

Uitwerkingen

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

Hoofdstuk 13

Antwoord 13.1

MLR is uitgevoerd op basis van een klassiek kalibratiemodel met niet-geschaalde variabelen in de X matrix en niet-geschaalde waarden in de y vector.

Lineaire regressie met het model $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9911
R-kwadraat	0,9824
Aangepaste kleinste kwadraat	0,9736
Standaardfout	2,4460
Waarnemingen	13

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F		Significantie F
Regressie	4	2667,90	666,97	111,5		4,76E-07
Storing	8	47,86	5,98			
Totaal	12	2715,76				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	62,4054	70,0710	0,8906	0,3991	-99,1786	223,9893
x1	1,5511	0,7448	2,0827	0,0708	-0,1663	3,2685
x2	0,5102	0,7238	0,7049	0,5009	-1,1589	2,1792
x3	0,1019	0,7547	0,1350	0,8959	-1,6385	1,8423
x4	-0,1441	0,7091	-0,2032	0,8441	-1,7791	1,4910

Alle regressiecoëfficiënten hebben een p -waarde groter dan 0,05 en zijn daarom niet significant.

Lineaire regressie met het model $y = \beta_0 + \beta_1x_1 + \beta_4x_4$

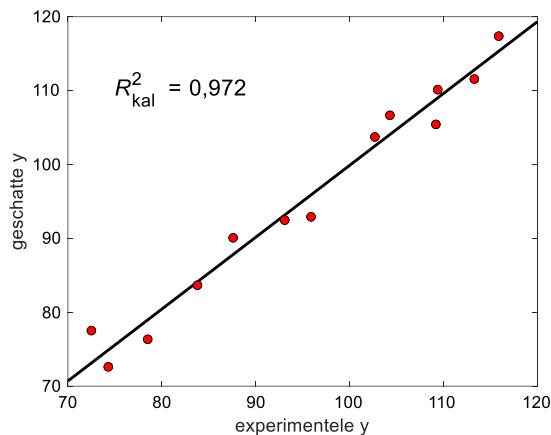
Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9861
R-kwadraat	0,9725
Aangepaste kleinste kwadraat	0,9670
Standaardfout	2,7343
Waarnemingen	13

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F		Significantie F
Regressie	2	2641,00	1320,50	176,627		1,58E-08
Storing	10	74,76	7,48			
Totaal	12	2715,76				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	103,0974	2,1240	48,5396	3,32E-13	98,3649	107,8299
x1	1,4400	0,1384	10,4031	1,11E-06	1,1315	1,7484
x4	-0,6140	0,0486	-12,6212	1,81E-07	-0,7223	-0,5056

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. De regressiecoëfficiënten β_0 , β_1 en β_4 hebben een p -waarde kleiner dan 0,05 en zijn daarom significant.

Uit voorgaande Exceluitvoer blijkt dat $R^2_{\text{kal}} = 0,9725$ en dat $RMSEC = s_r = 2,7343$.



Antwoord 13.2

MLR is uitgevoerd op basis van een klassiek kalibratiemodel met niet-geschaalde LSER-factoren in de X matrix en niet-geschaalde $\log k_w$ waarden in de y vector.

$$\text{Model } \log k_w = \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \beta_3 L_3 + \beta_4 L_4 + \beta_5 L_5$$

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9852
R-kwadraat	0,9706
Aangepaste kleinste kwadraat	0,9600
Standaardfout	0,3018
Waarnemingen	20

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	5	42,0345	8,4069	92,31	3,28E-10
Storing	14	1,2751	0,0911		
Totaal	19	43,3096			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,4983	0,3477	1,4333	0,1737	-0,2474	1,2440
L1	0,1246	0,2721	0,4578	0,6541	-0,4590	0,7081
L2	-0,8283	0,3065	-2,7024	0,0172	-1,4857	-0,1709
L3	-0,6354	0,2406	-2,6410	0,0194	-1,1514	-0,1194
L4	-2,3152	0,3066	-7,5519	0,0000	-2,9727	-1,6577
L5	3,3488	0,2486	13,4712	0,0000	2,8156	3,8820

Alleen de parameters $\beta_2, \beta_3, \beta_4$ en β_5 en hebben een p -waarde kleiner dan 0,05 en zijn daarom significant. Daarom wordt opnieuw een regressie uitgevoerd met een model zonder β_0 -term en zonder variabele L_1 : $\log k_w = \beta_2 L_2 + \beta_3 L_3 + \beta_4 L_4 + \beta_5 L_5$

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9950
R-kwadraat	0,9900
Aangepaste kleinste kwadraat	0,9256
Standaardfout	0,3043
Waarnemingen	20

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	4	146,7830	36,6958	396	5,19E-15
Storing	16	1,4814	0,0926		
Totaal	20	148,2644			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
		#N/B	#N/B	#N/B	#N/B	#N/B
Snijpunt	0	#N/B	#N/B	#N/B	#N/B	#N/B
L2	-0,5296	0,2090	-2,5345	2,21E-02	-0,9726	-0,0866
L3	-0,6121	0,2420	-2,5293	2,23E-02	-1,1251	-0,0991
L4	-2,4458	0,2949	-8,2925	3,47E-07	-3,0710	-1,8205
L5	3,6639	0,1332	27,5142	6,68E-15	3,3816	3,9462

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. De parameters $\beta_2, \beta_3, \beta_4$ en β_5 hebben een p -waarde kleiner dan 0,05 en zijn daarom significant.

Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 0,9900$ en dat $RMSEC = s_r = 0,3043$.

Met de regressiecoëfficiënten van dit model kunnen geschatte $\log k_w (\hat{y}_i)$, en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset worden berekend, zie volgende tabel.

	Log k_w = y_i	L2	L3	L4	L5	\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$	
Kalibratieset									
S1	n-Hexylbenzeen	5,4892	0,50	0,00	0,15	1,562	5,0913	-0,3979	0,158335
S2	1,3,5-Triisopropylbenzeen	6,0714	0,40	0,00	0,22	1,985	6,5229	0,4515	0,203816
S3	1,4-Dinitrobenzeen	1,5692	1,63	0,00	0,41	1,065	2,0360	0,4668	0,217885
S4	3-Trifluormethylfenol	2,5650	0,87	0,72	0,09	0,969	2,4287	-0,1363	0,018578
S5	4-Cyanofenol	1,1786	1,63	0,79	0,29	0,930	1,3513	0,1727	0,029826
S6	4-Joodfenol	2,3819	1,22	0,68	0,20	1,033	2,2333	-0,1486	0,022091
S7	Benzeen	2,0052	0,52	0,00	0,14	0,716	2,0055	0,0003	1,03E-07
S8	Chloorbenzeen	2,6725	0,65	0,00	0,07	0,839	2,5585	-0,1140	0,012989
S9	Cyclohexanon	1,0396	0,86	0,00	0,56	0,861	1,3295	0,2899	0,084037
S10	Fenol	1,0938	0,89	0,60	0,30	0,775	1,2672	0,1734	0,030053
S11	Hexachloorbutadien	4,5311	0,85	0,00	0,00	1,321	4,3898	-0,1413	0,019967
S12	Indazool	1,6175	1,25	0,54	0,34	0,905	1,4917	-0,1258	0,015827
S13	Cafeïne	0,8923	1,60	0,00	1,35	1,363	0,8447	-0,0476	0,002267
S14	4-Nitrobenzoëzuur	1,5758	1,07	0,62	0,54	1,106	1,7853	0,2095	0,043907
S15	N-Methyl-2-pyrrolidinon	0,3015	1,50	0,00	0,95	0,820	-0,1135	-0,4150	0,172248
S16	Naftaleen	3,0787	0,92	0,00	0,20	1,085	2,9989	-0,0798	0,006369
S17	4-Chloorfenol	1,9694	1,08	0,67	0,20	0,898	1,8189	-0,1505	0,022645
S18	Tolueen	2,6125	0,52	0,00	0,14	0,716	2,0055	-0,6070	0,368423
S19	Benzonitril	1,5680	1,11	0,00	0,33	0,871	1,7963	0,2283	0,052099
S20	Benzoëzuur	1,6027	0,90	0,59	0,40	0,932	1,5986	-0,0041	1,65E-05
Testset									
T1	3,5-Dichloorfenol	2,9099	1,10	0,83	0,00	1,020	2,6465	-0,2634	0,069362
T2	Methylfenylether	2,0436	0,75	0,00	0,29	0,916	2,2496	0,2060	0,042444
T3	Benzamide	0,8308	1,50	0,49	0,67	0,973	0,8319	0,0011	1,29E-06
T4	Dibenzothiofeen	4,0185	1,31	0,00	0,18	1,379	3,9184	-0,1001	0,010012
T5	1,3-Diisopropylbenzeen	4,8854	0,46	0,00	0,20	1,562	4,9902	0,1048	0,010980

De geschatte $\log k_w$ van de testmonsters staan vermeld in voorgaande tabel: 2,6465; 2,2496; 0,8319; 3,9184; 4,9902.

De correlatiecoëfficiënten voor de geschatte $\log k_w$ en de juiste $\log k_w$ kunnen afzonderlijk worden berekend voor de kalibratieset (S) en testset (T). De kwadraten van deze correlatiecoëfficiënten zijn: $R_{\text{kal}}^2 = 0,9663$ en $R_{\text{test}}^2 = 0,9871$.

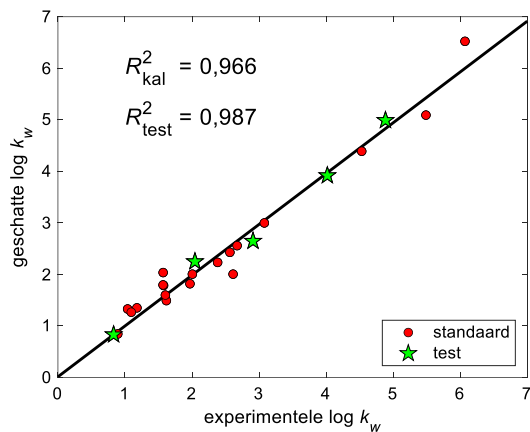
N.B. De $R_{\text{kal}}^2 = 0,9663$ die hierbij wordt berekend verschilt van $R_{\text{kal}}^2 = 0,9900$ die bij de lineaire regressie voor het model $\log k_w = \beta_2 L_2 + \beta_3 L_3 + \beta_4 L_4 + \beta_5 L_5$ wordt gevonden omdat het een andere regressie betreft.

Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^5 (\hat{y}_i - y_i)^2 = 0,132798$$

RMSEP kan worden berekend met (12.15):

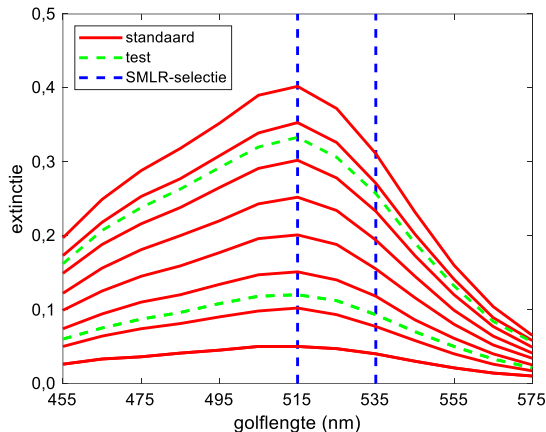
$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{0,132798}{5}} = 0,1630$$



Antwoord 13.3

SMLR is uitgevoerd op basis van een invers kalibratiemodel niet-geschaalde variabelen in de X matrix en met niet-geschaalde concentraties in de y vector.

De spectra van de standaarden en testmonsters van de kobaltoplossingen en de bij SMLR geselecteerde golflengten zijn afgebeeld in de volgende figuur.



Lineaire regressie voor model met β_0 -term

Door toepassing van *stapsgewijze meervoudige lineaire regressie* (SMLR) worden de golflengten bij 515 en 535 nm geselecteerd. De regressie is uitgevoerd met niet-gecentreerde data. De x - en y -waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel. Ook zijn hierin de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend.

Nr.	x1	x2	Co-concentratie y	\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
	515	535				
Kalibratieset						
1	0,050	0,040	0,010	0,0100	-2,844E-05	8,087E-10
2	0,102	0,077	0,020	0,0200	2,742E-05	7,516E-10
3	0,151	0,118	0,030	0,0301	8,486E-05	7,201E-09
4	0,201	0,155	0,040	0,0399	-1,111E-04	1,234E-08
5	0,252	0,194	0,050	0,0500	8,503E-06	7,229E-11
6	0,302	0,233	0,060	0,0600	2,198E-06	4,832E-12
7	0,353	0,271	0,070	0,0700	2,697E-05	7,275E-10
8	0,402	0,311	0,080	0,0800	-1,041E-05	1,084E-10
Testset						
1	0,120	0,093	0,024	0,0238	-1,890E-04	3,572E-08
2	0,333	0,257	0,066	0,0662	1,812E-04	3,284E-08

Het inverse lineaire kalibratiemodel is: $y = \beta_0 + \beta_{515} \cdot x_{515} + \beta_{535} \cdot x_{535}$

Lineaire regressie met Excel levert voor de geselecteerde golflengten met niet-gecentreerde data:

Gegevens voor de regressie	
Meervoudige	
correlatiecoëfficiënt R	1,0000
R-kwadraat	1,0000
Aangepaste kleinste kwadraat	1,0000
Standaardfout	6,64E-05
Waarnemingen	8

Variantieanalyse

	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	2	4,20E-03	2,10E-03	476878	6,29E-14
Storing	5	2,20E-08	4,40E-09		
Totaal	7	4,20E-03			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	-1,17E-04	5,26E-05	-2,2252	7,66E-02	-0,00025	1,82E-05
515	0,1259	0,0165	7,6147	6,21E-04	0,0834	0,1684
535	0,0948	0,0215	4,4151	6,92E-03	0,0396	0,1500

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. Alleen regressiecoëfficiënten β_{515} en β_{535} hebben een p -waarde kleiner dan 0,05 en zijn daarom significant.

Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 1,0000$ en dat $RMSEC = s_r = 6,64 \cdot 10^{-5}$. De geschatte kobaltconcentraties in de testmonsters zijn 0,0238 en 0,0662, zie twee tabellen terug.

Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^2 (\hat{y}_i - y_i)^2 = 6,856 \cdot 10^{-8}$$

RMSEP kan worden berekend met (12.15):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{6,856 \cdot 10^{-8}}{2}} = 1,85 \cdot 10^{-4}$$

De antwoorden in het boek corresponderen met dit regressiemodel.

Lineaire regressie voor model zonder β_0 -term

Voor de lineaire regressie van het model met β_0 -term, zoals hiervoor is beschreven, blijkt uit de Excel-uitvoer dat de β_0 -term niet significant is. Daarom is een nieuwe regressie uitgevoerd voor een model zonder β_0 -term, met de gemeten extincties bij de golflengten die door de SMLR-methode zijn geselecteerd, 515 en 535 nm.

Het inverse lineaire kalibratiemodel zonder β_0 -term is: $y = \beta_{515} \cdot x_{515} + \beta_{535} \cdot x_{535}$

De x - en y -waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel. Ook zijn hierin de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De regressie is uitgevoerd met niet-gecentreerde data.

Nr.	x1 515	x2 535	Co-concentratie y	\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
Kalibratieset						
1	0,050	0,040	0,010	0,0101	5,619E-05	3,157E-09
2	0,102	0,077	0,020	0,0201	1,163E-04	1,352E-08
3	0,151	0,118	0,030	0,0301	1,272E-04	1,619E-08
4	0,201	0,155	0,040	0,0399	-7,637E-05	5,833E-09
5	0,252	0,194	0,050	0,0500	2,507E-05	6,284E-10
6	0,302	0,233	0,060	0,0600	-5,338E-06	2,849E-11
7	0,353	0,271	0,070	0,0700	9,508E-06	9,039E-11
8	0,402	0,311	0,080	0,0799	-6,615E-05	4,376E-09
Testset						
1	0,120	0,093	0,024	0,0239	-1,249E-04	1,561E-08
2	0,333	0,257	0,066	0,0662	1,602E-04	2,567E-08

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	1,0000
R-kwadraat	1,0000
Aangepaste kleinste kwadraat	0,8333
Standaardfout	8,55E-05
Waarnemingen	8

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	2	0,0204	0,0102	1396559	4,29E-15	
Storing	6	4,38E-08	7,30E-09			
Totaal	8	0,0204				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0	#N/B	#N/B	#N/B	#N/B	#N/B
515	0,1318	0,0210	6,2740	0,0008	0,0804	0,1833
535	0,0866	0,0272	3,1782	0,0191	0,0199	0,1533

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. De regressiecoëfficiënten β_{515} en β_{535} hebben een p -waarde kleiner dan 0,05 en zijn daarom significant.

Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 1,0000$ en dat $RMSEC = s_r = 8,55 \cdot 10^{-5}$. De geschatte kobaltconcentraties in de testmonsters zijn 0,0239 en 0,0662, zie twee tabellen terug.

Voor de kalibratieset kan de correlatiecoëfficiënt worden berekend voor de geschatte \hat{y} en de juiste y . Het kwadraat van deze correlatiecoëfficiënt is: $R_{\text{kal}}^2 = 1,000$.

Voor de testset wordt voor de geschatte \hat{y} en de juiste y geen correlatiecoëfficiënt berekend omdat er slechts twee testmonsters zijn. Door twee meetpunten gaat één rechte lijn waarvoor altijd een $R_{\text{test}}^2 = 1,000$ zal worden gevonden.

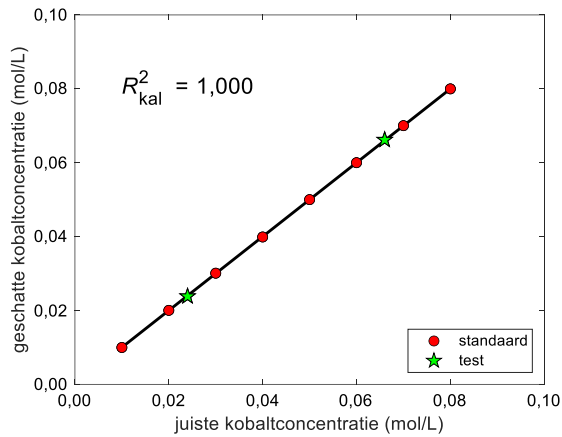
Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^2 (\hat{y}_i - y_i)^2 = 4,128 \cdot 10^{-8}$$

RMSEP kan worden berekend met (12.15):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{4,128 \cdot 10^{-8}}{2}} = 1,44 \cdot 10^{-4}$$

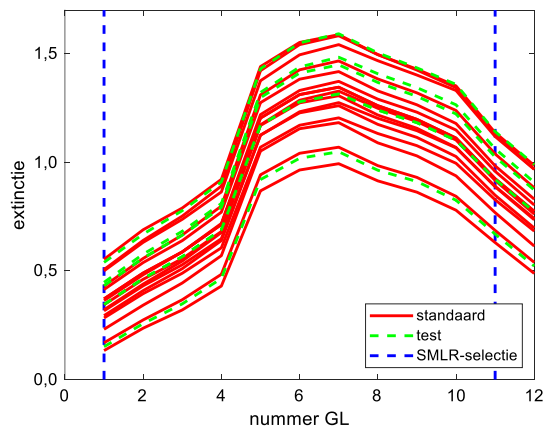
RMSEP is iets kleiner dan voor het model met β_0 -term.



Antwoord 13.4

SMLR is uitgevoerd op basis van een invers kalibratiemodel met niet-geschaalde variabelen in de X matrix en niet-geschaalde vetpercentages in de y vector.

De NIR spectra van de standaarden en testmonsters van kaas zijn afgebeeld in de volgende figuur.



De correlaties tussen de golflengten in de X -matrix van de kalibratieset zijn erg hoog zoals blijkt uit de volgende correlatiematrix. Er is dus sprake van een hoge mate van multicollineariteit in de kalibratiematrix.

	GL1	GL2	GL3	GL4	GL5	GL6	GL7	GL8	GL9	GL10	GL11	GL12
GL1	1											
GL2	0,9999	1										
GL3	0,9991	0,9993	1									
GL4	0,9975	0,9979	0,9995	1								
GL5	0,9935	0,9941	0,9971	0,9990	1							
GL6	0,9923	0,9930	0,9965	0,9985	0,9998	1						
GL7	0,9916	0,9923	0,9960	0,9979	0,9992	0,9997	1					
GL8	0,9914	0,9919	0,9956	0,9968	0,9973	0,9982	0,9993	1				
GL9	0,9915	0,9919	0,9955	0,9966	0,9968	0,9977	0,9989	0,9999	1			
GL10	0,9916	0,9921	0,9953	0,9957	0,9951	0,9962	0,9977	0,9995	0,9997	1		
GL11	0,9916	0,9919	0,9949	0,9951	0,9941	0,9952	0,9969	0,9990	0,9994	0,9999	1	
GL12	0,9927	0,9929	0,9953	0,9948	0,9928	0,9937	0,9955	0,9981	0,9986	0,9995	0,9997	1

Door toepassing van *stapsgewijze meervoudige lineaire regressie* (SMLR) worden de golflengten nummer 1 en 11 geselecteerd. De x - en y -waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel. Ook zijn hierin de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De regressie wordt uitgevoerd met niet-gecentreerde data.

nr.	x1	x11	vetgehalte y	\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
Kalibratieset						
1	0,371	0,924	41,7	41,6743	-0,025659	6,58E-04
2	0,133	0,630	41,5	41,4729	-0,027140	7,37E-04
3	0,553	1,136	41,7	41,6486	-0,051381	2,64E-03
4	0,231	0,768	41,7	41,7934	0,093352	8,71E-03
5	0,430	1,038	42,2	42,3008	0,100793	1,02E-02
6	0,294	0,848	41,9	41,8772	-0,022798	5,20E-04
7	0,169	0,686	41,7	41,6650	-0,035012	1,23E-03
8	0,500	1,118	42,2	42,2693	0,069329	4,81E-03
9	0,414	0,983	41,7	41,7932	0,093219	8,69E-03
10	0,283	0,827	41,8	41,7640	-0,036030	1,30E-03
11	0,344	0,887	41,6	41,6003	0,000349	1,22E-07
12	0,316	0,889	42,0	42,0896	0,089644	8,04E-03
13	0,506	1,126	42,4	42,2827	-0,117344	1,38E-02
14	0,363	0,958	42,3	42,2828	-0,017163	2,95E-04
15	0,321	0,916	42,5	42,3858	-0,114160	1,30E-02
Testset						
1	0,345	0,921	41,9	42,0606	0,160585	2,58E-02
2	0,432	1,024	42,0	42,0716	0,071554	5,12E-03
3	0,447	1,062	42,3	42,3572	0,057247	3,28E-03
4	0,151	0,669	41,6	41,7232	0,123153	1,52E-02
5	0,539	1,141	42,1	41,9493	-0,150650	2,27E-02

Het inverse lineaire kalibratiemodel is: $y = \beta_0 + \beta_1 \cdot x_1 + \beta_{11} \cdot x_{11}$

Lineaire regressie met Excel levert voor de geselecteerde golflengten met niet-gecentreerde data:

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9732
R-kwadraat	0,9471
Aangepaste kleinste kwadraat	0,9383
Standaardfout	0,0788
Waarnemingen	15

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	2	1,3348	0,6674	107,38	2,2E-08	
Storing	12	0,0746	0,0062			
Totaal	14	1,4093				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	34,8307	0,5137	67,8068	7,03E-17	33,7115	35,9499
GL1	-16,4733	1,3503	-12,1994	4,02E-08	-19,4154	-13,5312
GL11	14,0208	1,0663	13,1491	1,73E-08	11,6976	16,3441

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. Alle regressiecoëfficiënten in dit model hebben een p -waarde kleiner dan 0,05 en zijn daarom significant.

Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 0,9471$ en dat $RMSEC = s_r = 0,0788$

De geschatte vetpercentages in de testmonsters zijn hiervoor vermeld, zie twee tabellen terug.

De correlatiecoëfficiënten voor het geschatte vetgehalte en het juiste vetgehalte kunnen afzonderlijk worden berekend voor de kalibratieset en testset. De kwadraten van deze correlatiecoëfficiënten zijn: $R_{\text{kal}}^2 = 0,9471$ en $R_{\text{test}}^2 = 0,7824$.

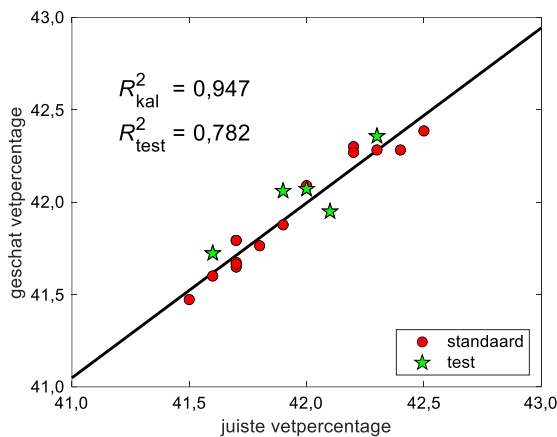
De correlatiecoëfficiënt is afhankelijk van het concentratiebereik. R_{test}^2 is in dit geval relatief laag mede omdat het concentratiebereik erg klein is: $42,3\% - 41,6\% = 0,7\%$. Bij dezelfde residuen en een groter concentratiebereik zou R_{test}^2 hoger zijn.

Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^5 (\hat{y}_i - y_i)^2 = 7,205 \cdot 10^{-2}$$

RMSEP kan worden berekend met (12.15):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{7,205 \cdot 10^{-2}}{5}} = 1,20 \cdot 10^{-1}$$



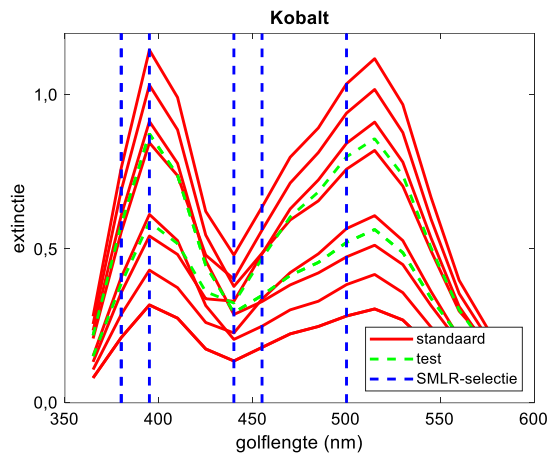
Antwoord 13.5

SMLR is uitgevoerd op basis van een invers kalibratiemodel met niet-geschaalde variabelen in de X matrix en niet-geschaalde concentraties in de y vector.

Kobalt

Door toepassing van *stapsgewijze meervoudige lineaire regressie* (SMLR) worden de golflengten 380, 395, 440, 455 en 500 nm geselecteerd. Deze zijn aangegeven in de volgende afbeelding. De x - en y -waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel.

nr.	Co-concentratie					y
	x1	x2	x3	x4	x5	
	380	395	440	455	500	
Kalibratieset						
1	0,212	0,318	0,136	0,180	0,282	0,060
2	0,408	0,612	0,206	0,249	0,384	0,080
3	0,609	0,914	0,287	0,329	0,474	0,100
4	0,287	0,431	0,227	0,337	0,566	0,120
5	0,690	1,032	0,378	0,486	0,759	0,160
6	0,358	0,542	0,330	0,484	0,843	0,180
7	0,555	0,845	0,402	0,561	0,941	0,200
8	0,763	1,146	0,480	0,636	1,036	0,220
Testset						
1	0,579	0,874	0,293	0,349	0,521	0,110
2	0,387	0,587	0,324	0,469	0,799	0,170



Het inverse lineaire kalibratiemodel is:

$$y = \beta_0 + \beta_{380} \cdot x_{380} + \beta_{395} \cdot x_{395} + \beta_{440} \cdot x_{440} + \beta_{455} \cdot x_{455} + \beta_{500} \cdot x_{500}$$

De regressie wordt uitgevoerd met niet-gecentreerde data.

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	1,0000
R-kwadraat	1,0000
Aangepaste kleinste kwadraat	1,0000
Standaardfout	2,66E-05
Waarnemingen	8

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	5	2,40E-02	4,80E-03	6787051	1,47E-07	
Storing	2	1,41E-09	7,07E-10			
Totaal	7	2,40E-02				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	-0,0011	3,51E-05	-31,1197	1,03E-03	-0,0012	-0,0009
380	-0,0190	4,44E-03	-4,2792	5,05E-02	-0,0381	0,0001
395	-0,0148	3,06E-03	-4,8212	4,04E-02	-0,0279	-0,0016
440	0,1256	3,04E-03	41,3256	5,85E-04	0,1125	0,1387
455	0,0575	3,44E-03	16,7181	3,56E-03	0,0427	0,0723
500	0,1502	1,86E-03	80,8748	1,53E-04	0,1422	0,1582

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop 'Coëfficiënten' en de significantie ervan onder de kop 'P-waarde'. Uit de Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 1,0000$ en dat $RMSEC = s_r = 2,66 \cdot 10^{-5}$.

Op basis van deze regressie zijn de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De resultaten staan in de volgende tabel. De geschatte kobaltconcentraties in de testmonsters zijn in deze tabel vermeld.

nr.	\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
Kalibratieset			
1	0,059980	-2,038E-05	4,154E-10
2	0,079999	-1,262E-06	1,592E-12
3	0,100017	1,665E-05	2,773E-10
4	0,120005	5,365E-06	2,878E-11
5	0,160005	4,807E-06	2,311E-11
6	0,180018	1,770E-05	3,133E-10
7	0,199995	-4,615E-06	2,130E-11
8	0,219982	-1,827E-05	3,337E-10
Testset			
1	0,110141	1,406E-04	1,977E-08
2	0,170577	5,771E-04	3,330E-07

De correlatiecoëfficiënt voor de geschatte en juiste kobaltconcentraties kan worden berekend voor de kalibratieset. Het kwadraat van deze correlatiecoëfficiënt is: $R_{\text{kal}}^2 = 1,0000$.

Voor de testset wordt voor de geschatte \hat{y} en de juiste y geen correlatiecoëfficiënt berekend omdat er slechts twee testmonsters zijn.

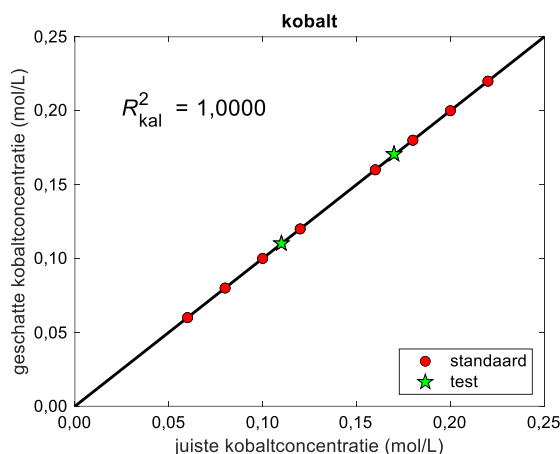
Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^2 (\hat{y}_i - y_i)^2 = 3,528 \cdot 10^{-7}$$

RMSEP kan worden berekend met (12.15):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{3,528 \cdot 10^{-7}}{2}} = 4,20 \cdot 10^{-4}$$

Deze RMSEP-waarde voor kobalt is kleiner dan die voor het klassieke regressiemodel met twee golflengten (395 en 515 nm), waarvoor $\text{RMSEP} = 6,38 \cdot 10^{-4}$, zie bladzijde 303.

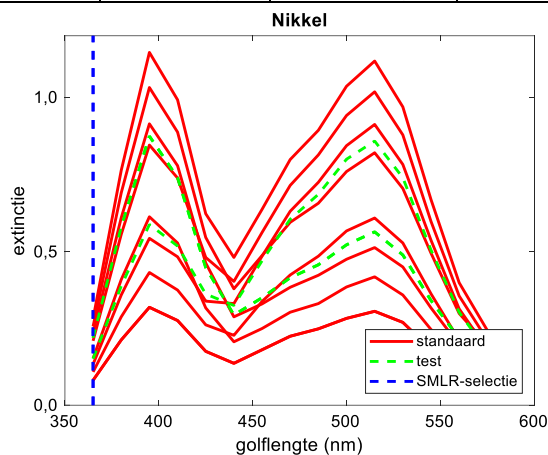


Nikkel

Door toepassing van *stapsgewijze meervoudige lineaire regressie* (SMLR) wordt alleen de golflengte 365 nm geselecteerd. Deze is aangegeven in de volgende afbeelding. De x - en y -

waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel. Ook zijn hierin de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De regressie is uitgevoerd met niet-gecentreerde data.

Nr.	x (365 nm)	Ni-concentratie y		\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
Kalibratieset						
1	0,081	0,060		0,0626	0,00262	6,87E-06
2	0,152	0,120		0,1175	-0,00249	6,20E-06
3	0,235	0,180		0,1817	0,00168	2,81E-06
4	0,108	0,080		0,0835	0,00349	1,22E-05
5	0,256	0,200		0,1979	-0,00209	4,36E-06
6	0,133	0,100		0,1028	0,00282	7,96E-06
7	0,209	0,160		0,1616	0,00158	2,49E-06
8	0,281	0,220		0,2172	-0,00276	7,62E-06
Testset						
1	0,219	0,170		0,1693	-0,00069	4,79E-07
2	0,151	0,110		0,1167	0,00674	4,54E-05



Eerst is een regressie uitgevoerd met het inverse lineaire kalibratiemodel: $y = \beta_0 + \beta_{365} \cdot x_{365}$. Daarbij bleek dat de β_0 -term niet significant was. Daarom is een nieuwe regressie uitgevoerd met niet-gecentreerde data met het volgende inverse lineaire kalibratiemodel zonder β_0 -term: $y = \beta_{365} \cdot x_{365}$

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9999
R-kwadraat	0,9997
Aangepaste kleinste kwadraat	0,8569
Standaardfout	2,686E-03
Waarnemingen	8

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	1	1,807E-01	1,807E-01	25047	4,29E-12
Storing	7	5,052E-05	7,216E-06		
Totaal	8	1,808E-01			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0	#N/B	#N/B	#N/B	#N/B	#N/B
365	0,7731	0,0049	158,2622	1,06E-13	0,7615	0,7846

De regressiecoëfficiënt is af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 0,9997$ en dat $RMSEC = s_r = 2,69 \cdot 10^{-3}$.

Op basis van deze regressie zijn de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De resultaten staan in de volgende tabel. De geschatte nikkelconcentraties in de testmonsters zijn in deze tabel vermeld.

De correlatiecoëfficiënt voor de geschatte en juiste nikkelconcentraties kan worden berekend voor de kalibratieset. Het kwadraat van deze correlatiecoëfficiënt is: $R_{\text{kal}}^2 = 0,9989$.

N.B. De $R_{\text{kal}}^2 = 0,9989$ die hierbij wordt berekend verschilt van $R_{\text{kal}}^2 = 0,9997$ die bij de lineaire regressie voor het model $y = \beta_{365} \cdot x_{365}$ wordt gevonden omdat het een andere regressie betreft.

Voor de testset wordt voor de geschatte \hat{y} en de juiste y geen correlatiecoëfficiënt berekend omdat er slechts twee testmonsters zijn.

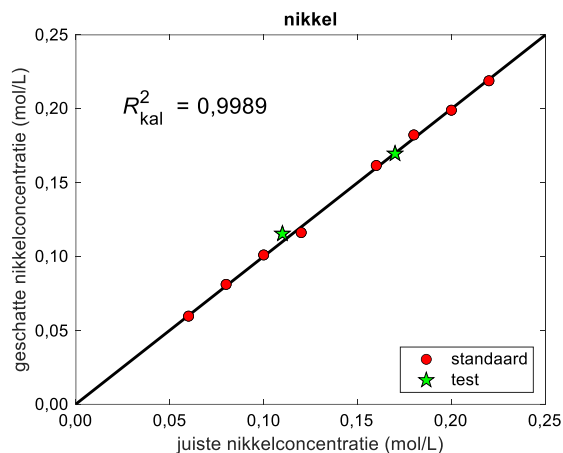
Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^2 (\hat{y}_i - y_i)^2 = 4,587 \cdot 10^{-5}$$

RMSEP kan worden berekend met (12.15):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{4,587 \cdot 10^{-5}}{2}} = 4,79 \cdot 10^{-3}$$

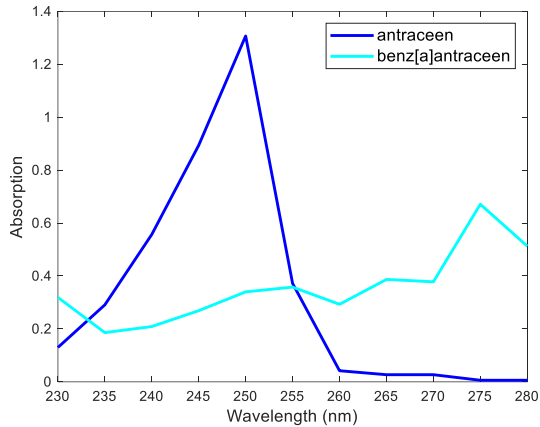
Deze RMSEP-waarde voor nikkel is groter dan die voor het klassieke regressiemodel met twee golflengten (395 en 515 nm), waarvoor $RMSEP = 9,15 \cdot 10^{-4}$, zie bladzijde 303.



Antwoord 13.6

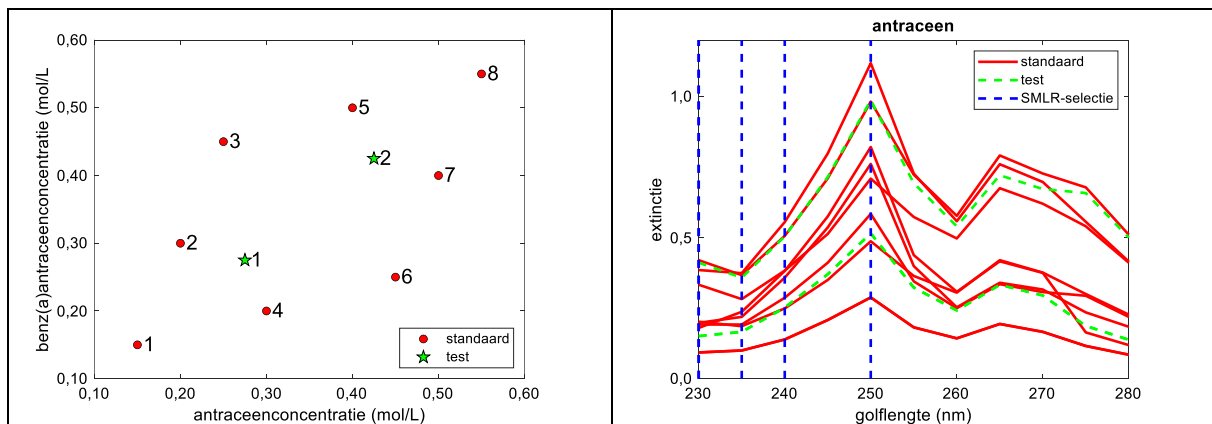
SMLR is uitgevoerd op basis van een invers kalibratiemodel met niet-geschaalde variabelen in de X matrix en niet-geschaalde concentraties in de y vector.

De spectra van de zuivere componenten zijn gegeven in de volgende afbeelding.



Antraceen

Door toepassing van *stapsgewijze meervoudige lineaire regressie (SMLR)* worden de golflengten 230, 235, 240, en 250 nm geselecteerd. Deze zijn rechts in de volgende afbeelding aangegeven. Links is het experimenteel design afgebeeld. De x - en y -waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel.



nr.	x1	x2	x3	x4	Antraceen-concentratie y
	230	235	240	250	
Kalibratieset					
1	0,093	0,100	0,139	0,288	0,150
2	0,202	0,186	0,249	0,487	0,200
3	0,333	0,282	0,383	0,709	0,250
4	0,192	0,192	0,287	0,584	0,300
5	0,420	0,367	0,506	0,982	0,400
6	0,198	0,219	0,358	0,761	0,450
7	0,180	0,237	0,382	0,821	0,500
8	0,385	0,374	0,555	1,118	0,550
Testset					
1	0,151	0,166	0,252	0,515	0,275
2	0,411	0,358	0,504	0,987	0,425

Het inverse lineaire kalibratiemodel is:

$$y = \beta_0 + \beta_{230} \cdot x_{230} + \beta_{235} \cdot x_{235} + \beta_{240} \cdot x_{240} + \beta_{250} \cdot x_{250}$$

De regressie is uitgevoerd met niet-gecentreerde data.

Gegevens voor de regressie	
Meervoudige	
correlatiecoëfficiënt R	1,0000
R-kwadraat	1,0000
Aangepaste kleinste kwadraat	0,9999
Standaardfout	1,563E-03
Waarnemingen	8

Variantieanalyse					
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F
Regressie	4	1,500E-01	3,750E-02	15346	8,54E-07
Storing	3	7,331E-06	2,444E-06		
Totaal	7	1,500E-01			

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0,0080	0,0021	3,8554	3,08E-02	0,0014	0,0146
230	-0,3698	0,0661	-5,5927	1,13E-02	-0,5802	-0,1594
235	-1,0590	0,1376	-7,6970	4,56E-03	-1,4969	-0,6212
240	-0,6664	0,1875	-3,5539	3,80E-02	-1,2631	-0,0697
250	1,2964	0,0704	18,4121	3,50E-04	1,0723	1,5205

De regressiecoëfficiënten zijn af te lezen in de Excel-uitvoer onder de kop ‘Coëfficiënten’ en de significantie ervan onder de kop ‘P-waarde’. Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 1,0000$ en dat $RMSEC = s_r = 1,563 \cdot 10^{-3}$

Op basis van deze regressie zijn de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De resultaten staan in de volgende tabel.

nr.	\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
Kalibratieset			
1	0,1485	-1,545E-03	2,386E-06
2	0,2018	1,754E-03	3,075E-06
3	0,2502	1,503E-04	2,258E-08
4	0,2995	-4,753E-04	2,259E-07
5	0,3999	-8,713E-05	7,591E-09
6	0,4509	8,607E-04	7,408E-07
7	0,5002	2,443E-04	5,967E-08
8	0,5491	-9,017E-04	8,130E-07
Testset			
1	0,2761	1,092E-03	1,193E-06
2	0,4206	-4,413E-03	1,948E-05

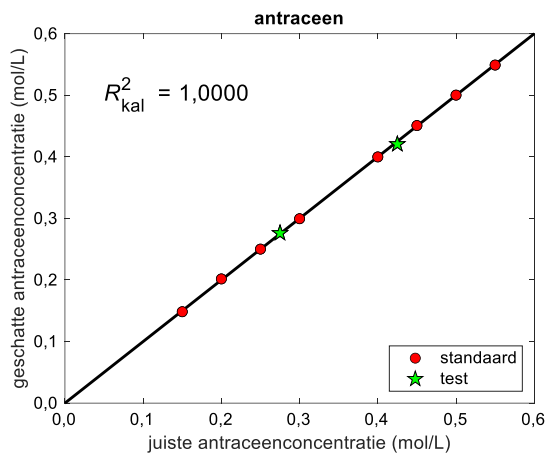
De correlatiecoëfficiënt voor de geschatte en juiste antraceenconcentraties kan worden berekend voor de kalibratieset. Het kwadraat van deze correlatiecoëfficiënt is: $R_{\text{kal}}^2 = 1,0000$. Voor de testset wordt voor de geschatte \hat{y} en de juiste y geen correlatiecoëfficiënt berekend omdat er slechts twee testmonsters zijn.

Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^2 (\hat{y}_i - y_i)^2 = 2,067 \cdot 10^{-5}$$

RMSEP kan worden berekend met (12.15):

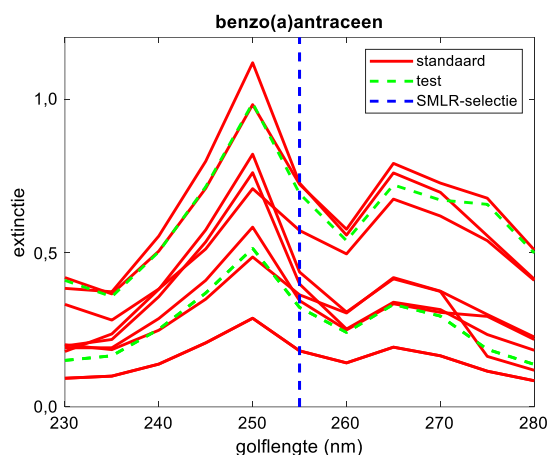
$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{2,067 \cdot 10^{-5}}{2}} = 3,21 \cdot 10^{-3}$$



Benzo(a)antraceen

Door toepassing van *stapsgewijze meervoudige lineaire regressie* (SMLR) wordt alleen de golflengte 255 nm geselecteerd. Deze is aangegeven in de volgende afbeelding. De x - en y -waarden van het inverse lineaire regressiemodel voor de kalibratieset en de testset zijn gegeven in de volgende tabel. Ook zijn hierin de geschatte y -waarden (\hat{y}_i), en de bijbehorende residuen en kwadraten van de residuen voor de kalibratieset en de testset berekend. De regressie is uitgevoerd met niet-gecentreerde data.

Nr.	x (255 nm)	Benzo(a)antraceen -concentratie y		\hat{y}_i	$r_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
Kalibratieset						
1	0,182	0,150		0,1354	-1,457E-02	2,122E-04
2	0,364	0,300		0,2709	-2,913E-02	8,487E-04
3	0,573	0,450		0,4264	-2,361E-02	5,573E-04
4	0,345	0,200		0,2567	5,673E-02	3,218E-03
5	0,723	0,500		0,5380	3,801E-02	1,445E-03
6	0,399	0,250		0,2969	4,691E-02	2,201E-03
7	0,438	0,400		0,3259	-7,407E-02	5,486E-03
8	0,727	0,550		0,5410	-9,010E-03	8,117E-05
Testset						
1	0,323	0,275		0,2404	-3,464E-02	1,200E-03
2	0,693	0,425		0,5157	9,069E-02	8,225E-03



Eerst is een regressie uitgevoerd met het inverse lineaire kalibratiemodel:

$$y = \beta_0 + \beta_{255} \cdot x_{255}$$

Daarbij bleek dat de β_0 -term niet significant was. Daarom is een nieuwe regressie uitgevoerd met het volgende inverse lineaire kalibratiemodel zonder β_0 -term. :

$$y = \beta_{255} \cdot x_{255}$$

Gegevens voor de regressie	
Meervoudige correlatiecoëfficiënt R	0,9938
R-kwadraat	0,9876
Aangepaste kleinste kwadraat	0,8447
Standaardfout	0,0448
Waarnemingen	8

Variantieanalyse						
	Vrijheidsgraden	Kwadratensom	Gemiddelde kwadraten	F	Significantie F	
Regressie	1	1,1160	1,1160		556	3,82E-07
Storing	7	0,0140	0,0020			
Totaal	8	1,1300				

	Coëfficiënten	Standaardfout	T-statistische gegevens	P-waarde	Laagste 95%	Hoogste 95%
Snijpunt	0	#N/B	#N/B	#N/B	#N/B	#N/B
255	0,7441	0,0316	23,5802	6,27E-8	0,6695	0,8188

De regressiecoëfficiënt is af te lezen in de Excel-uitvoer onder de kop 'Coëfficiënten' en de significantie ervan onder de kop 'P-waarde'. Uit voorgaande Exceluitvoer blijkt dat $R_{\text{kal}}^2 = 0,9876$ en dat $RMSEC = s_r = 0,0448$

De correlatiecoëfficiënt voor de geschatte en juiste benzo(a)antraceenconcentraties kan worden berekend voor de kalibratieset. Het kwadraat van deze correlatiecoëfficiënt is: $R_{\text{kal}}^2 = 0,9068$.

N.B. De $R_{\text{kal}}^2 = 0,9068$ die hierbij wordt berekend verschilt van $R_{\text{kal}}^2 = 0,9876$ die bij de lineaire regressie voor het model $y = \beta_{255} \cdot x_{255}$ wordt gevonden omdat het een andere regressie betreft.

Voor de testset is de som van de kwadraten van de residuen:

$$\sum_{i=1}^2 (\hat{y}_i - y_i)^2 = 9,425 \cdot 10^{-3}$$

Voor de testset wordt voor de geschatte \hat{y} en de juiste y geen correlatiecoëfficiënt berekend omdat er slechts twee testmonsters zijn.

RMSEP kan worden berekend met (12.15):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{test}}} (\hat{y}_i - y_i)^2}{n_{\text{test}}}} = \sqrt{\frac{9,425 \cdot 10^{-3}}{2}} = 6,86 \cdot 10^{-2}$$

