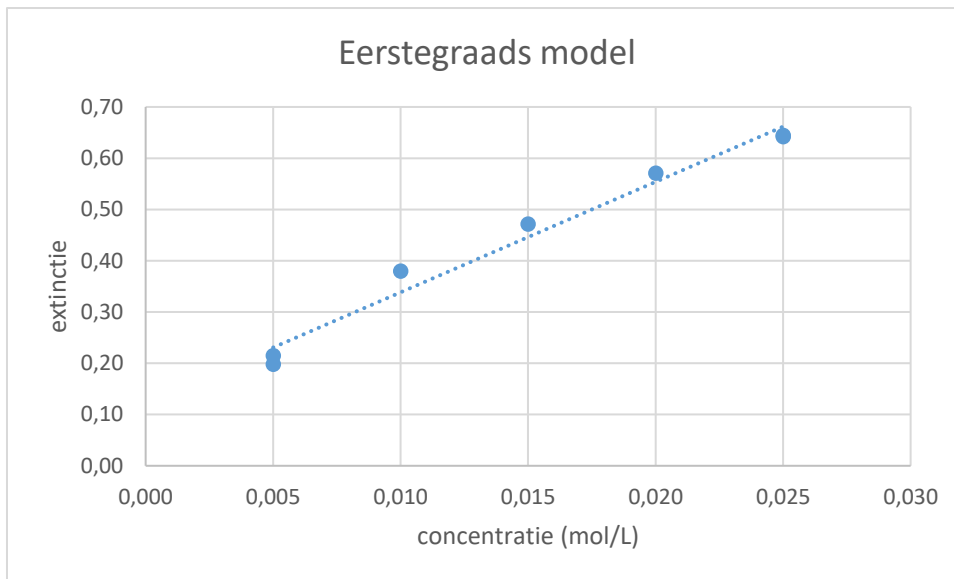
De as-afsnede  $b_0$  wordt berekend met (4.16):

$$b_0 = \bar{y} - b_1 \bar{x} = 0,4461 - 21,5444 \cdot 0,0150 = 0,1230$$

De vergelijking voor de eerstegraads kalibratielij is,

$$y = b_0 + b_1 x = 0,1230 + 21,54 \cdot x$$

De kalibratielij is in de grafiek hierna getekend.



De standaarddeviatie van de residuen, berekend m.b.v. (4.20) is,

$$s_r = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n-p)}} = \sqrt{\frac{S_{yy}}{(n-p)}} = \sqrt{\frac{0,2135}{(7-2)}} = 3,05 \cdot 10^{-2}$$

De standaarddeviaties van de regressieparameters  $b_0$  en  $b_1$  kunnen worden berekend met (4.24) en (4.25):

$$s_{b_0} = s_r \cdot \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} = s_r \cdot \sqrt{\frac{\sum x_i^2}{n \cdot SS_{xx}}} = 3,05 \cdot 10^{-2} \cdot \sqrt{\frac{2,025 \cdot 10^{-3}}{7 \cdot 4,50 \cdot 10^{-4}}} = 2,45 \cdot 10^{-2}$$

$$s_{b_1} = \frac{s_r}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{s_r}{\sqrt{SS_{xx}}} = \frac{3,05 \cdot 10^{-2}}{\sqrt{4,50 \cdot 10^{-4}}} = 1,44$$

Het bijbehorende 95%-betrouwbaarheidsinterval  $BI(b)$  voor beide parameters kan worden berekend met (4.26) en  $t = 2,571$ :

$$BI(b_0) = b_0 \pm t_{(0,05;n-p)} s_{b_0} = 0,1230 \pm 2,571 \cdot 2,45 \cdot 10^{-2} = 0,1230 \pm 0,0630$$

$$BI(b_1) = b_1 \pm t_{(0,05;n-p)} s_{b_1} = 21,54 \pm 2,571 \cdot 1,44 = 21,54 \pm 3,70$$

Merk op dat beide parameters significant zijn omdat nul niet ingesloten ligt in het betrouwbaarheidsinterval.

De geschatte concentratie voor  $y = 0,20$  is met (4.29):

$$\hat{x}_0 = \frac{y - b_0}{b_1} = \frac{0,20 - 0,1230}{21,54} = 8,22 \cdot 10^{-3} \text{ mol/L}$$

Het betrouwbaarheidsinterval voor de geschatte concentratie bij  $y = 0,20$  is met (4.31)

$$BI(\hat{x}_0) = \hat{x}_0 \pm \frac{t_{(0,05;n-p)} s_r}{b_1} \cdot \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b_1^2 \sum (x_i - \bar{x})^2}} =$$

$$8,22 \cdot 10^{-3} \pm \frac{2,571 \cdot 3,05 \cdot 10^{-2}}{21,54} \cdot \sqrt{\frac{1}{1} + \frac{1}{7} + \frac{(0,20 - 0,4461)^2}{21,54^2 \cdot 4,50 \cdot 10^{-4}}}$$

$$= 8,22 \cdot 10^{-3} \pm 4,07 \cdot 10^{-3} \text{ mol/L}$$

De regressie kan ook worden uitgevoerd met Excel:

Gegevens voor de regressie						
Meervoudige correlatiecoëfficiënt R						0,9890
R-kwadraat						0,9782
Aangepaste kleinste kwadraat						0,9738
Standaardfout						0,0305
Waarnemingen						7

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	1	2,089E-01	2,089E-01	223,8504	2,415E-05	
Storing	5	4,665E-03	9,331E-04			
Totaal	6	2,135E-01				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,1230	0,0245	5,0211	0,0040	0,0600	0,1859
Variabele X 1	21,5444	1,4400	14,9616	0,0000	17,8429	25,2460

Merk op dat  $b_0$  en  $b_1$  beide significant zijn omdat  $p(b) < 0,05$ .

Deze uitvoer bevat de resultaten voor  $s_r$ ,  $SS_{\text{Reg}}$ ,  $SS_{\text{Res}}$ ,  $SS_{\text{T}}$ ,  $b_0$ ,  $b_1$ ,  $s(b_0)$  en  $s(b_1)$  en de 95% betrouwbaarheidsintervallen van  $b_0$  en  $b_1$ . Deze gegevens kunnen worden gebruikt voor de berekening van de geschatte concentratie voor  $y = 0,20$  en voor het bijbehorende betrouwbaarheidsinterval zoals hiervoor is beschreven.

Test op lack of fit met (4.67)

$$F_{\text{Lof}} = \frac{SS_{\text{Lof}}/(k-p)}{SS_{\text{Pe}}/(n-k)} = \frac{4,516 \cdot 10^{-3}/(5-2)}{1,490 \cdot 10^{-4}/(7-5)} = 20,21$$

De kritische eenzijdige waarde  $F_{(0,05;3;2)} = 19,16$ .  $F_{\text{Lof}} > F_{\text{krit}}$ . Dit betekent dat de variantie voor *lack of fit* voor het *eerstegraadsmodel* significant groter is dan de zuivere experimentele fout. Het *eerstegraadsmodel* past dus *niet* bij de data.

#### Antwoord 4.5

De regressie, uitgevoerd met Excel, levert het volgende resultaat:

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9984
R-kwadraat	0,9967
Aangepaste kleinste kwadraat	0,9951
Standaardfout	0,0132
Waarnemingen	7



## Variantieanalyse

	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	2	2,128E-01	1,064E-01	607,3961	1,07711E-05
Storing	4	7,008E-04	1,752E-04		
Totaal	6	2,135E-01			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,0318	0,0219	1,4521	0,2201	-0,0290	0,0926
x	38,5618	3,6314	10,6190	0,0004	28,4795	48,6442
x <sup>2</sup>	-567,2464	119,2458	-4,7570	0,0089	-898,3258	-236,1669

Hieruit kunnen de volgende waarden worden afgelezen:

- de modelparameters:  $b_0 = 0,0318$ ;  $b_1 = 38,56$ ;  $b_2 = -567$
- de standaarddeviatie van de residuen  $s_r = 0,0132$ .
- $SS_T = 2,135 \cdot 10^{-1}$  ( $f = 6$ );  $SS_{\text{Reg}} = 2,128 \cdot 10^{-1}$  ( $f = 2$ );  $SS_{\text{Res}} = 7,008 \cdot 10^{-4}$  ( $f = 4$ )

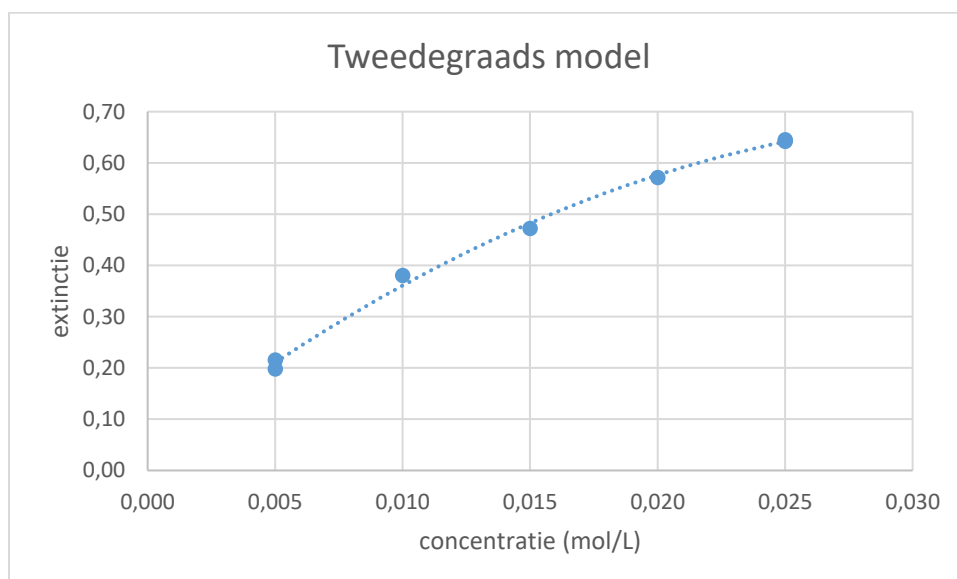
De uiterste grenzen van de 95%-betrouwbaarheidsintervallen van de modelparameters staan in de Excel regressie-uitvoer. Hieruit kan de helft van het betrouwbaarheidsinterval (die bij de antwoorden staat vermeld) worden berekend door het verschil te nemen van de hoogste waarde van het 95% betrouwbaarheidsinterval en de waarde van de parameter.

De resultaten zijn:  $b_0 = 0,0318 \pm 0,0608$ ;  $b_1 = 38,56 \pm 10,08$ ;  $b_2 = -567 \pm 331$ .

Merk op dat voor dit model  $b_0$  niet significant is omdat nul opgesloten is in het betrouwbaarheidsinterval en omdat  $p(b_0) > 0,05$ .

De vergelijking voor de tweedegraads kalibratielijn is,

$$y = b_0 + b_1x + b_2x^2 = 0,0318 + 38,56x - 567x^2$$



i	j	$x_i$	$y_{ij}$	$x_i^2$	$(y_{ij} - \bar{y})^2$	$\hat{y}_i$	$\bar{y}_i$	$(\hat{y}_i - \bar{y})^2$	$r_i^2 = (y_{ij} - \hat{y}_i)^2$	$(\hat{y}_i - \bar{y}_i)^2$	$(y_{ij} - \bar{y}_i)^2$
1	1	0,005	0,215	2,50E-05	0,0534	0,2104	0,2065	0,0556	2,080E-05	1,552E-05	7,225E-05
1	2	0,005	0,198	2,50E-05	0,0616	0,2104	0,2065	0,0556	1,547E-04	1,552E-05	7,225E-05
2	1	0,010	0,380	1,00E-04	0,0044	0,3607	0,3800	0,0073	3,723E-04	3,723E-04	0E+00
3	1	0,015	0,472	2,25E-04	0,0007	0,4826	0,4720	0,0013	1,125E-04	1,125E-04	0E+00
4	1	0,020	0,571	4,00E-04	0,0156	0,5761	0,5710	0,0169	2,652E-05	2,652E-05	0E+00
5	1	0,025	0,645	6,25E-04	0,0395	0,6413	0,6435	0,0381	1,348E-05	4,715E-06	2,250E-06
5	2	0,025	0,642	6,25E-04	0,0384	0,6413	0,6435	0,0381	4,509E-07	4,715E-06	2,250E-06
som		0,1050	3,1230	2,025E-03	0,2135			2,128E-01	7,008E-04	5,518 · 10 <sup>-4</sup>	1,490E-04
gem		0,0150	0,4461		SS <sub>yy</sub> = SS <sub>T</sub>			SS <sub>Reg</sub>	SS <sub>Res</sub>	SS <sub>Lof</sub>	SS <sub>Pe</sub>

Test op lack of fit met (4.67)

$$F_{Lof} = \frac{SS_{Lof}/(k-p)}{SS_{Pe}/(n-k)} = \frac{5,518 \cdot 10^{-4}/(5-3)}{1,490 \cdot 10^{-4}/(7-5)} = 3,70$$

De kritische eenzijdige waarde  $F_{(0,05;2;2)} = 19,00$ .  $F_{Lof} < F_{krit}$ . Dit betekent dat de variantie voor *lack of fit* voor het *tweedegraadsmodel* niet significant groter is dan de zuivere experimentele fout. Er is geen lack of fit. Het *tweedegraadsmodel* past dus bij de data.

De geschatte concentratie voor  $y = 0,30$  kan worden berekend met (4.40) met een plusteken vóór het wortelteken:

$$\hat{x}_0 = \frac{-b_1 + \sqrt{b_1^2 - 4b_2(b_0 - y_0)}}{2b_2} = \frac{-38,56 + \sqrt{38,56^2 - 4 \cdot (-567) \cdot (0,0318 - 0,30)}}{2 \cdot (-567)}$$

$$= 7,86 \cdot 10^{-3} \text{ mol/L}$$

De test of het tweedegraads model significant beter is dan het eerstegraadsmodel kan worden uitgevoerd op basis van de testen op lineariteit.

Test1: Test op lack of fit

Bij de hiervoor beschreven test op lack of fit is gebleken dat het *tweedegraadsmodel* past bij de data. Bij deze test is echter geen vergelijking gemaakt met het eerstegraads model.

Test2: Vergelijking van de residuvarianties (paragraaf 4.7.2) met (4.73):

$$F = \frac{s_{r1}^2}{s_{r2}^2} = \frac{MS_{Res,1egr}}{MS_{Res,2egr}} = \frac{9,331 \cdot 10^{-4}}{1,752 \cdot 10^{-4}} = 5,33$$

De gemiddelde kwadratensommen staan in de Excel uitvoer voor beide modellen.

De eenzijdige kritische waarde is  $F_{(0,05;5;4)} = 6,26$ .

De variantie van het eerstegraadsmodel is wel groter, maar niet significant groter, dan die van het tweedegraadsmodel,  $F < F_{(0,05;5;4)}$ . De test op lineariteit slaagt. Het eerstegraadsmodel is niet significant slechter dan het tweedegraads model.

Test3: Test van Mandel (paragraaf 4.7.3) met (4.76):

$$F = \frac{SS_{Res,1egr} - SS_{Res,2egr}}{SS_{Res,2egr}/(n-3)} = \frac{4,665 \cdot 10^{-3} - 7,008 \cdot 10^{-4}}{7,008 \cdot 10^{-4}/(7-3)} = 22,63$$

De kwadratensommen staan in de Excel uitvoer voor beide modellen.  
De eenzijdige kritische waarde is  $F_{(0,05;1;4)} = 7,71$ .

De test op lineariteit faalt omdat de verschilvariantie in de teller significant groter is dan de variantie van het tweedegraadsmodel,  $F > F_{(0,05;1;4)}$ . Het tweedegraadsmodel is dus significant beter dan het eerstegraadsmodel.

Test4: Test op significantie van de  $b_2$ -coëfficiënt (paragraaf 4.7.4):

De regressiecoëfficiënt  $b_2$  is significant omdat:

- $p(b_2) = 0,0089$ . Dit is kleiner dan 0,05;
- het 95%-betrouwbaarheidsinterval van  $b_2$  omvat niet het getal nul:  $BI(b_2) = [-898; -236]$ ;

Echter  $b_0$  is niet significant. Er zou een nieuwe regressie moeten worden uitgevoerd met een tweedegraads model zonder  $b_0$  term.

#### Antwoord 4.6

Correctie op antwoorden:

b:  $s_r = 3,83 \cdot 10^{-4}$  moet zijn  $s_r = 1,96 \cdot 10^{-2}$ .

d:  $s_r = 5,49 \cdot 10^{-5}$  moet zijn  $s_r = 7,41 \cdot 10^{-3}$ .

Eerstegraads model (a, b en c)

i	$x_i$	$y_i$	$x_i^2$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(y_i - \bar{y})^2$	$\hat{y}_i$	$(\hat{y}_i - \bar{y})^2$	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1	1	0,050	1	43,560	1,54968	0,055131	0,0632	0,049096	-0,0132	1,749E-04
2	2	0,085	4	31,360	1,11888	0,039920	0,0968	0,035345	-0,0118	1,391E-04
3	5	0,216	25	6,760	0,17888	0,004733	0,1975	0,007619	0,0185	3,418E-04
4	10	0,384	100	5,760	0,23808	0,009841	0,3654	0,006492	0,0186	3,470E-04
5	20	0,689	400	153,760	5,01208	0,163378	0,7011	0,173301	-0,0121	1,463E-04
som	38	1,4240	530	241,200	8,09760	0,273003		0,271854		1,149E-03
gem	7,60	0,2848		$SS_{xx}$	$SS_{xy}$	$SS_{yy} = SS_T$		$SS_{Reg}$		$SS_{Res}$

met  $n = 5$ ;  $p = 2$ ;  $t_{(0,05;3)} = 3,182$

$SS_T$ ,  $SS_{Reg}$ , en  $SS_{Res}$ , zie tabel.

De helling  $b_1$  wordt berekend met (4.15):

$$b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{8,09760}{241,200} = 0,0336$$

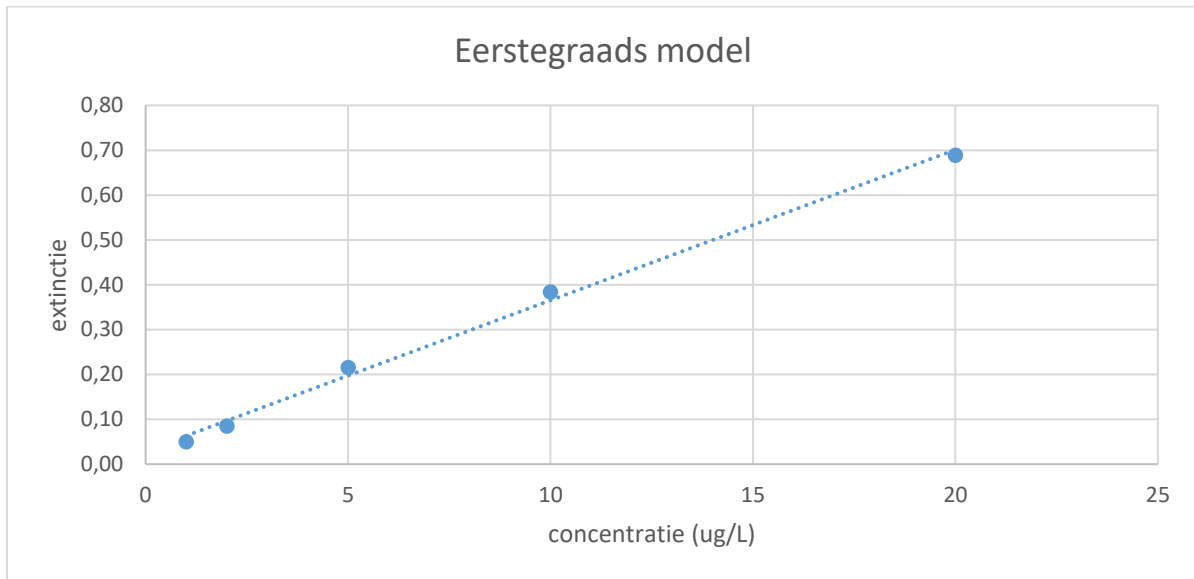
De as-afsnede  $b_0$  wordt berekend met (4.16):

$$b_0 = \bar{y} - b_1 \bar{x} = 0,2848 - 0,0336 \cdot 7,60 = 0,0297$$

De vergelijking voor de eerstegraads kalibratielijn is,

$$y = b_0 + b_1 x = 0,0297 + 0,0336 \cdot x$$

De kalibratielij is in de grafiek hierna getekend.



De standaarddeviatie van de residuen, berekend m.b.v. (4.20) is,

$$s_r = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n-p)}} = \sqrt{\frac{S_{yy}}{(n-p)}} = \sqrt{\frac{0,273003}{(5-2)}} = 1,96 \cdot 10^{-2}$$

De standaarddeviaties van de regressieparameters  $b_0$  en  $b_1$  kunnen worden berekend met (4.24) en (4.25):

$$s_{b_0} = s_r \cdot \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} = s_r \cdot \sqrt{\frac{\sum x_i^2}{n \cdot SS_{xx}}} = 1,96 \cdot 10^{-2} \cdot \sqrt{\frac{530}{5 \cdot 241,200}} = 1,30 \cdot 10^{-2}$$

$$s_{b_1} = \frac{s_r}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{s_r}{\sqrt{SS_{xx}}} = \frac{1,96 \cdot 10^{-2}}{\sqrt{241,200}} = 1,26 \cdot 10^{-3}$$

Het bijbehorende 95%-betrouwbaarheidsinterval  $BI(b)$  voor beide parameters kan worden berekend met (4.26) en  $t = 3,182$ :

$$BI(b_0) = b_0 \pm t_{(0,05;n-p)} s_{b_0} = 0,0297 \pm 3,182 \cdot 1,30 \cdot 10^{-2} = 0,0297 \pm 0,0413$$

$$BI(b_1) = b_1 \pm t_{(0,05;n-p)} s_{b_1} = 0,0336 \pm 3,182 \cdot 1,26 \cdot 10^{-3} = 0,0336 \pm 0,0040$$

Merk op dat nul ingesloten ligt in het betrouwbaarheidsinterval van  $b_0$ .

De geschatte concentratie voor  $y = 0,30$  is met (4.29):

$$\hat{x}_0 = \frac{y - b_0}{b_1} = \frac{0,30 - 0,0297}{0,0336} = 8,05 \text{ ug/L}$$

Het betrouwbaarheidsinterval voor de geschatte concentratie bij  $y = 0,30$  is met (4.31)

$$BI(\hat{x}_0) = \hat{x}_0 \pm \frac{t_{(0,05;n-p)} s_r}{b_1} \cdot \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b_1^2 \sum (x_i - \bar{x})^2}} =$$

$$8,05 \pm \frac{3,182 \cdot 1,96 \cdot 10^{-2}}{0,0336} \cdot \sqrt{\frac{1}{1} + \frac{1}{5} + \frac{(0,30 - 0,2848)^2}{0,0336^2 \cdot 241,200}} = 8,05 \pm 2,03$$

De regressie kan ook worden uitgevoerd met Excel:

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9979
R-kwadraat	0,9958
Aangepaste kleinste kwadraat	0,9944
Standaardfout	0,0196
Waarnemingen	5

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	1	2,719E-01	2,719E-01	709,7733	0,000116	
Storing	3	1,149E-03	3,830E-04			
Totaal	4	2,730E-01				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,0297	1,30E-02	2,2855	0,1064	-0,0116	0,0709
Variabele X 1	0,0336	1,26E-03	26,6416	0,0001	0,0296	0,0376

Merk op dat  $b_0$  niet significant is omdat  $p(b_0) > 0,05$ . Bij opgave 4.7 wordt een nieuwe regressie uitgevoerd zonder  $b_0$  term.

Deze uitvoer bevat de resultaten voor  $s_r$ ,  $SS_{Reg}$ ,  $SS_{Res}$ ,  $SS_T$ ,  $b_0$ ,  $b_1$ ,  $s(b_0)$  en  $s(b_1)$  en de 95% betrouwbaarheidsintervallen van  $b_0$  en  $b_1$ . Deze gegevens kunnen worden gebruikt voor de berekening van de geschatte concentratie voor  $y = 0,30$  en voor het bijbehorende betrouwbaarheidsinterval zoals hiervoor is beschreven.

### Tweedegraads model (d)

#### Regressie met Excel

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9998
R-kwadraat	0,9996
Aangepaste kleinste kwadraat	0,9992
Standaardfout	7,41E-03
Waarnemingen	5

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	2	2,729E-01	1,364E-01	2484,362	0,0004024	
Storing	2	1,098E-04	5,492E-05			
Totaal	4	2,730E-01				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	7,487E-03	7,078E-03	1,0578	0,4010	-0,0230	0,0379
x	4,212E-02	2,022E-03	20,8272	0,0023	0,0334	0,0508
x <sup>2</sup>	-4,038E-04	9,284E-05	-4,3499	0,0490	-0,0008	-0,000004

Hieruit kunnen de volgende waarden worden afgelezen:

- de modelparameters:  $b_0 = 7,487 \cdot 10^{-3}$ ;  $b_1 = 4,212 \cdot 10^{-2}$ ;  $b_2 = -4,038 \cdot 10^{-4}$
- de standaarddeviatie van de residuen  $s_r = 7,41 \cdot 10^{-3}$ .
- $SS_T = 2,730 \cdot 10^{-1}$  ( $f = 4$ );  $SS_{\text{Reg}} = 2,729 \cdot 10^{-1}$  ( $f = 2$ );  $SS_{\text{Res}} = 1,098 \cdot 10^{-4}$  ( $f = 2$ )

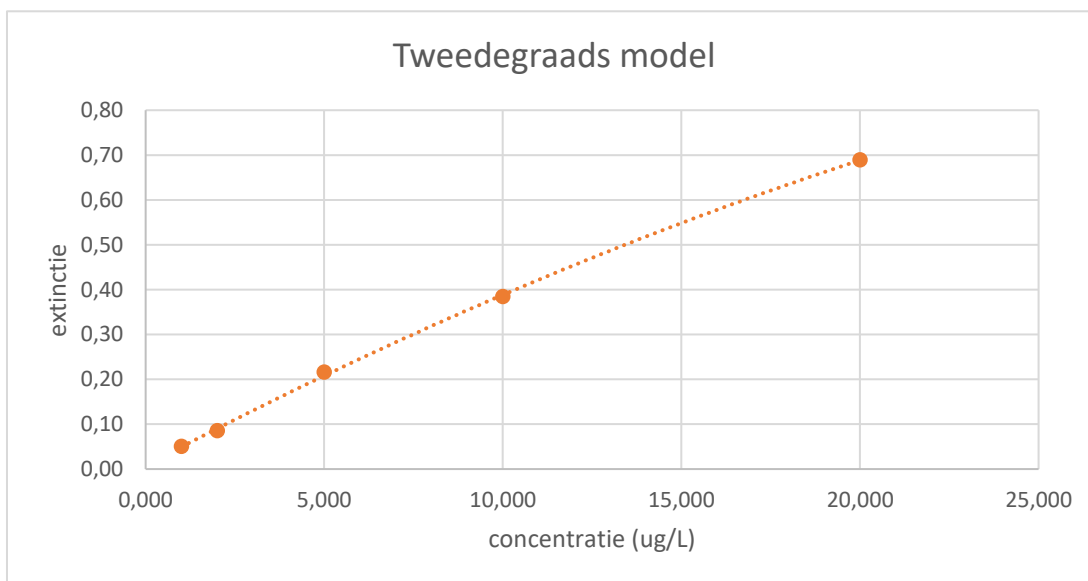
De uiterste grenzen van de 95%-betrouwbaarheidsintervallen van de modelparameters staan in de Excel regressie-uitvoer. Hieruit kan de helft van het betrouwbaarheidsinterval (die bij de antwoorden staat vermeld) worden berekend door het verschil te nemen van de hoogste waarde van het 95% betrouwbaarheidsinterval en de waarde van de parameter.

De resultaten zijn:  $b_0 = 7,487 \cdot 10^{-3} \pm 3,05 \cdot 10^{-2}$ ;  $b_1 = 4,212 \cdot 10^{-2} \pm 8,70 \cdot 10^{-3}$ ;  $b_2 = -4,038 \cdot 10^{-4} \pm 3,99 \cdot 10^{-4}$

Merk op dat voor dit model  $b_0$  niet significant is omdat nul opgesloten is in het betrouwbaarheidsinterval en omdat  $p(b_0) > 0,05$ . Bij opgave 4.7 wordt een nieuwe regressie uitgevoerd zonder  $b_0$  term.

De vergelijking voor de tweedegraads kalibratielijn is,

$$y = b_0 + b_1x + b_2x^2 = 7,49 \cdot 10^{-3} + 4,21 \cdot 10^{-2}x - 4,04 \cdot 10^{-4}x^2$$



Test2: Vergelijking van de residuvarianties (paragraaf 4.7.2) met (4.73):

$$F = \frac{s_{r1}^2}{s_{r2}^2} = \frac{MS_{\text{Res},1\text{egr}}}{MS_{\text{Res},2\text{egr}}} = \frac{3,830 \cdot 10^{-4}}{5,492 \cdot 10^{-5}} = 6,97$$

De gemiddelde kwadratensommen staan in de Excel uitvoer voor beide modellen.

De eenzijdige kritische waarde is  $F_{(0,05;3;2)} = 19,16$ .

De variantie van het eerstegraadsmodel is wel groter, maar niet significant groter, dan die van het tweedegraadsmodel,  $F < F_{(0,05;3;2)}$ . De test op lineariteit slaagt.

Test3: Test van Mandel (paragraaf 4.7.3) met (4.76):

$$F = \frac{SS_{\text{Res},1\text{egrd}} - SS_{\text{Res},2\text{egrd}}}{SS_{\text{Res},2\text{egrd}}/(n-3)} = \frac{1,149 \cdot 10^{-3} - 1,098 \cdot 10^{-4}}{1,098 \cdot 10^{-4}/(5-3)} = 18,92$$

De kwadratensommen staan in de Excel uitvoer voor beide modellen.

De eenzijdige kritische waarde is  $F_{(0,05;1;2)} = 18,51$ .

De test op lineariteit faalt omdat de verschilvariantie in de teller significant groter is dan de variantie van het tweedegraadsmodel,  $F > F_{(0,05;1;2)}$ . Het tweedegraadsmodel is dus significant beter dan het eerstegraadsmodel.

Test4: Test op significantie van de  $b_2$ -coëfficiënt (paragraaf 4.7.4):

De regressiecoëfficiënt  $b_2$  is significant omdat:

- $p(b_2) = 0,049$ . Dit is kleiner dan 0,05;
- het 95%-betrouwbaarheidsinterval van  $b_2$  omvat niet het getal nul:  $BI(b_2) = [-0,0008; -0,000004]$ ;

Echter  $b_0$  is niet significant. Er zou een nieuwe regressie moeten worden uitgevoerd met een tweedegraads model zonder  $b_0$  term, zie opgave 4.7.

#### Antwoord 4.7

#### Eerstegraads model door de oorsprong

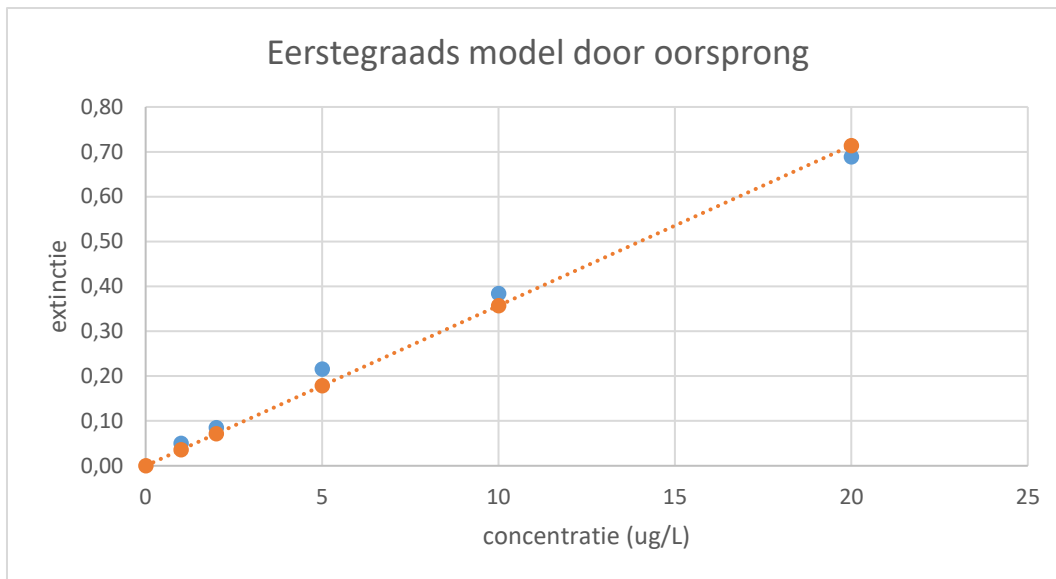
Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9977
R-kwadraat	0,9954
Aangepaste kleinste kwadraat	0,7454
Standaardfout	0,0281
Waarnemingen	5

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	1	6,754E-01	6,754E-01	857,7435	8,74E-05
Storing	4	3,150E-03	7,874E-04		
Totaal	5	6,786E-01			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0	#N/B	#N/B	#N/B	#N/B	#N/B
Variabele X 1	3,570E-02	1,219E-03	2,929E+01	8,092E-06	3,231E-02	3,908E-02

Hieruit kunnen de volgende waarden worden afgelezen:

- de modelparameter:  $b_1 = 3,570 \cdot 10^{-2}$ ;  $s(b_1) = 1,219 \cdot 10^{-3}$ ;  $p(b_1) = 8,09 \cdot 10^{-6}$
- de standaarddeviatie van de residuen  $s_r = 0,0281$ .



### Tweedegraads model door de oorsprong

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9999
R-kwadraat	0,9997
Aangepaste kleinste kwadraat	0,6663
Standaardfout	7,556E-03
Waarnemingen	5

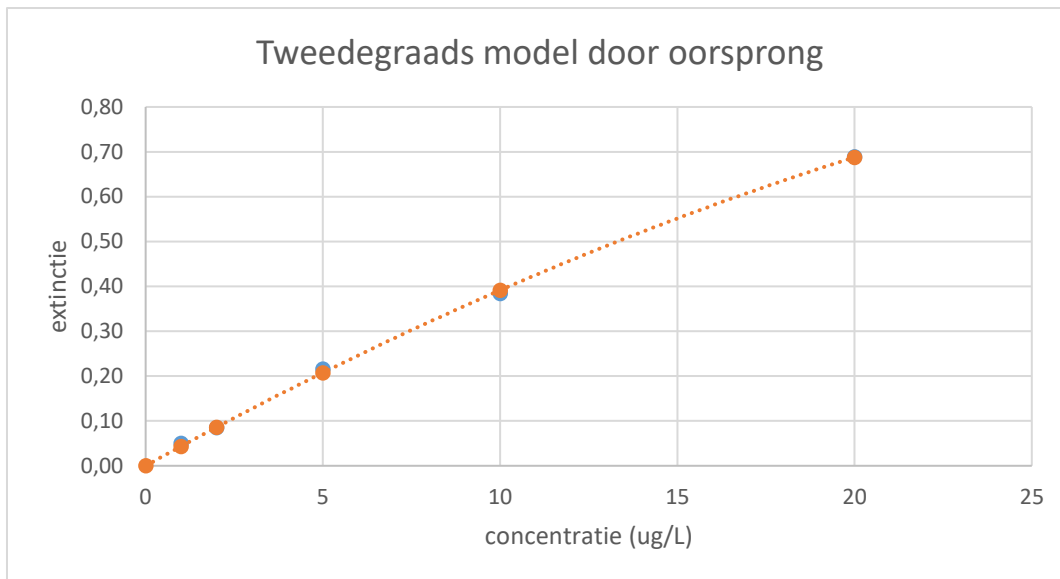
Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	2	6,784E-01	3,392E-01	5940,452	0,0001683
Storing	3	1,713E-04	5,710E-05		
Totaal	5	6,786E-01			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0	#N/B	#N/B	#N/B	#N/B	#N/B
x	4,388E-02	1,179E-03	37,22	4,267E-05	4,012E-02	4,763E-02
x <sup>2</sup>	-4,745E-04	6,570E-05	-7,22	5,473E-03	-6,836E-04	-2,654E-04

Hieruit kunnen de volgende waarden worden afgelezen:

- de modelparameters:  $b_1 = 4,388 \cdot 10^{-2}$ ;  $b_2 = -4,745 \cdot 10^{-4}$ ;  $s(b_1) = 1,179 \cdot 10^{-3}$ ;  $s(b_2) = 6,570 \cdot 10^{-5}$ ;  $p(b_1) = 4,27 \cdot 10^{-5}$ ;  $p(b_2) = 5,47 \cdot 10^{-3}$
- de standaarddeviatie van de residuen  $s_r = 7,556 \cdot 10^{-3}$ .





Test2: Vergelijking van de residuvarianties (paragraaf 4.7.2) met (4.73):

$$F = \frac{s_{r1}^2}{s_{r2}^2} = \frac{MS_{Res,1egr}}{MS_{Res,2egr}} = \frac{7,874 \cdot 10^{-4}}{5,710 \cdot 10^{-5}} = 13,79$$

De gemiddelde kwadratensommen staan in de Excel uitvoer voor beide modellen.  
De eenzijdige kritische waarde is  $F_{(0,05;4;3)} = 9,12$ .

De variantie van het eerstegraadsmodel is significant groter dan die van het tweedegraadsmodel,  $F > F_{(0,05;4;3)}$ . De test op lineariteit faalt.

Test3: Test van Mandel (paragraaf 4.7.3) met (4.76), aangepast met  $(n-2)$ :

$$F = \frac{SS_{Res,1egr} - SS_{Res,2egr}}{SS_{Res,2egr}/(n-2)} = \frac{3,150 \cdot 10^{-3} - 1,713 \cdot 10^{-4}}{1,713 \cdot 10^{-4}/(5-2)} = 52,16$$

De kwadratensommen staan in de Excel uitvoer voor beide modellen.  
De eenzijdige kritische waarde is  $F_{(0,05;1;3)} = 10,13$ .

De test op lineariteit faalt omdat de verschilvariantie in de teller significant groter is dan de variantie van het tweedegraadsmodel,  $F > F_{(0,05;1;3)}$ . Het tweedegraadsmodel is dus significant beter dan het eerstegraadsmodel.

Test4: Test op significantie van de  $b_2$ -coëfficiënt (paragraaf 4.7.4):

De regressiecoëfficiënt  $b_2$  is significant omdat:

- $p(b_2) = 5,47 \cdot 10^{-3}$ . Dit is kleiner dan 0,05;
- het 95%-betrouwbaarheidsinterval van  $b_2$  omvat niet het getal nul:  
 $BI(b_2) = [-6,84 \cdot 10^{-4}; -2,65 \cdot 10^{-4}]$ ;

Dus test op lineariteit faalt.

Er kan worden geconcludeerd dat het tweedegraads model zonder as-afsnede het beste model is voor de kalibratieset uit opgave 4.6.

#### Antwoord 4.8

Bij de antwoorden in het boek worden de gestandaardiseerde residuen en de Cook's afstanden voor alle meetpunten gegeven. Alleen de laatste waarden in beide reeksen zijn van toepassing op deze opgave.

Tip: Bij deze opgave kan de regressie eenvoudig worden uitgevoerd met behulp van de volgende Excel-functies:

$b_0$  =SNIJPUNT(y-blok;x-blok)

$b_1$  =RICHTING(y-blok;x-blok)

$s_r$  =STAND.FOUT.YX(y-blok;x-blok)

#### Berekening van de gestandaardiseerde residuen.

Eerst wordt een regressie uitgevoerd *met alle meetpunten*, inclusief de potentiële uitbijter. Het resultaat hiervan is:  $b_0 = -0,0290$ ;  $b_1 = 0,0326$ ;  $s_r = 4,046 \cdot 10^{-2}$

nr	$x_i$	$y_i$	$\hat{y}_i$	$r_i = (y_i - \hat{y}_i)$	$ r_i/s_r $
1	5	0,16	0,1340	0,0260	0,64
2	10	0,29	0,2970	-0,0070	0,17
3	15	0,44	0,4600	-0,0200	0,49
4	20	0,58	0,6230	-0,0430	1,06
5	25	0,83	0,7860	0,0440	1,09

De gestandaardiseerde residuwaarde voor het 5<sup>e</sup> meetpunt is 1,09. Dit is kleiner dan de meest gebruikte grenswaarde 3. Meetpunt 5 wordt met deze methode dus niet als uitbijter gedetecteerd.

#### Berekening van de Cook's afstand voor het 5<sup>e</sup> meetpunt.

De berekening van de Cook's afstand voor een verdacht meetpunt verloopt in drie stappen:

1. Regressie *met alle meetpunten, inclusief* de potentiële uitbijter, gevolgd door berekening van  $\hat{y}_{i1}$  voor alle meetpunten met de hierbij geschatte modelparameters. Het resultaat is:  $b_0 = -0,0290$ ;  $b_1 = 0,0326$ ;  $s_r = 4,046 \cdot 10^{-2}$  met  $\hat{y}_{i1}$  in onderstaande tabel.
2. Regressie *met alle meetpunten, exclusief* de potentiële uitbijter, gevolgd door berekening van  $\hat{y}_{i2}$  voor alle meetpunten met de hierbij geschatte modelparameters. Het resultaat is:  $b_0 = 0,0150$ ;  $b_1 = 0,0282$  met  $\hat{y}_{i2}$  in onderstaande tabel.  
N.B.  $s_r$  hoeft hierbij niet te worden berekend.
3. Berekening van de som van de kwadraten van de verschillen tussen  $\hat{y}_{i1}$  en  $\hat{y}_{i2}$  en invulling daarvan in formule (4.79) voor berekening van de Cook's afstand.

			Stap 1	Stap 2		
nr	$x_i$	$y_i$	$\hat{y}_{i1}$	$\hat{y}_{i2}$	$\hat{y}_{i1} - \hat{y}_{i2}$	$(\hat{y}_{i1} - \hat{y}_{i2})^2$
1	5	0,16	0,1340	0,1560	-2,200E-02	4,840E-04
2	10	0,29	0,2970	0,2970	0,000E+00	0,000E+00
3	15	0,44	0,4600	0,4380	2,200E-02	4,840E-04
4	20	0,58	0,6230	0,5790	4,400E-02	1,936E-03
5	25	0,83	0,7860	0,7200	6,600E-02	4,356E-03
					som	7,260E-03

Berekening van de Cook's afstand voor meetpunt 5 met (4.79):

$$CD_i = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{y}_j^{(-i)})^2}{ps_r^2} = \frac{7,260 \cdot 10^{-3}}{2 \cdot (4,046 \cdot 10^{-2})^2} = 2,20$$

#### Antwoord 4.9

Correctie antwoorden: b.  $\hat{x} = 1,73 \cdot 10^{-4} \pm 2,15 \cdot 10^{-5}$

i	$x_i$	$y_i$	$(x_i - \bar{x})^2$
1	0,00E+00	0,171	1,56E-08
2	5,00E-05	0,226	5,63E-09
3	1,00E-04	0,262	6,25E-10
4	1,50E-04	0,323	6,25E-10
5	2,00E-04	0,369	5,63E-09
6	2,50E-04	0,420	1,56E-08
som	7,50E-04	1,771	4,375E-08
gem	1,25E-04	0,2952	SS <sub>xx</sub>

Gegevens voor de regressie	
Meervoudige	
correlatiecoëfficiënt R	0,9988
R-kwadraat	0,9975
Aangepaste kleinste kwadraat	0,9969
Standaardfout	5,187E-03
Waarnemingen	6

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	1	4,300E-02	4,300E-02	1598,3496	2,339E-06
Storing	4	1,076E-04	2,690E-05		
Totaal	5	4,311E-02			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,1712	0,0038	45,6141	1,38E-06	0,1608	0,1817
Variabele X 1	991,43	24,7985	39,9794	2,34E-06	923	1060

Deze uitvoer bevat de resultaten voor  $s_r$ ,  $SS_{\text{Reg}}$ ,  $SS_{\text{Res}}$ ,  $SS_T$ ,  $b_0$ ,  $b_1$ ,  $s(b_0)$  en  $s(b_1)$  en de 95% betrouwbaarheidsintervallen van  $b_0$  en  $b_1$ . Hieruit kan de helft van het betrouwbaarheidsinterval (die bij de antwoorden staat vermeld) worden berekend door het verschil te nemen van de hoogste waarde van het 95% betrouwbaarheidsinterval en de waarde van de parameter. De resultaten zijn:  $b_0 = 0,171 \pm 0,010$ ;  $b_1 = 991 \pm 69$ .

Deze gegevens kunnen worden gebruikt voor de berekening van de concentratie uit de as-afsnede en het bijbehorende betrouwbaarheidsinterval.

Berekening van de concentratie uit de as-afsnede met (4.88) levert:

$$x_0 = b_0/b_1 = 0,1712/991,43 = 1,73 \cdot 10^{-4}$$

Berekening van het bijbehorende betrouwbaarheidsinterval met (4.89):

$$BI(\hat{x}_0) = \hat{x}_0 \pm \frac{t_{(0,05;n-2)} S_r}{b_1} \cdot \sqrt{\frac{1}{n} + \frac{\bar{y}^2}{b_1^2 \sum (\Delta x_i - \Delta \bar{x})^2}}$$

$$\begin{aligned}
 \text{BI}(\hat{x}_0) &= 1,73 \cdot 10^{-4} \pm \frac{2,776 \cdot 5,187 \cdot 10^{-3}}{991,43} \cdot \sqrt{\frac{1}{6} + \frac{0,2952^2}{991,43^2 \cdot 4,375 \cdot 10^{-8}}} \\
 &= 1,73 \cdot 10^{-4} \pm 2,15 \cdot 10^{-5}
 \end{aligned}$$

