
Statistiek in het laboratorium met Excel

Dr. J.W.A. Klaessens

Derde druk

Syntax Media – Utrecht



© 2016 Uitgeverij Syntax Media, Utrecht

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voorzover het maken van kopieën uit deze uitgave is toegestaan op grond van artikel 16b Auteurswet 1912 jo het Besluit van 20 juni 1974, Stb. 351, zoals gewijzigd bij Besluit van 23 augustus 1985, Stb. 471 en artikel 17 Auteurswet 1912, dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 882, 1180 AW Amstelveen). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) dient men zich tot de uitgever te wenden.

ISBN: 978 94 91764 14 1

www.syntaxmedia.nl

Ontwerp omslag: Lapis Vivus grafisch ontwerp, Oosterbeek

Illustratieverantwoording

Alle afbeeldingen: J.W.A. Klaessens



Inhoud

Voorwoord	7
Bij de tweede druk	7
Hoofdstuk 1	
Inleiding	9
1.1 Statistiek in het laboratorium	9
1.2 Het gebruik van Excel	11
Functies	11
Gegevensanalyse	12
Noot	12
Hoofdstuk 2	
Spreading in meetwaarden	13
2.1 Inleiding	13
2.2 Benadering van de ware waarde	13
2.3 Maat voor het uitdrukken van spreading	15
2.4 Grafische weergave van de spreading van meetwaarden	19
2.5 De normale verdeling en de toepassing ervan	21
Standaardafwijking van een gemiddelde waarde	24
2.6 Covariantie en correlatie	26
2.7 De T-verdeling	31
2.8 De χ^2 - en de F-verdeling	36
Noten	39
Hoofdstuk 3	
Hulpmiddelen voor het nemen van beslissingen	41
3.1 Beslissen	41
3.2 De methode van hypothesetoetsing	43
3.3 Z-toets	45
Z-toets voor een enkele meting	46
Z-toets voor een groep van n metingen	47
Z-toets voor twee groepen metingen	49
3.4 T-toets voor vergelijking van een meting	
met een groep	52
T-toets voor een enkele meting	52
T-toets voor de vergelijking van het gemiddelde van	
een groep met een theoretische waarde	53
3.5 T-toets voor vergelijking van groepen metingen	55

3.6	Evaluatie van spreiding Vergelijking van de spreidingen van twee groepen meetwaarden	59 60
3.7	Toetsen van de verdeling van een groep waarnemingen χ^2 -toets	62 63
	Noten	67

Hoofdstuk 4**Lineaire regressie****69**

4.1	Inleiding	69
4.2	Theorie in het kort	70
4.3	Excel-hulpmiddelen voor regressie Benadering 1: functies RICHTING en SNIJPUNT Benadering 2: functie LIJNSCH Benadering 3: met Gegevensanalyse Benadering 4: met de grafische module	71 71 73 74 75
4.4	Kalibratie met lineaire regressie	76
4.5	Regressie-analyse met Excel Regressie-analyse met LIJNSCH Regressie-analyse met Gegevensanalyse Uitgebreidere modellen	80 80 83 86
4.6	Regressierekening tijdens methode-ontwikkeling	87
4.7	Regressierekening bij onderzoek en ontwikkeling	95
	Noten	105

Hoofdstuk 5**Variantie-analyse en ruisvermindering****107**

5.1	Bronnen van spreiding	107
5.2	Uitvoering van variantie-analyse	108
5.3	Uitvoering met Excel	110
5.4	Multifactoriële variantie-analyse Tweevoudige variantie-analyse zonder herhaling Tweevoudige variantie-analyse met herhaling Lopend gemiddelde Exponentieel gewogen gemiddelde	115 116 118 124 125
	Noten	127

Register**129**

	Nederlands – Engels	130
	Gegevensanalyse	131
	Engels – Nederlands	132
	Data analysis	133

Voorwoord

Er is geen twijfel over het nut van statistiek als hulpmiddel in het laboratorium. Wat soms wel eens toepassing in de weg staat, is de benodigde hoeveelheid rekenwerk, zeker voor mensen voor wie statistiek geen dagelijkse activiteit is. De stap naar een gespecialiseerd statistiekpakket is te groot. Maar velen realiseren zich niet dat ze al een statistiekpakket binnen hun bereik hebben. Namelijk in Excel. Iedereen kan wel over Excel beschikken op zijn werk. En daarin zit een groot aantal statistische hulpmiddelen.

Dit boek is geschreven om deze statistische hulpmiddelen van Excel over het voetlicht te brengen. Maar we kiezen wel de vraagstellingen van het laboratorium als uitgangspunt, het routinelaboratorium en het onderzoekslaboratorium.

Den Dolder, februari 2009

Dr. Jo Klaessens

Bij de tweede druk

Deze tweede druk is volledig aangepast aan Excel 2010. Ten opzichte van Excel 2007 en eerdere versies is er veel veranderd in de beschikbare statistische functies. In een aantal gevallen is alleen de naam gewijzigd, maar er zijn ook verschillende nieuwe functies toegevoegd. De opzet van het boek is echter ongewijzigd gebleven.

Den Dolder, februari 2013

Dr. Jo Klaessens

Bij de derde druk

Bij deze derde druk konden de wijzigingen beperkt blijven tot enkele kleine correcties. Na Excel 2010 zijn er inmiddels twee nieuwe versies verschenen: 2013 en 2016. De statistische functies zijn hierin echter gelijk gebleven. Daarom is besloten om de versie niet langer in de titel van het boek te vermelden.

Den Dolder, februari 2016

Dr. Jo Klaessens

Omdat alle analysesresultaten bij hetzelfde monster horen, gebruiken we de ongepaarde T-toets. Er is geen aanleiding om te veronderstellen dat het ene laboratorium altijd meer zou vinden dan het andere. Dus we gebruiken een dubbelzijdige toets. Merk op dat hier de analysemethoden van beide laboratoria worden vergeleken.

- 1 H_0 : de beide groepen zijn niet verschillend.
- 2 Kies het betrouwbaarheidsniveau $\alpha = 0,05$.
- 3 Stel de H_1 -hypothese vast: beide groepen zijn wel significant verschillend: dubbelzijdig.
- 4 Bereken $p = T.TEST(A2:A6; B2:B6; 2; 2) = 0,033$.
- 5 Omdat $p < \alpha$ verwerpen we H_0 .

We concluderen dat de methoden van beide laboratoria verschillend zijn. Vermoedelijk is dat ook de oorzaak van de klacht. De volgende stap is te onderzoeken waar de verschillen door veroorzaakt worden en hoe deze kunnen worden opgelost.

De drie varianten van T.TEST kunnen ook worden uitgevoerd met onderdelen van de module *Gegevensanalyse*:

- *T-toets: twee gepaarde steekproeven voor gemiddelden.*
- *T-toets: twee ongepaarde steekproeven met gelijke varianties* (opmerking: deze variant staat aangeduid als ‘gepaarde steekproeven’; dit moet ‘ongepaard’ zijn).
- *T-toets: twee steekproeven met ongelijke varianties.*

De aanroep van de functies spreekt voor zich. De uitvoer is een stuk uitgebreider. Tabel 3.7 geeft als voorbeeld de uitvoer die hoort bij toepassing 3.5.

	<i>Certificaat</i>	<i>Resultaat</i>	<i>Toelichting</i>
Gemiddelde	12,9	12,7	
Variantie	190	175	
Waarnemingen	11	11	
Pearson-correlatie	0,986		correlatiecoëfficiënt volgens formule 2.13
Schatting van verschil tussen gemiddelden	0		instelling: getoetst kan worden op een verschil
Vrijheidsgraden	10		
T- statistische gegevens	0,256		T-waarde zoals berekend volgens formule 3.3
P(T<=t) eenzijdig	0,402		overschrijdingskans p , enkelzijdig
Kritiek gebied van T-toets: eenzijdig	1,812		enkelzijdige kritieke waarde
P(T<=t) tweezijdig	0,803		overschrijdingskans p , dubbelzijdig
Kritiek gebied van T-toets: tweezijdig	2,228		dubbelzijdige kritieke waarde

Tabel 3.7

Uitvoer van de *T-toets: twee gepaarde steekproeven voor gemiddelden* bij toepassing op de gegevens van toepassing 3.5. De drie linker kolommen vormen de Excel-uitvoer; de rechter kolom bevat toelichtingen op de uitvoer.

De uitvoer wordt niet automatisch geüpdatet wanneer er iets wordt veranderd in het invoerbereik. Bij de functie T.TEST is dat wel het geval.

3.6 Evaluatie van spreiding

Vergelijking van een geschatte spreiding met een theoretische waarde

Om te onderzoeken of de standaardafwijking van een steekproef afwijkt van een verwachte of een theoretische waarde kunnen we gebruik maken van de χ^2 -verdeling. Immers, zoals we zagen in paragraaf 2.8, volgt de variabele $v \cdot s^2 / \sigma^2$ deze verdeling.

We kunnen in dit geval niet gebruikmaken van de Excel-functie CHI.TEST. Deze is bestemd voor een andere toepassing; zie paragraaf 3.7. In plaats daarvan gebruiken we de functies CHIKW.VERD en CHIKW.VERD.RECHTS.

1	Formuleer de nulhypothese H_0 : de gevonden standaardafwijking wijkt niet af van de theoretische waarde, d.w.z. $\sigma = \sigma_{th}$.		
2	Kies het betrouwbaarheidsniveau α (meestal $\alpha = 0,05$).		
3	H_1 -hypothese: $\sigma < \sigma_{th}$ (linkszijdig)	H_1 -hypothese: $\sigma \neq \sigma_{th}$ (dubbelzijdig)	H_1 -hypothese: $\sigma > \sigma_{th}$ (rechtszijdig)
4	Bereken: $\chi^2 = \frac{v \cdot s^2}{\sigma_{th}^2}$ met v het aantal vrijheidsgraden en s^2 de berekende variantie van de groep metingen.		
	$p =$ CHIKW.VERD($\chi^2, v; 1$)	Als $s < \sigma_{th}$ bereken: $p =$ 2·CHKW.VERD($\chi^2, v; 1$) Als $s > \sigma_{th}$ bereken: $p = 2 \cdot$ CHIKW.VERD.RECHTS(χ^2, v)	$p =$ CHIKW.VERD.RECHTS(χ^2, v)
5	Beslis: accepteer H_0 als $p > \alpha$; anders accepteer H_1 .		

Opmerking: in de middelste kolom wordt de enkelzijdige kans omgezet in een dubbelzijdige kans door met 2 te vermenigvuldigen.

Toepassing 3.7

Evaluatie van de herhaalbaarheid bij validatie

Het laboratorium is bezig met de implementatie van een nieuwe analysemethode die wordt uitgevoerd conform een ISO-norm. Bij de ontwikkeling van de norm is een methode-evaluerend ringonderzoek uitgevoerd volgens ISO 5725. De herhaalbaarheid die is vastgesteld bij dit ringonderzoek is in de norm gepubliceerd: $\sigma_{r,th} = 1,5$ ppm.

Als onderdeel van de validatie onderzoekt het laboratorium of de eigen herhaalbaarheid in overeenstemming is met de gepubliceerde waarde. De herhaalbaarheid is bepaald met acht metingen.

We vinden een kleinere waarde dan de gepubliceerde herhaalbaarheid geen probleem. Dus we voeren een enkelzijdige toets uit met als H_1 -hypothese dat de gevonden herhaalbaarheid groter is dan de gepubliceerde waarde.

Het is onze taak om de experimentele resultaten te analyseren. Bij eerste inspectie van de getallen zien we al meteen dat er sprake is van een ingewikkeld patroon. Met de twee parameters melk (x) en bitter (y) kunnen we de resultaten niet goed beschrijven. Het model $z = a + bx + cy$ is niet geschikt. Dit wordt bevestigd door de regressie-analyse op basis van dit model (zie tabel 4.19).

	A	B	C	D	E	F	G
24	Gegevens voor de regressie						
25	Meervoudige corr	0,4311					
26	R-kwadraat	0,1859					
27	Aangepaste klein	0,0050					
28	Standaardfout	1,7671					
29	Waarnemingen	12					
30							
31	Variantie-analyse						
32		<i>Vrijheidsgraden</i>	<i>Kwadratensom</i>	<i>Gemiddelde kwadraten</i>	<i>F</i>	<i>Significantie F</i>	
33	Regressie	2	6,416	3,208	1,027	0,396	
34	Storing	9	28,103	3,123			
35	Totaal	11	34,519				
36							
37		<i>Coëfficiënten</i>	<i>Standaardfout</i>	<i>T-statistische geg.</i>	<i>P-waarde</i>	<i>Laagste 95%</i>	<i>Hoogste 95%</i>
38	Snijpunt	4,419	1,711	2,582	0,030	0,548	8,289
39	melk (x)	0,046	0,034	1,376	0,202	-0,030	0,122
40	bitter (y)	0,502	1,250	0,402	0,697	-2,324	3,329
41							

Tabel 4.19

Uitvoer van de regressie-analyse van de gegevens van tabel 4.18 met het model $z = a + bx + cy$ (model 1).

Excel heeft dus het volgende model berekend:

$z = 4,419 + 0,046 \cdot x + 0,502 \cdot y$. We noemen dit model 1. Er is sprake van een kleine correlatiecoëfficiënt (aangepaste r^2 van 0,0050). Bij het onderdeel variantie-analyse zien we bij *Significantie F* een waarde $> 0,05$. De restspreiding (*Storing*) is niet significant kleiner dan de spreiding die verklaard wordt door het regressiemodel. Er is dus geen sprake van een significante relatie. Ook de parameters voor Melk en Bitter (*Coëfficiënten*) zijn niet significant.

Als we naar de resultaten kijken, ontstaat de indruk dat er een interactie bestaat tussen melk en bitter. Bijvoorbeeld: bij 10% melk neemt de waardering af bij toenemend gehalte bittercomponent, terwijl deze bij 50% melk dan juist toeneemt. Deze interactie kunnen we als volgt in het model opnemen: $z = a + bx + cy + dxy$. Nadat we eerst de kolom met de waarden voor $x \cdot y$ hebben berekend, voeren we opnieuw een regressie-analyse uit (tabel 4.20). Excel heeft nu het volgende model berekend:

$z = 9,825 - 0,127 \cdot x - 4,904 \cdot y + 0,173 \cdot x \cdot y$. We noemen dit model 2. Hiermee kunnen we de resultaten al beter beschrijven. De meeste spreiding wordt verklaard door het regressiemodel, maar het is nog (net) niet significant (zie cel F35). Het is duidelijk dat de interactieterm een belangrijke rol speelt.

	A	B	C	D	E	F	G
24	Gegevens voor de regressie						
25	Meervoudige cor	0,7643					
26	R-kwadraat	0,5842					
27	Aangepaste klei	0,4282					
28	Standaardfout	1,3395					
29	Waarnemingen	12					
30							
31	Variantie-analyse						
32		<i>Vrijheidsgraden</i>	<i>Kwadraten-som</i>	<i>Gemiddelde kwadraten</i>	<i>F</i>	<i>Significantie F</i>	
33	Regressie	3	20,165	6,722	3,746	0,060	
34	Storing	8	14,354	1,794			
35	Totaal	11	34,519				
36							
37		<i>Coëfficiënten</i>	<i>Standaardfout</i>	<i>T-statistische geg.</i>	<i>P-waarde</i>	<i>Laagste 95%</i>	<i>Hoogste 95%</i>
38	Snijpunt	9,825	2,344	4,191	0,003	4,418	15,231
39	melk-bitter(xy)	0,173	0,062	2,768	0,024	0,029	0,317
40	melk (x)	-0,127	0,068	-1,877	0,097	-0,282	0,029
41	bitter (y)	-4,904	2,171	-2,259	0,054	-9,909	0,102
42							

Tabel 4.20

Uitvoer van de regressie-analyse van de gegevens van tabel 4.15 met het model $z = a + bx + cy + dxy$ (model 2).

We besluiten het model uit te breiden met kwadratische termen. Dit model kan rondingen in het responsvlak beter volgen.

We krijgen dan een model met zes parameters:

$z = a + bx + cy + dxy + ex^2 + fy^2$. Omdat we twaalf metingen hebben, kunnen we dit bepalen. We houden nog zes vrijheidsgraden over om informatie over de restspreiding te bepalen. Na berekening van twee kolommen met de kwadraten voor x en y voert Excel de regressie-analyse uit.

	A	B	C	D	E	F	G
24	Gegevens voor de regressie						
25	Meervoudige cor	0,9722					
26	R-kwadraat	0,9452					
27	Aangepaste klei	0,8995					
28	Standaardfout	0,5615					
29	Waarnemingen	12					
30							
31	Variantie-analyse						
32		<i>Vrijheidsgraden</i>	<i>Kwadraten-som</i>	<i>Gemiddelde kwadraten</i>	<i>F</i>	<i>Significantie F</i>	
33	Regressie	5	32,63	6,525	20,70	0,001	
34	Storing	6	1,892	0,315			
35	Totaal	11	34,52				
36							
37		<i>Coëfficiënten</i>	<i>Standaardfout</i>	<i>T-statistische geg.</i>	<i>P-waarde</i>	<i>Laagste 95%</i>	<i>Hoogste 95%</i>
38	Snijpunt	3,920	1,634	2,399	0,053	-0,079	7,919
39	melk-kw (y ²)	-2,580	1,375	-1,876	0,110	-5,945	0,786
40	bitter-kw (x ²)	-0,006	0,001	-6,001	0,001	-0,008	-0,003
41	melk-bitter (xy)	0,173	0,026	6,603	0,001	0,109	0,237
42	melk (x)	0,213	0,063	3,368	0,015	0,058	0,368
43	bitter (y)	0,256	2,897	0,088	0,933	-6,834	7,345
44							

Tabel 4.21

Uitvoer van de regressie-analyse van de gegevens van tabel 4.18 met het model $z = a + bx + cy + dxy + ex^2 + fy^2$ (model 3).

Een voorbeeld van de eerste toepassing vinden we bij productontwikkeling: heeft een bewerkingsstap (bijvoorbeeld om de houdbaarheid van een product te verlengen) invloed op de smaakbeleving? Als dat zo is, zal deze een significante bijdrage hebben aan de totale spreiding van de meetwaarden van de smaakbeleving. Bij interlaboratoriumonderzoek zien we vaak beide stappen. Eerst wordt vastgesteld of er inderdaad sprake is van significante verschillen tussen de laboratoria. Als dat zo is, wordt vervolgens berekend hoe groot die bijdrage aan de totale spreiding is (de interlaboratoriumspreiding).

Variantie-analyse wordt ook vaak gebruikt als een soort uitbreiding op de T-toets, namelijk als er meer dan twee groepen waarnemingen zijn. Dit is mogelijk zowel voor de gepaarde als de ongepaarde T-toets. De enkelvoudige variantie-analyse kan dienen als de uitbreiding op de ongepaarde T-toets. Als de H_0 -hypothese opgaat dat er geen significante verschillen zijn, zal er geen significante bijdrage zijn van tussen-groep spreiding aan de totale spreiding. Voor de uitbreiding op de gepaarde T-toets moeten we een extra variabele inbrengen. Zie paragraaf 5.4 en toepassing 5.6.

5.2 Uitvoering van variantie-analyse

Excel heeft een functie voor variantie-analyse. Dit verhoogt de toepasbaarheid aanzienlijk, omdat de techniek relatief veel rekenwerk vraagt. De toelichting in Excel is vrij summier. Daarom moeten we eerst de theorie uitwerken voordat we ingaan op de toepassingen.

Algemeen gesproken onderzoeken we of variantie van een bepaalde factor een significante bijdrage geeft aan de uiteindelijke spreiding (bijvoorbeeld de verschillende laboratoria). De ANOVA terminologie spreekt dan van verschillende behandelingen ('treatments') voor deze factor. Het aantal behandelingen noemen we n . Per behandeling wordt een aantal van m herhalingen ('replicates') uitgevoerd (zie tabel 5.1). In totaal voeren we $n \cdot m$ metingen uit.

Behandeling	Herhaling			
	1	2	..	m
1	x_{11}	x_{12}	..	x_{1m}
2	x_{21}	x_{22}	..	x_{2m}
..
..
n	x_{n1}	x_{n2}	..	x_{nm}

Tabel 5.1

Experimenten behorend bij een ANOVA-onderzoek met n behandelingen en m herhalingen per behandeling.

Als we één factor onderzoeken, spreken we van een enkelvoudige ANOVA (in het Engels: ‘one factor’ of ‘one way’ ANOVA). Met dezelfde systematiek kunnen ook twee factoren of meer worden onderzocht. Een belangrijke eis voor de toepassing van variantie-analyse is dat de resultaten normaal verdeeld zijn.

Uit de resultaten van tabel 5.1 kunnen we de spreiding binnen behandeling 1 berekenen:

$$s = \sqrt{\frac{1}{m-1} \sum_{j=1}^m (x_{1j} - \bar{x}_1)^2}$$

Hierin is \bar{x}_1 het gemiddelde van behandeling 1. De beste schatting van de basisspreiding ‘binnen-de-behandelingen’ krijgen we als we gebruikmaken van de gegevens van alle behandelingen:

$$s_0 = \sqrt{\frac{1}{n \cdot (m-1)} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_{i1})^2} \quad (5.1)$$

Voor de spreiding die het gevolg is van de behandelingen, zullen we de gemiddelde waarden van de behandelingen als uitgangspunt nemen:

$$s_{tussen} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{x}_i - \bar{x})^2} \quad (5.2)$$

\bar{x} is de gemiddelde waarde van alle resultaten. Omdat we werken met gemiddelden van m waarnemingen, is s_{tussen} opgebouwd uit de spreiding ‘tussen-de-behandelingen’ (s_a) vermeerderd met een aandeel van s_0 :

$$s_{tussen} = \sqrt{s_a^2 + \frac{s_0^2}{m}}$$

ANOVA begint met de volgende opdeling van de kwadratenom:

$$\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})^2 = m \cdot \sum_{i=1}^n (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2 \quad (5.3)$$

Afgezien van de vrijheidsgraden en de worteltekens, herkennen we in het rechterdeel m maal vergelijking 5.2 en eenmaal vergelijking 5.1. In tabel 5.2 wordt alles systematisch bijeen gebracht. Hierbij worden ook de gebruikelijke namen van de kwadraten-sommen $SS_1 - SS_3$ (‘sum of squares’) opgenomen en de gebruikelijke namen van de spreidingen s_3 (zie vergelijking 5.1) en s_2 (zie vergelijking 5.2). Merk op dat $s_2^2 = m \cdot s_{tussen}^2$. De uitvoer van de Excel-functie volgt dit schema.

Indien significant kunnen we de bijbehorende spreidingen ook bepalen op basis van de formules in de rechter kolom van tabel 5.12. Dit is echter niet gebruikelijk. We beperken ons bijna altijd tot de evaluatie of effecten significant zijn. In de klinische chemie wordt zo'n schema wel eens gebruikt om een opsplitsing te maken van binnen-run, tussen-run en tussen-dag precisie.

Interpretatie van een tweevoudige variantie-analyse kan behoorlijk complex zijn. Het is aan te bevelen om altijd de gemiddelde waarden per behandeling grafisch uit te zetten. Dit geeft extra informatie voor de interpretatie. Wanneer lijnen parallel lopen, is er geen interactie. Gekruiste lijnen duiden op interactie. Tabel 5.13 toont enkele verschillende situaties en hun betekenis. Effecten spelen natuurlijk pas een rol wanneer zij in de F-toets significant zijn.

	<ul style="list-style-type: none"> • Verschillen voor factor A • Geen verschillen voor factor B • Geen interactie
	<ul style="list-style-type: none"> • Geen verschillen voor factor A • Verschillen voor factor B • Geen interactie <p>Merk op dat deze afbeelding getransponeerd is ten opzichte van de bovenstaande.</p>
	<ul style="list-style-type: none"> • Verschillen voor factor A • Verschillen voor factor B • Geen interactie
	<ul style="list-style-type: none"> • Interactie tussen de factoren • Verschillen voor factor A • Verschillen voor factor B

Tabel 5.13

Grafische weergave van de gemiddelde waarde per behandeling, vastgelegd door factor A met instellingen A1, A2 en A3 (langs de x-as) en factor B met instellingen B1 (—), B2 (---) en B3 (---).

Toepassing 5.5

Onderzoek effect hittebehandeling legeringen op de treksterkte

Om het juiste materiaal te vinden voor een bepaalde toepassing worden verschillende varianten met elkaar vergeleken op treksterkte². De varianten verschillen in de samenstelling van de legering (factor 1) en in de hittebehandeling die het materiaal ondergaat (factor 2). Het is goed mogelijk dat er een interactie bestaat tussen beide factoren. Om deze reden wordt een tweevoudige variantie-analyse uitgevoerd met herhaling.

Onderzocht worden drie verschillende hittebehandelingen (HB 1, HB 2 en HB 3) en drie verwante legeringen (A, B en C). Afbeelding 5.3 toont de meetwaarden en het invoervenster van de variantie-analyse (*Multifactoriële variantie-analyse met herhaling*).

	A	B	C	D	E	F	G	H	I	J	K
1	treksterkte legeringen										
2		legering	A	B	C						
3	hitte	HB 1	24,2	27,7	28,4						
4	behandeling		23,1	26,8	29,8						
5		HB 2	25,6	27,1	27,0						
6			23,2	26,8	27,8						
7		HB 3	23,9	25,6	26,4						
8			24,6	25,9	27,4						
9											
10											
11											
12											
13											
14											
15											

Multifactoriële variantie-analyse met her...			
Invoer			
Invoerbereik:	\$B\$2:\$E\$8		
Aantal rijen per steekproef:	2		
Alfa:	0,05		
Uitvoeropties			
<input checked="" type="radio"/> Uitvoerbereik:	\$A\$11		
<input type="radio"/> Nieuw werkblad:			
<input type="radio"/> Nieuwe werkmapij:			
		OK	Annuleren
		Help	

Afbeelding 5.3

Herhaalde metingen voor drie hittebehandelingen en drie legeringen, samen met het invoervenster voor de tweevoudige variantie-analyse.

Het invoervenster heeft geen keuzeveld met betrekking tot *labels*. Toch moeten zij altijd aan het invoerbereik worden toegevoegd. Anders wordt een foutmelding gegeven. De herhalingen per behandeling dienen onder elkaar gezet te worden, zoals in afbeelding 5.3. Het aantal herhalingen dient te worden ingevuld bij *Aantal rijen per steekproef*. Voor de Excel-functie is het noodzakelijk dat de gegevens op deze manier gegroepeerd worden.

	A	B	C	D	E	F	G
40	Variantie-analyse						
41	<i>Bron van variatie</i>	<i>Kwadratensom</i>	<i>Vrijheidsgraden</i>	<i>Gemiddelde kwadraten</i>	<i>F</i>	<i>P-waarde</i>	<i>Kritische gebied van F-toets</i>
42	Steekproef	3,24	2	1,62	2,422	0,144	4,256
43	Kolommen	43,0	2	21,5	32,1	0,000	4,256
44	Interactie	5,23	4	1,31	1,952	0,186	3,633
45	Binnen	6,02	9	0,67			
46	Totaal	57,5	17				
47							

Tabel 5.14

Uitvoer van de multifactoriële variantie-analyse met herhaling, uitgevoerd op de gegevens van afbeelding 5.3.