TECHNISCHE HOGESCHOOL EINDHOVEN
Afdeling Algemene Wetenschappen
Onderafdeling der Wiskunde

Vraagstukken en antwoorden behorende bij het college STATISTISCHE THEORIE van PROEFOPZETTEN

van

Prof. Dr. R. Doornbos

samengesteld door

J.Th.M. Wijnen

Najaarssemester 1980

Bibl Wsk

Technische Hogeschool
Eindhoven

Dictaatnummer 2.244
Prijs f. 6,00

Onderafdeling der Wiskunde en Informatica

Vraagstukken en antwoorden behorende bij het college

Statistische theorie van proefopzetten

Van prof.dr. R. Doornbos

Samengesteld door J.Th.M. Wijnen

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

Vraagstukken en antwoorden behorende bij het college

STATISTISCHE THEORIE VAN PROEFOPZETTEN

van Prof.dr. R. Doornbos

Samengesteld door J.Th.M. Wijnen

Najaarssemester 1980

Inhoudsbeschrijving

Vraagstukken en antwoorden bij

STATISTISCHE THEORIE van PROEFOPZETTEN

R. Doornbos en J.Th.M. Wijnen

najaarssemester 1980

De vindplaats van de antwoorden onder "Antwoorden blz" is achterin het dictaat.
De nummering der bladzijden aldaar is gelijk aan die der vraagstukken

JdG, 25 Juli 2005

Inhoudsopgave

1. Inleiding

1.1. Een proefopzet had ten doel de axiale kracht op een boor te bepalen als functie van de omwentelingssnelheid S, de voedingssnelheid V en het materiaal M.

Factoren en niveau's

1) Omwentelingssnelheid S: 100, 220, 475, 715, 870 omw./min.
2) Voedingssnelheid V: 0,04, 0,08, 0,14 mm/sec.
3) Materiaal: M_1, M_2.
4) Herhalingen H_1, H_2.

Responsie: de axiale kracht in eenheden van 10 gr.
Codering : alle waarnemingen zijn met 200 verminderd.
Resultaat:

		S				
	V	100	220	475	715	870
M_1	0.04	122 110	108 85	108 60	66 50	89 60
	0.08	332 330	276 310	248 295	248 275	276 310
	0.14	640 500	612 500	543 450	612 610	696 610
M_2	0.04	192 170	136 130	122 85	108 75	136 75
	0.08	386 365	333 330	318 330	472 350	499 390
	0,14	810 725	779 670	810 750	893 890	1820 890

Volgens opgaaf was de proef volledig verloot.
Iedere waarneming werd uitgevoerd met een apart proefstuk uit het betreffende materiaal.

a) Geef de waarnemingen weer in een nette overzichtelijke grafiek. Denk daarbij aan schaal, eenheden, onderschrift.
b) Door verloting toe te passen kan men de invloed van 2 (niet genoemde) factoren beschouwen als toevallige fouten. Welke factoren zijn dat?
c) De waarnemingen doen twijfelen aan de juiste uitvoering van volledige verloting. Waarom?

1.2. Geef een omschrijving of definitie van de volgende statistische begrippen:

a) experimentele eenheid;
b) factor;
c) niveau;
d) behandeling;
e) verloting;
f) blokvorming;
g) vrijheidsgraden;
h) contrast;
j) restvariantie;
k) lineair model.

1.3. a) Wat is het doel van

1) verloting,
2) blokvorming.

b) Wat zijn de onderstellingen bij een lineair model?

1.4. Er moeten drie werkmethoden worden vergeleken. Voor elke methode worden 8 metingen van de opbrengst van één manuur verricht.
Splits het totale aantal vrijheidsgraden in overeenstemming met de opzet van het experiment:

a) zonder blokvorming,
b) met blokvorming.

1.5. Er moeten 4 methoden A, B, C en D worden vergeleken. Met elke methode worden 9 monsters gemeten.
Geef de splitsing van het totale aantal vrijheidsgraden bij een opzet van het experiment:

a) zonder blokvorming,
b) met blokvorming.

1.6. Een fabrikant wil graag 2 methoden, A en B, vergelijken die de enzymengroei in een fermentatieproces meten. Daartoe kan hij kiezen uit 2 proefopzetten:

I Neem aselect 18 monsters uit het vat met goed geroerde vloeistof en verloot deze over de methoden A en B (elk 9 monsters).

II Neem aselect 9 monsters uit het vat, deel ieder van deze monsters in 2 delen en meet in het ene m.b.v. methode A en in het andere m.b.v. methode B.

a) Geef van beide proefopzetten het model.

b) Welke opzet zou U de fabrikant aanraden? Waarom? (Het kostenverschil tussen het nemen van 18 monsters en het nemen van 9 monsters is verwaarloosbaar.)

c) De fabrikant volgt Uw raad op en past de door U gekozen opzet toe. De matrix der (gecodeerde) waarnemingen is:

Methode A	6	7	7	9	4	7	8	8	4
Methode B	6	7	6	8	4	6	8	7	3

Toets m.b.v. de t-toets de nulhypothese H_0, dat de methoden A en B gelijk zijn, tegen de alternatieve hypothese dat ze verschillen ($\alpha = 0.05$).

d) Wat moet U in Uw model aannemen, opdat U met recht de t-toets mag gebruiken?

N.B. Degenen die het college Toegepaste Statistiek niet hebben gevolgd kunnen de u-toets toepassen en daarbij gebruik maken van het volgende gegeven:

$$\text{proefopzet I : var } \underline{x}_{A_i} = \text{var } \underline{x}_{B_i} = 3 \qquad (i = 1,\ldots,9)$$

$$\text{proefopzet II: var}(\underline{x}_{A_i} - \underline{x}_{B_i}) = 0{,}3 \qquad (i = 1,\ldots,9)\ .$$

1.7. Een psychologisch experiment bestaat uit het vergelijken van responsietijden (in sec.) van 2 verschillende stimuli A en B. Daartoe kregen 9 personen eerst stimulus A en later stimulus B toegediend.
De waarnemingen zijn

	Personen								
stimulus A	9.4	7.8	5.6	12.1	6.9	4.2	8.8	7.7	6.4
stimulus B	10.3	8.9	4.1	14.7	8.7	7.1	11.3	5.2	7.8

a) Geef het model.
b) Ga na of er verschil is tussen de gemiddelde responsietijden van de 2 stimuli ($\alpha = 0.05$).
c) Geef een 95% betrouwbaarheidsinterval voor het gemiddelde verschil in responsietijden.
d) Geef kritiek op de opzet van dit experiment.

N.B. Voor degenen die geen Toegepaste Statistiek hebben gedaan:
$\mathrm{var}(\underline{\bar{x}}_A - \underline{\bar{x}}_B) = 0{,}36.$

1.8. In zeker bedrijf dienen twee werkmethoden te worden vergeleken. De bedrijfsleiding stelt zich voor bij elk der methoden 8 maal de opbrengst van één manuur te bepalen. Er zijn 50 personeelsleden aanwezig.

a) Beschrijf hoe U het experiment denkt uit te voeren.

De matrix van waarnemingen die behoort bij de door U gekozen opzet van het experiment, luidt:

methode A:	277	240	253	299	324	322	310	353
methode B:	295	233	305	309	354	353	305	367

b) Toets H_0: $\mu_A = \mu_B$ tegen H_{alt}: $\mu_A \neq \mu_B$ (met $\alpha = 0{,}05$).

N.B. Voor deze toets dient gebruik te worden gemaakt van de $\underline{t}$-verdeling; zij die het college Toegepaste Statistiek niet hebben gevolgd mogen gebruik maken van het volgende gegeven:

$$\mathrm{var}(\underline{x}_{A_i}) = \mathrm{var}(\underline{x}_{B_j}) = 1600 \qquad (\text{voor } i,j = 1,\ldots,8)$$

en/of van

$$\mathrm{var}(\underline{x}_{A_i} - \underline{x}_{B_i}) = 400 \qquad (\text{voor } i = 1,\ldots,8)$$

(zodat het mogelijk is in plaats van de $\underline{t}$-verdeling de normale verdeling te benutten).

1.9. Om twee rekenmachines te vergelijken werd aan zes geoefende rekenaars gevraagd op beide machines een bepaald type berekening uit te voeren. In onderstaande tabel staan de waargenomen tijden.

Rekenaar	Rekenmachine A	Rekenmachine B
1	22	13
2	12	17
3	12	9
4	15	10
5	23	5
6	24	12

Benodigde tijden (aantal seconden verminderd met 100) voor het uitvoeren van een berekening op mamachines A en B door zes rekenaars.

a) Geef het model.

b) Ga na of er verschil is tussen de gemiddelde rekentijden van de twee rekenmachines ($\alpha = 0.05$).

c) Geef een 95% betrouwbaarheidsinterval voor het gemiddelde verschil in rekentijd.

N.B. Voor degenen die geen Toegepaste Statistiek hebben gedaan:
$\mathrm{var}(\bar{\underline{x}}_A - \bar{\underline{x}}_B) = 13.$

2. Lineaire modellen

2.1. Een variabele y is een functie van 3 onafhankelijke variabelen x_1, x_2 en x_3. De functie luidt:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \underline{e} \ .$$

a) Pas dit model aan bij de volgende waarnemingen (d.w.z. geef schattingen voor β_0, β_1, β_2 en β_3):

x_1 =	-3	-2	-1	0	1	2	3
x_2 =	5	0	-3	-4	-3	0	5
x_3 =	-1	1	1	0	-1	-1	1
y =	1	0	0	1	2	3	3

b) Geef een 95%-betrouwbaarheidsinterval voor σ_0^2.

2.2. Gegeven is de volgende serie waarnemingen

y_i =	1	0	0	-1	-1	0	0
x_i =	-3	-2	-1	0	1	2	3

We beschouwen het model

$$\underline{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \underline{e} \ .$$

a) Geef schattingen van de parameters van bovenstaand model.

b) Wat is de schatting voor σ_0^2 en wat is het aantal vrijheidsgraden bij deze schatting?

c) Geef een 95% betrouwbaarheidsinterval voor σ_0.

2.3. We beschouwen nog eens de serie waarnemingen van opgave 2.2, maar nu het model

$$\underline{y} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \underline{e}$$

waarin

$$x_1 = x$$

$$x_2 = x^2 - 4 \ .$$

a) Geef schattingen van α_0, α_1, α_2 en σ_0^2.

b) Wat is het effect van deze codering?

c) Wat is het verband tussen α en β (α en β zijn hier de kolomvectoren met elementen α_0, α_1 en α_2 resp. β_0, β_1 en β_2)

2.4. Pas een rechte lijn aan aan de volgende vijf gegeven punten:

x	-1	0	1	2	3
y	3	2	1	1	0,5

a) Teken de punten en de grafiek van de aangepaste rechte.

b) Bereken $\hat{e} = y - Xb$ en verifieer $X'\hat{e} = 0$ (dat is: $\hat{e}$ staat loodrecht op de door de verklarende vectoren opgespannen ruimte).

c) Verifieer $y'y = (Xb)'Xb + \hat{e}'\hat{e}$ (dat is totale kwadr.som = door regressie verklaarde kwadr.som + restkwadr.som)
Met welke aantallen vrijheidsgraden corresponderen de 3 kwadratensommen?

d) Pas eveneens een rechte lijn aan aan de volgende 5 gegeven punten:

x_1	-2	-1	0	1	2
y	3	2	1	1	0,5

Wat is het effect vande codering $x_1 = x - 1$?

2.5. We beschouwen de volgende waarnemingen

y_{1i}: 12, 81, 37, 15 ;
y_{2i}: 35, 82, 51, 3 .

a) Onderstel dat de waarnemingen afkomstig zijn uit normaal verdeelde populaties met dezelfde variantie.
   i) Geef een schatting voor σ^2.
   ii) Toets met de t-toets de hypothese $\mu_1 = \mu_2$ ($\alpha = 0.05$).
   iii) Geef een geschikt lineair model.
   iv) Bereken schattingen voor de parameters van het model (β en σ_0^2).

b) Onderstel dat de waarnemingen y_{1i} en y_{2i} gepaard zijn (blokvorming).
   i) Wat wordt nu de schatting voor σ^2?
   ii) Toets de hypothese $\mu_1 = \mu_2$ ($\alpha = 0.05$).
   iii) Geef een geschikt lineair model.
   iv) Wat zijn de schattingen voor de modelparameters β en σ_0^2 ?

c) Vergelijk de verschillende resultaten.

2.6. Ter vergelijking van drie werkmethoden worden voor elke methode 4 waarnemingen van de opbrengst van één manuur uitgevoerd:

I zonder blokvorming, totaal verloot,

II met blokvorming, en verloting binnen blokken.

a) Detailleer voor elk der beide beschreven experimenten het model $y = X\beta + \underline{e}$.

b) Stel de matrix X op en bereken X'X (voor I èn II).

c) Bepaal voor het experiment onder I $(X'X)^{-1}$ en bereken b uit de volgende waarnemingen:

meth.	A	307	[illegible]	270	312
	B	[illegible]	306	303	325
	C	317	287	274	303

$\longrightarrow j$ $\quad j = 1,2,3,4$

$i \downarrow$ $\quad i = 1,2,3$

Vergelijk b met Uw schatting van de parameters in het model

$$\underline{y}_{ij} = \mu + \alpha_i + \underline{e}_{ij} \qquad (\text{met } \sum_i \alpha_i = 0 \text{ en } \mathcal{E}\underline{e}_{ij} = 0, \text{ dus ook } \mathcal{E}\sum_j \underline{e}_{ij} = 0) .$$

d) Geef de splitsing van het totale aantal ter beschikking gestelde vrijheidsgraden voor beide modellen.

2.7. De opbrengst van een chemisch produktieproces is bepaald bij twee concentraties van een katalysator en bij twee temperaturen.

De resultaten waren

		Concentratie	
		0,2	0,4
Temperatuur	60°	66	50
	80°	72	61

Het bijbehorende model luidt:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underline{e} .$$

Hierin is

$x_1 = -1$ als de concentratie = 0,2

$x_1 = +1$ als de concentratie = 0,4

$x_2 = -1$ als de temperatuur = 60°

$x_2 = +1$ als de temperatuur = 80°.

a) Bereken $\hat{\beta}$ en $\hat{\sigma}_0^2$.

b) Bereken r en verifieer dat $r'r = KS_r$.

c) Uit vroegere bepalingen van opbrengsten is een schatting bekend van σ_0^2:

$$\hat{\sigma}_0^2 = 4 \; , \quad \nu = 5 \; .$$

Toets of de schatting uit a) significant verschilt van de schatting uit vroegere bepalingen.

d) Geef het 95% betrouwbaarheidsinterval voor β_1 als $\hat{\sigma}_0^2 = 4$, $\nu = 5$.

2.8. Om het effect van 4 factoren (temperatuur T_1; druk P; katalysator K; temperatuur T_2) op de opbrengst van een chemisch produkt te onderzoeken werd een zogeheten 2^4-factorieel experiment (een proefopzet met 4 factoren elk op 2 niveaus) uitgevoerd. De niveaus van de factoren worden als volgt gecodeerd:

T_1	x_1
50°C	-1
70°C	1

P	x_2
10 atm	-1
20 atm	1

K	x_3
1	-1
2	1

T_2	x_4
100°C	-1
200°C	1

a) Geef voor iedere factor de relatie tussen originele en gecodeerde variabele.

b) Bereken uit de volgende waarnemingen schattingen voor β en σ_0^2 als het model luidt:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \underline{e} \; .$$

x_1	-1	x_2	-1	22.2	24.5	24.4	25.9
			+1	19.4	24.1	25.2	18.4
	+1	x_2	-1	22.1	19.6	23.5	16.5
			+1	14.2	12.7	19.3	16.0
				-1	+1	-1	+1
				x_3		x_3	
				-1		+1	
				x_4			

c) Wat is de schatting van y ?

d) Bereken $r'r$ $\quad (r = y - \hat{y})$ en vergelijk dit met KS_r (restkwadratensom).

e) Toets met de t-toets (α = 0.05) de nulhypothese dat T_1 geen invloed heeft. Geef ook een 95%-betrouwbaarheidsinterval voor β_1, de parameter behorende bij T_1.

f) Geef een 90%-betrouwbaarheidsinterval voor de verwachting van $\hat{\underline{y}}$ als $T_1 = 50$, $P = 20$, $K = 1$ en $T_2 = 200$. Merk op, dat $\hat{\underline{y}}$ een lineaire functie is van de b_i.)

2.9. Een experiment om 3 werkmethoden te vergelijken leverde de volgende gegevens op (zie ook vraagstuk 2.6):

methode A:	307	271	270	312
methode B:	318	306	303	325
methode C:	317	287	274	303 .

We beschouwen een model zonder blokvorming

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underline{e}$$

met de matrices:

$$y' = [307, 271, 270, 312, 318, 306, 303, 325, 317, 287, 274, 303]$$

en

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -2 & -2 & -2 & -2 & 1 & 1 & 1 & 1 \end{bmatrix} .$$

a) Bepaal $(X'X)$, $(X'X)^{-1}$, $X'y$, b, $\hat{\sigma}_0^2$.

b) Bepaal het 95% betrouwbaarheidsinterval voor σ_0.

c) x_1 en x_2 bepalen samen de verschillen tussen de 3 methoden. Bereken het gemiddelde kwadraat (GK) voor de verschillen tussen de methoden.
(GK = KS/ν, ν = het aantal vrijheidsgraden.)

d) Vergelijk het onder c) berekende GK met $\hat{\sigma}_0^2$.

2.10. Als het experiment van opgaven 2.6 en 2.9 zou zijn uitgevoerd met blokvorming, dan moeten we in het model nog 3 variabelen toevoegen om de verschillen tussen blokken te kunnen beschrijven:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \underline{e} .$$

De "design-matrix" wordt dan bv.:

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -2 & -2 & -2 & -2 & 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{bmatrix}$$

a) Bepaal $(X'X)$, $(X'X)^{-1}$, $X'y$, b, $\hat{\sigma}_0^2$.

b) Bereken de GK's voor methoden en voor blokken en vergelijk beide met $\hat{\sigma}_0^2$.

c) Vergelijk het resultaat met dat van opgave 2.9.

2.11. Bij een veresteringsproces werd het omzettingspercentage gemeten bij 6 combinaties van temperatuur en concentratie van een katalysator.

T(emperatuur) in °C	C(oncentratie) in mol/gr $\times 10^{-4}$	Omzettingspercentage
175	4	58
175	8	63
200	4	72
200	12	82
225	8	85
225	12	90

We beschouwen het model

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underline{e} .$$

Hierin is x_1 := een geschikte codering voor de temperatuurniveaus
x_2 := " " " " " concentratieniveaus
y := omzettingspercentage.

a) Bepaal schattingen voor de onbekende parameters en voor σ_0^2.

b) Toets de hypothese: $\beta_2 = 0$ $(\alpha = 0{,}05)$.

c) Welke onderstellingen liggen aan deze toets ten grondslag?

d) Geef een 95% betrouwbaarheidsinterval voor het verwachte omzettingspercentage bij een temperatuur van 200°C en een concentratie van 10^{-3} mol/gr.

2.12. Gegeven zijn de volgende (gemakshalve reeds gecodeerde) waarnemingen (aantal bakteriën per mm^3) voor 2 typen bacteriën A_1 en A_2 op tijdstippen t = 10, 20, 30, 40, 50.

	t				
	10	20	30	40	50
A_1	8.0	9.0	9.1	10.2	10.4
A_2	10.0	10.3	12.2	12.6	13.9

I. Uitgangspunt: de toename van het aantal bacteriën per tijdseenheid is voor beide typen gelijk.

a) Pas het volgende lineaire model aan:

$$\underline{y} = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \underline{e}$$

waarin:

x_1 voor het type bacteriën
x_2 voor de tijd.

Codeer liefst zo, dat X'X een diagonaalmatrix wordt.

b) Teken de punten in het platte vlak, teken de grafieken van de aangepaste rechten en geef de componenten van Uw schatting *b* in de figuur aan. Hoeveel vrijheidsgraden zijn beschikbaar voor de schatting van de restvariantie?

II. Uitgangspunt: we willen nu de groeisnelheden van de bacteriesoorten vergelijken.

a) Pas daartoe het volgende model aan:

$$\underline{y} = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \underline{e}$$

waarin:

x_1 voor het type bacteriën
x_2 voor de tijd
$x_3 = x_1 x_2$ de interactie van de (kwalitatieve) factor type bacterie en de (kwantitatieve) factor tijd.

b) Teken in een tweede figuur de punten, de aangepaste rechten en de componenten van Uw schatting b.
Hoeveel vrijheidsgraden zijn nu beschikbaar voor de restvariantie?

Opmerking. Het experiment is een volledig factorieel experiment (er zijn waarnemingen verricht bij elk der mogelijke niveaucombinaties der in het geding zijnde factoren).

2.13. Bij een bepaald type radiobuis heeft men de verzadigingsemissie onderzocht als functie van de gewichtshoeveelheid Ba $Sr(CO_3)_2$ op de kathode. De uitkomsten waren:

x = hoeveelheid Ba $Sr(CO_3)_2$ in mg	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4
$y = \bar{I}_S$ in mA	10	13	16	18	19	21	19	17

Iedere waarneming is het gemiddelde van 10 buizen.
We beschouwen het model:

$$\underline{y} = \alpha_0 + \alpha_1 x + \underline{e} .$$

a) Door een geschikte codering krijgt men het model

$$\underline{y} = \beta_0 x_0 + \beta_1 x_1 + \underline{e}$$

waarbij $x_0 \equiv 1$ en $(x_0,x_1) = 0$.
Bepaal b, KS_t, KS_v, KS_r, $\hat{\sigma}_0^2$, $\hat{var}(\underline{b})$.
Wat is de schatting voor de variantie tussen buizen (voor buizen met eenzelfde hoeveelheid Ba $Sr(CO_3)_2$)?

b) Bepaal ook $\hat{var}(\underline{a})$ als $\underline{a}$ de schatter is van α.
Geef een 95%-betrouwbaarheidsinterval voor α_1.

c) Bereken $\hat{y}$ en r en verifieer $r'r = KS_r$.

d) Geven de gevonden resultaten aanleiding tot het verwerpen van het beschouwde model?
Motiveer Uw antwoord.

2.14. De opbrengst van een chemisch proces hangt af van reactietijd en temperatuur. De niveaus van de factoren worden gegeven door

$$x_1 = \frac{\text{tijd} - 90}{10}$$

$$x_2 = \frac{\text{temperatuur} - 205}{10} .$$

De waarnemingen zijn:

x_1	0	0	-1	+1	-1	+1	0	0	$\sqrt{2}$	$-\sqrt{2}$	0	0
x_2	0	0	-1	-1	+1	+1	0	0	0	0	$\sqrt{2}$	$-\sqrt{2}$
y	88	87	79	84	91	77	90	85	81	83	80	81

Men wil het volgende model aanpassen:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \underline{e} .$$

a) Bereken $\hat{\beta}$ en $\hat{\sigma}_0^2$.

b) Toets de hypothese: $\beta_3 = \beta_4 = \beta_5 = 0$.

2.15. In een electrisch apparaat wilde men op een gegeven plaats de spanning meten onder twee omstandigheden A en B. Men organiseerde daartoe de volgende proefopzet.

Men koos uit de aanwezige voorraad aselect 10 apparaten. Vervolgens werd van ieder apparaat eerst de spanning gemeten onder omstandigheid A; daarna werd gemeten onder omstandigheid B.

Het omschakelen van A naar B vergde nogal wat tijd zodat het hele experiment 2 dagen duurde.

Bij het toepassen van de t-toets op de gemiddelde spanningen onder A en B bleek: $t_8 = 3.07$.

Men besloot toen de nulhypothese: omstandigheden A en B verschillen niet, te verwerpen.

a) Wat is Uw commentaar op deze proefopzet en op de conclusies?

b) Geef een alternatieve proefopzet en maak daarbij een schema in welke volgorde de metingen worden verricht.

2.16. Gegeven zijn de volgende waarnemingen, verkregen uit een factorieel experiment met 2 factoren:

	B_1	B_2	B_3	B_4
A_1	5.2	5.9	7.7	7.9
A_2	8.2	9.0	9.1	10.5
A_3	10.0	10.3	12.1	12.7

a) Bereken de variantie-analyse-tabel.

b) Toets welke effecten significant zijn als gegeven is dat er geen interactie is tussen A en B.

c) Zij nu B een kwantitatieve factor:

B_1: t = 10 sec $\qquad$ B_3: t = 40 sec

B_2: t = 25 sec $\qquad$ B_4: t = 55 sec.

De onderstelling is dat

$$\mathcal{E}\underline{y}_{A_i}(t) = \mu_{A_i} + \alpha_1 t + \alpha_2 t^2 .$$

Codeer:

$$t^* = \frac{t - 32.5}{7.5}$$

en

$$\begin{array}{llll} x_0 \equiv 1 & & & \\ x_1 = -1 & \text{voor } A_1 ; & x_2 = 1 & \text{voor } A_1 \\ x_1 = 0 & \text{voor } A_2 ; & x_2 = -2 & \text{voor } A_2 \\ x_1 = 1 & \text{voor } A_3 ; & x_2 = 1 & \text{voor } A_3 \\ x_3 = t^* & ; & x_4 = \dfrac{t^{*2} - 5}{4} . & \end{array}$$

Bepaal schattingen voor β en var $\hat{\underline{\beta}}$ bij het model $\mathcal{E}\underline{y} = X\beta$.

d) Toets H_0: $\beta_4 = 0$ ($\alpha = 0.05$).

e) Toets H_0: $\beta_3 = \beta_4 = 0$ ($\alpha = 0.05$).

f) Hoe zou U de resultaten van deze analyse overzichtelijk kunnen samenvatten?

2.17. Twee rattendiëten, A en B, bevatten elk eenzelfde hoeveelheid gegeven chemicalie. Men weet dat de opname van dit chemisch product in het organisme van een rat lineair afhangt van het gewicht van de rat. Beide diëten werden door verloting aan vier ratten toegewezen en het begingewicht van iedere rat werd bepaald.

(Het begingewicht is een medeveranderlijke (engels: covariate) naast de dieetveranderlijke. De vergelijking van de diëten rekening houdend met de medeveranderlijke heet een covariantie-analyse.)

Na verloop van tijd werd het gewicht van de ratten opnieuw gemeten.

De gegevens zijn

Begingewicht	12	10	13	16	15	11	14	13
Dieet	A	A	A	A	B	B	B	B
Responsie	14.1	12.5	14.1	17.3	14.8	10.0	15.1	13.4

a) Geef het model (codeer de diëten op (-1,1)).

b) Pas het model aan bij de waarnemingen.

c) Bepaal een 95%-betrouwbaarheidsinterval voor β_2 als x_2 de veranderlijke is die het verschil in de diëten aangeeft. Welke conclusie kunt U hieruit trekken t.a.v. de nulhypothese $\beta_2 = 0$?

d) Toets de hypothese dat het begingewicht geen invloed heeft op de responsie ($\alpha = 0.05$).

e) Geef een 90%-betrouwbaarheidsinterval voor de verwachte responsie bij een rat met begingewicht 12 en dieet B.

f) Stel dat we geen rekening houden met het begingewicht. Is er nu een verschil in responsie tussen dieet A en dieet B ($\alpha = 0.05$)? Geef ook het model.

g) Het is ook denkbaar dat iemand U als waarnemingen de gewichtstoenamen overhandigt, maar verder niets weet van de lineaire afhankelijkheid tussen opnamecapaciteit en gewicht. Welke conclusie trekt U nu t.a.v. het verschil tussen dieet A en dieet B ($\alpha = 0.05$)?

h) Vergelijk de verschillende resultaten en geef Uw commentaar.

2.18. a) Pas een lineair model

$$\underline{y} = \alpha_0 + \alpha_1 x + \underline{e}$$

aan bij de volgende waarnemingen

x	1	2	3	4	5
y	3	2	4	6	5

Geef schattingen voor de parameters α_0 en α_1 en voor de restvariantie σ_0^2.

b) Doe hetzelfde voor het model

$$\underline{y} = \beta_0 + \beta_1 x + \underline{e}$$

bij de waarnemingen

x	1	2	3	4	5
y	4	8	9	10	14

c) Neem aan dat de toevallige fouten $\underline{e}$ in beide gevallen dezelfde normale verdeling hebben.
Toets de hypothese H_0: $\alpha_1 = \beta_1$ ($\alpha = 0{,}05$).

2.19. Er zijn wasproeven verricht met 3 doseringen van een wasmiddel en bij 3 temperaturen. De resultaten zijn in de volgende tabel samengevat (het effect y is gemeten in een maat die verder niet ter zake doet). Twee proeven zijn mislukt.

		Temperatuur		
		70°	80°	90°
Dosering	35g	56	60	-
	40g	59	60	61
	45g	-	60	64

a) Pas een lineair model aan van de vorm

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underline{e} ,$$

waarin x_1 met de temperatuur en x_2 met de hoeveelheid wasmiddel overeenkomt.

Denk aan een verstandige codering!

b) Toets de volgende hypothesen (onbetrouwbaarheid $\alpha = 0,05$):

i) $H_0 : \beta_2 = 0$

ii) $H_0 : \beta_1 = \beta_2 = 0$.

2.20. In een landbouwkundige proef werd het effect onderzocht van de rijafstand en van de bemesting op de opbrengst van suikerbieten.
De rijafstand d werd op 30, 40 en 50 cm. genomen en bemest werd met de meststoffen A,B en C.
De opbrengsten waren:

		Rijafstand (cm) 30	40	50
	A	23	24	22
Meststof	B	18	20	25
	C	22	25	28

Het model is:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \underline{e},$$

waarin $x_1 = \frac{d - 40}{10}$

$$\begin{cases} x_2 = 1 \text{ als meststof B werd toegepast,} \\ x_2 = 0 \text{ anders,} \end{cases}$$

$$\begin{cases} x_3 = 1 \text{ als meststof C werd toegepast,} \\ x_3 = 0 \text{ anders.} \end{cases}$$

a) Bereken schattingen voor de parameters en voor var $\underline{e} =: \sigma_0^2$.

b) Toets de hypothese dat de meststoffen A en B dezelfde invloed hebben op de opbrengst.

c) Idem voor B en C.

2.21. Gegeven het lineaire model

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \underline{e}.$$

Om de parameters te schatten beschouwen we twee proefopzetten, I en II, waarvan de waarnemingen in de volgende tabellen zijn weergegeven

x_2 \ x_1	-1	0	1
-1	y_1	y_2	y_3
0	y_4	y_5	y_6
1	y_7	y_8	y_9

Proefopzet I

x_2 \ x_1	-1	0	1
-1	y_1, y_2		y_3, y_4
0		y_5	
1	y_6, y_7		y_8, y_9

Proefopzet II

Aangenomen mag worden dat in beide gevallen de fouten $\underline{e}$ onafhankelijk zijn en dezelfde variantie σ_0^2 hebben.

Vraag a: Welke van de twee proefopzetten verdient de voorkeur op grond van het criterium dat de som van de varianties van de schatters van de parameters:

$$(\text{var } \underline{b}_0 + \text{var } \underline{b}_1 + \text{var } \underline{b}_2 + \text{var } \underline{b}_3)$$

zo klein mogelijk moet zijn ?

Vraag b: Een van beide proefopzetten is niet geschikt om een uitgebreider model:

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \underline{e}$$

aan te passen. Welke proefopzet is dit? Motiveer Uw antwoord.

3. Latijnse vierkanten

3.1. In een electronisch systeem moest de spanning worden gemeten onder 4 verschillende omstandigheden A, B, C en D. De metingen werden echter beïnvloed door de netspanning in het laboratorium. Omdat werd aangenomen dat de netspanning gedurende iedere dag hetzelfde patroon volgde (door het verbruik van nabijgelegen industrieën) en omdat de gemiddelde netspanning van dag tot dag varieerde, werd een Latijns vierkant toegepast.

De gegevens (in Volts) waren:

perioden gedurende een dag	Dagen			
	1	2	3	4
1	A 104	B 108	C 126	D 112
2	C 111	D 124	A 110	B 121
3	B 120	C 115	D 126	A 97
4	D 118	A 113	B 116	C 127

a) Geef het meest volledige model bij deze proefopzet.

b) Pas variantie-analyse toe op deze gegevens (codeer eerst!).

c) Toets of er verschillen zijn tussen de omstandigheden en geef Uw conclusies overzichtelijk weer.

d) Indien gevraagd wordt of de drie factoren samen (dagen, perioden, omstandigheden) van invloed zijn, formuleer dan H_0, toets H_0 en geef het aangepaste model.

3.2. We willen de slijtage van 4 typen (A, B, C en D) autobanden vergelijken d.m.v. een rijtest. Hiervoor nemen we 16 banden (van elke soort 4) en 4 auto's van verschillend type om een grotere geldigheid voor onze conclusies te verkrijgen.

a) Hoe kan men het beste de 16 banden over de 4 auto's verdelen zodat het "auto-effect" geen systematische invloed op de conclusies heeft. Schrijf ook het (lineaire) model voor de waarnemingsresultaten op.

b) Omdat we echter vermoeden dat de bandenslijtage niet alleen per auto verschilt maar ook van de wielpositie afhangt, gebruiken we een Latijns vierkant als proefopzet. Aan welke voorwaarde moet dan worden voldaan? Geef het bijpassende volledige model.

c) We trekken een (willekeurig) 4 × 4 Latijns vierkant, monteren volgens de banden op de wijze voorgeschreven door het vierkant en na de rijtest krijgen we de volgende resultaten:

auto / wielpositie	1	2	3	4
links voor	C : 4	A : 5	D : 6	B : 2
rechts voor	B : 2	D : 7	A : 4	C : 4
links achter	A : 3	C : 8	B : 6	D : 5
rechts achter	D : 7	B : 4	C : 9	A : 3

Toets de hypothese dat er geen verschil tussen de 4 bandensoorten bestaat. Zijn er ook redenen aan te nemen dat er "auto-effect" en/of "wielpositie-effecten" zijn?

3.3. Om gaten te maken in stripmateriaal doet men een electrische ontlading plaats vinden tussen een electrode en het materiaal.
Men wil meten met welke van 5 verschillende vormen (A, B, C, D en E) van electroden het snelst gaten kunnen worden gemaakt.
Om deze vraag te kunnen beantwoorden worden in 5 verschillende strips 5 gaten gemaakt, zodanig dat bij dezelfde positie op ieder van de 5 strips slechts één electrode vorm wordt gebruikt.
We hebben dus te maken met een Latijns vierkant, waarbij de strips en de posities op de strips de blokken vormen.
De resultaten van de proef zijn gegeven in onderstaande tabel.

	positie				
strip	1	2	3	4	5
1	A 3.5	B 2.1	C 2.5	D 3.5	E 2.4
2	E 2.6	A 3.3	B 2.1	C 2.5	D 2.7
3	D 2.9	E 2.6	A 3.5	B 2.7	C 2.9
4	C 2.5	D 2.9	E 3.0	A 3.3	B 2.3
5	B 2.1	C 2.3	D 3.7	E 3.2	A 3.5

a) Geef de variantie-analyse en toets met een betrouwbaarheid van 95% welke effecten significant zijn.

b) Als proefopzet is hier gekozen voor een Latijns vierkant. Welke veronderstellingen moet men maken om deze keuze te rechtvaardigen?

3.4. Vier tarwerassen, aangeduid met A, B, C, D en vier meststoffen α, β, γ, δ zijn met elkaar vergeleken op 16 proefvelden. De proefopzet bestond uit het volgende Grieks-Latijns vierkant

Kolommen

Rijen				
	Aα	Bβ	Cγ	Dδ
	Bγ	Aδ	Dα	Cβ
	Cδ	Dγ	Aβ	Bα
	Dβ	Cα	Bδ	Aγ

De opbrengsten waren als volgt:

26	20	18	24
14	20	20	18
10	18	24	20
22	18	22	26

Toets met onbetrouwbaarheid van 0,05 welke effecten significant zijn.

3.5. In een electronisch systeem moest de spanning worden gemeten onder 4 verschillende omstandigheden A, B, C en D. De metingen werden echter onafhankelijk van de omstandigheden beïnvloed door de netspanning.
Aangenomen werd dat de netspanning gedurende iedere dag hetzelfde patroon volgde wegens het verbruik van nabij gelegen industrieën; de gemiddelde netspanning varieerde bovendien van dag tot dag.
Er werd daarom gekozen voor een Latijns vierkant: De resultaten in Volts waren de volgende:

perioden gedurende een dag	Dagen 1	2	3	4
1	A 120	B 122	C 123	D 123
2	B 121	A 125	D 120	C 126
3	C 126	D 124	A 123	B 123
4	D 121	C 125	B 122	A 124

a) Geef het meest volledige model dat bij dit schema mogelijk is.
b) Voer een variantie-analyse uit.
c) Wat zijn Uw conclusies?

3.6. De kwaliteit van een beschermingslaag hangt af van 4 factoren: vernisconcentratie, acryloidconcentratie, acetonconcentratie en dompeltijd. Om de invloed vast te stellen werd een Grieks-Latijns vierkant toegepast met de volgende resultaten.

	vernisconcentratie			
dompeltijd	½	1	1½	2
30	Cβ(10)	Bγ(6)	Dδ(11)	Aα(5)
20	Bα(8)	Cδ(8)	Aγ(10)	Dβ(8)
10	Aδ(6)	Dα(1)	Bβ(8)	Cγ(7)
5	Dγ(4)	Aβ(3)	Cα(2)	Bδ(3)

A, B, C en D zijn de 4 niveau's van de concentratie acryloid, α, β, γ en δ van de concentratie aceton.
Gebruiken we voor de factoren tijd, vernis, aceton en acryloid resp. de indices i, j, k en ℓ dan geldt:

$$\Sigma x^2_{ijk\ell} = 762 \qquad \Sigma x^2_{\cdot\cdot k\cdot}/4 = 652\tfrac{1}{2}$$
$$\Sigma x^2_{i\cdot\cdot\cdot}/4 = 702 \qquad \Sigma x^2_{\cdot\cdot\cdot\ell}/4 = 626\tfrac{1}{2}$$
$$\Sigma x^2_{\cdot j\cdot\cdot}/4 = 649\tfrac{1}{2} \qquad \Sigma x^2_{\cdot\cdot\cdot\cdot}/16 = 625 .$$

a) Analyseer deze gegevens en toets welke effecten significant zijn ($\alpha = 0.05$).

b) Aan welke onderstellingen moet voldaan zijn voor het toepassen van een Grieks-Latijns vierkant?

c) Aan welke onderstelling moet voldaan zijn voor de onder a) gebruikte toets?

d) Waarom is een 3 × 3-Grieks-Latijns vierkant een weinig geschikte proefopzet?

4. Factoriële proefopzetten

4.1. Voor de bestudering van graasgewoonten van vee zijn er waarnemers nodig. Om na te gaan of er grote verschillen zijn tussen waarnemers werd het volgende experiment opgezet. Op een middag werden in een afgesloten wei 10 stuks vee gezet. Aan 5 waarnemers werd gevraagd de dieren gedurende anderhalf uur te observeren en te noteren hoeveel minuten elk dier aan het grazen was.
Om het rekenwerk te beperken volgt hieronder slechts een gedeelte van de waarnemingsuitkomsten.

waarnemer	dier			
	1	2	3	4
A	34	76	75	31
B	33	76	72	29
C	35	78	76	30

a) Hoe luidt het bijbehorende model?
b) Toets met een onbetrouwbaarheid van 5% of er verschillen zijn tussen de waarnemers.

4.2. Er werd een experiment uitgevoerd om het effect van 3 kwaliteiten fosfor en 2 typen glas op de lichtopbrengst (in μA) van een TV-buis te bepalen. Per niveaucombinatie werden 3 waarnemingen verricht om σ_0^2 te kunnen schatten. Bovendien werd volledige verloting toegepast. De gecodeerde resultaten waren:

		fosfor		
		A	B	C
Glas	I	4	8	2
		6	10	5
		5	7	6
	II	-6	0	-8
		-5	-4	-7
		-4	-5	-6

a) Geef de variantie-analyse-tabel.
b) Toets welke effecten significant zijn.

4.3. Een onderzoeker verzamelde onderstaande gegevens over gemiddelde uurlonen die in verschillende bedrijfstakken in verschillende steden van de Verenigde Staten werden uitbetaald.

Stad	Handel	Industrie	Dienstensector
	4.17	8.92	4.51
San Francisco	2.87	9.85	6.50
	3.93	7.51	6.55
	4.12	8.61	5.66
	5.99	9.23	6.20
Chicago	6.38	9.84	5.20
	3.72	8.69	3.75
	3.85	9.78	4.16
	9.65	6.96	4.59
New York	9.39	6.54	5.38
	9.02	6.57	5.68
	8.90	5.90	5.42

Analyseer deze gegevens en geef de resultaten en conclusies overzichtelijk weer.

4.4. Er is een experiment uitgevoerd om schimmelgroei op houten afwerkingen van buitenmuren van huizen te vergelijken voor drie soorten verf (V_1, V_2, V_3), twee soorten zonbelichting (Z_1, Z_2; denk hierbij aan noord- en zuidmuur) en op drie verschillende plaatsen (P_1, P_2, P_3) in de Amerikaanse staat Florida. Voor ieder der niveaucombinaties der factoren werden 4 experimentele oppervlakken genomen waarop de schimmelgroei na een jaar werd gemeten.
De waarnemingen zijn:

	Z_1			Z_2		
	V_1	V_2	V_3	V_1	V_2	V_3
P_1	8	6	5	14	12	9
	7	6	8	12	11	5
	9	7	7	14	11	8
	8	8	7	15	11	7
P_2	6	5	1	14	10	7
	7	7	2	15	10	6
	7	7	2	13	10	7
	8	6	1	12	9	8
P_3	10	10	7	23	20	10
	11	10	5	22	21	8
	13	12	4	20	21	9
	11	11	4	21	19	8

a) Analyseer deze gegevens en geef Uw conclusies samenvattend weer.

b) Geef een schatter van het verschil in gemiddelde schimmelgroei van twee plaats-verf-combinaties en bepaal de variantie van deze schatter (de twee combinaties zijn willekeurig te kiezen).

c) Veronderstel nu, dat de waarnemingen zijn uitgevoerd in blokken. Van boven naar beneden gerekend behoren de waarnemingen van iedere niveaucombinatie van Z, P en V bij blok 1 t/m blok 4. Is het blokeffect significant?

4.5. Er werd een experiment uitgevoerd om te bepalen hoe de levensduur van snijgereedschap afhangt van metaalsoort en snijsnelheid. De waarnemingen werden volledig verloot. De resultaten van het experiment staan in de volgende tabel (levensduur in eenheden van honderd uren):

	snijsnelheid		
	laag	matig	hoog
metaal I	9 8 10	14 15 13	14 12 10
metaal II	13 12 11	17 15 16	12 14 13

a) Hoe luidt het volledige model?

b) Geef de variantie-analyse tabel.

c) Toets welke effecten significant zijn.

d) Geeft kort en overzichtelijk het resultaat van de analyse.

4.6. Er worden 2 soorten lijm voor glasverbindingen beproefd. Daarbij zijn 3 methoden te onderscheiden. Omdat er wellicht een interactie bestaat tussen de methode van lijmen en de lijmsoort worden herhalingen uitgevoerd.
De gecodeerde resultaten van de sterkteproeven zijn:

		Methode		
		1	2	3
Soort lijm	047	3 2	5 9	2 6
	00 T	10 5	11 7	10 8

a) Hoe luidt het model voor deze proefopzet? (Inclusief onderstellingen!)

b) Voer een variantie-analyse uit.

c) Toets welke effecten significant zijn ($\alpha = 0.05$).

5. Codering en orthogonale polynomen

5.1. Van 3 verschillende typen radiobuizen werd de opwarmtijd T van de kathode bepaald in seconden.
De volgorde der waarnemingen is door loting bepaald.
De resultaten van de metingen waren:

Type	T in sec.
A	26, 19, 23, 35
B	40, 24, 20
C	16, 18, 19, 17, 19, 26

a) Geef 2 onderling orthogonale contrasten voor de verschillen tussen de 3 typen.

b) Bereken de 2 bijbehorende kwadratensommen.

c) Bereken uit de 3 series waarnemingen een schatting voor σ_0^2 met bijbehorend aantal vrijheidsgraden.

d) Bereken de totale kwadratensom en de kwadratensom voor de typen buizen (rijen). Vergelijk het resultaat met de uitkomsten bij b).

e) Geef de variantie-analyse en toets of er verschil is tussen de 3 typen buizen.

5.2. a) Pas met behulp van orthogonale polynomen een geschikt lineair model aan bij de volgende serie waarnemingen:

$$x_i = 25;\ 30;\ 35;\ 40;\ 45;\ 50;\ 55;\ 60$$

$$y_i = 1.20;\ 1.26;\ 1.30;\ 1.30;\ 1.32;\ 1.36;\ 1.34;\ 1.28\ .$$

b) Geef een 95%-betrouwbaarheidsinterval voor de restvariantie σ_0^2.

5.3. Voor verschillende snelheden is de gemiddelde afstand bepaald waarbinnen een auto tot stilstand komt bij plotseling hard remmen:

x = snelheid (mijlen/uur)	20	30	40	50	60	70
y = afstand (voet)	54	90	138	206	292	396

a) Bepaal m.b.v. SC-tabel 9.6 een polynoom $y = f(1, x, .., x^n)$ van zo laag mogelijke graad, dat in statistische zin voldoende goed past bij deze waarnemingen (zie voorbeeld polynoomaanpassing, SC pag. 23).

b) Voor welke waarde(n) van x is var $\hat{\underline{y}}$ minimaal?

5.4. Er werd een 2 * 3-factoriële proefopzet uitgevoerd om de invloed van temperatuur en druk te bepalen op de opbrengst van een chemikalie. De druk werd ingesteld op 50 resp. 80 eenheden, de temperatuur op 100, 200 en 300 °F. Het is bekend dat er geen interactie is tussen deze twee factoren. De opbrengsten waren:

		Temperatuur (°F)		
		100	200	300
Druk	50	21	23	26
	80	22	23	28

a) Geef het meest uitgebreide lineaire model voor deze proefopzet met bijbehorende coderingen en onderstellingen.

b) Bereken schattingen voor de parameters en voor σ_0^2.

c) Ga na welke parameters op statistische gronden uit het model verwijderd kunnen worden ($\alpha = 0.05$).
Hoe luidt nu het gereduceerde aangepaste model?

d) Geef een 95%- betrouwbaarheidsinterval voor de opbrengst bij een druk van 60 eenheden en een temperatuur van 150 °F. Ga daarbij uit van het gereduceerde model.

5.5. Om het effect van de concentratie van een katalysator (C) en van de temperatuur (T) op de conversie van veresteringsproces te onderzoeken werd de volgende proef uitgevoerd. De temperatuur werd ingesteld op 180, 200 en 220 graden Celcius. De concentratie werd gekozen op 4 niveau's: 0.5, 0.6, 0.7 en 0.8 mmol/gr.

Beschouwd wordt het volgende model:

$$\mathcal{E}\underline{y} = \alpha_0 + \alpha_1 C + \alpha_2 T + \alpha_3 CT + \alpha_4 C^2 .$$

Om tijd en kosten te sparen werd geen volledige proef uitgevoerd. De (gecodeerde) waarnemingen zijn:

		Temperatuur in °C		
		180	200	220
	0.5	3		-5
concentratie	0.6		8;6	
in mmol/gr.	0.7		10;8	
	0.8	11		39

a) Kies met behulp van orthogonale polynomen een codering van x_1, x_2, x_3 en x_4 voor het aequivalente model:

$$\mathcal{E}\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

zodanig dat X'X een diagonale matrix wordt.

b) Bepaal $\hat{\beta}$ en $\hat{\sigma}_0^2$.

c) Uit andere proeven is een schatting bekend van de toevallige fout bij het bepalen van de conversie: $\hat{\sigma}_0^2 = 13$, $\nu = 7$.
Bereken uit de beide schattingen een nieuwe schatting voor σ_0^2 en toets de hypothese dat de concentratie geen kwadratisch verband heeft met de conversie ($\alpha = 0.05$).

d) Bepaal, uitgaande van het eventueel gereduceerde model een 95%-betrouwbaarheidsinterval voor de conversie bij een temperatuur van 220°C en een concentratie van 0,7 mmol/gr.

5.6. Onderzocht werd het effect van de concentratie van een katalysator en van de temperatuur op de conversie van een veresteringsproces.
Bij elke niveaucombinatie werden 2 bepalingen verricht. De gecodeerde waarnemingen zijn:

		Temperatuur in °C		
		175	200	225
Concentratie in 10^{-4} mol/gr	4	2; 0	16; 20	25; 27
	8	6; 8	19; 21	28; 32

Beschouwd wordt het model

$$\underline{y} = \alpha_0 + \alpha_1 C + \alpha_2 T + \alpha_3 T^2 + \underline{e} .$$

a) Bepaal voor het equivalente model

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \underline{e}$$

het verband tussen C, T resp. x_1, x_2 en x_3, als gegeven is dat de contrasten x_1, x_2 en x_3 onderling orthogonaal zijn.

b) Bepaal schattingen van β en σ_0^2.

c) Toets met een onbetrouwbaarheid van 5% de hypothesen $\beta_1 = 0$, $\beta_2 = 0$ en $\beta_3 = 0$.

d) Geef een 95%-betrouwbaarheidsinterval voor de verwachte responsie bij een concentratie van 5×10^{-4} mol/gr en een temperatuur van 200°C.

6. 2^n-proefopzetten

6.1. Gegeven zijn in verlote volgorde de uitkomsten van een 2^3-experiment:

a	22	ac	30
(1)	12	bc	24
abc	40	b	18
ab	30	c	20

a) Bereken met behulp van de algoritme van Yates de effecten en kwadratensommen.

b) Toets welke effecten significant zijn en geef een schatting van de restvariantie als aangenomen mag worden dat interacties verwaarloosbaar zijn.

c) Geef een 95%-betrouwbaarheidsinterval van de verwachte responsie in de situatie dat A en B op het hoge niveau zijn en C op het lage niveau is.

6.2. We beschouwen nog eens het 2^4-factorieel experiment van opgave 2.8.
De waarnemingen waren

				T_2			
				100°		200°	
				K		K	
				1	2	1	2
T_1	50°	P	10	22.2	24.5	24.4	25.9
			20	19.4	24.1	25.2	18.4
	70°	P	10	22.1	19.6	23.5	16.5
			20	14.2	12.7	19.3	16.0

a) Analyseer deze gegevens m.b.v. het algoritme van Yates.

b) We nemen aan dat interacties van 3 of meer factoren te verwaarlozen zijn. Toets welke effecten (zowel hoofdeffecten als interacties) significant zijn.

c) Vergelijk de resultaten met die van opgave 2.8.

6.3. Gegeven zijn de resultaten van een 2^3-experiment met 2 herhalingen in blokken. Binnen elk blok zijn de waarnemingen verloot.

	blok I	blok II
(1)	39	40
a	27	29
b	37	39
ab	27	30
c	45	47
ac	30	34
bc	45	47
abc	33	35

Interacties van 3 en meer factoren zijn verwaarloosbaar.

a) Hoe luidt het meest uitgebreide lineaire model?

b) Bereken de effecten met bijbehorende kwadratensommen.

c) Toets welke effecten significant zijn.

d) Hoe luidt het gereduceerde aangepaste model?

e) Bereken de schatting van de variantie van de geschatte parameters.

6.4. In een factorieel 3-factoren experiment zijn 16 waarnemingen uitgevoerd. De resultaten zijn:

niveaucombinatie	H_1	H_2
(1)	74	78
a	111	118
b	89	86
ab	101	129
c	78	100
ac	116	138
bc	69	86
abc	98	131

a) Voer een Yates analyse uit voor het geval elk tweetal bepalingen bij dezelfde niveaucombinaties van A,B,C als duplo's zijn op te vatten.

b) Idem voor het geval de herhaling in 2 blokken heeft plaatsgevonden.

c) Idem voor het geval H_1 en H_2 twee niveaus vertegenwoordigen van een factor H, die ook interacties met A, B en C kan hebben.

d) Vergelijk de drie analysen. Welk model lijkt U juist? Waarom?

e) Wat is nu de schatting van de restvariantie?
Welke effecten zijn significant?

f) Geef het aangepaste model; geef de schatting voor de variantie van $\hat{\underline{\beta}}_A$.

6.5. Een 2^3-factorieel experiment werd tweemaal in zijn geheel uitgevoerd (herhaling in 2 blokken). Analyse van de resultaten leverde het volgende aangepaste model

$$\hat{y} = 2 + 3x_1 + 0.3\,x_2 + 4.1\,x_3 + 0.3\,x_1x_2 + 2.3\,x_2x_3 + x_1x_3 + 0.5\,x_4\,,$$

waarbij x_1, x_2 en x_3 de niveaus van de 3 factoren (gecodeerd op -1, $+1$) weergeven. Verder is $x_4 = -1$ voor het eerste experiment en $x_4 = +1$ voor het tweede.

a) Hoeveel vrijheidsgraden zijn beschikbaar voor het schatten van σ_0^2 ?

b) Is er aanleiding om aan te nemen dat er een x_1x_2-interactie is, als $\hat{\sigma}_0^2 = 1$? (Toets met onbetrouwbaarheid van 5%.)

c) Stel een 95%-betrouwbaarheidsinterval op voor β_1 (i.e. de coëfficiënt van x_1).

d) Als het 2^3-experiment wordt uitgevoerd in k blokken, dan coderen we de k-1 contrasten voor blokken zodanig dat deze orthogonaal zijn t.o.v. de anderen.
   i) Hoeveel vrijheidsgraden zijn nu beschikbaar voor de schatting van σ_0^2 ?
   ii) Als $\hat{\sigma}^2 = 1$, hoe groot moet k dan minstens zijn opdat β_1 geschat kan worden met een nauwkeurigheid van $\pm$ 0.2 of minder?
   (Nauwkeurigheid = $t_{\nu(\frac{1}{2}\alpha)}s_{\hat{\beta}1}$, $\alpha = 0.05$.)

6.6. Een 2^5-factorieel experiment wordt uitgevoerd in 4 blokken van 8 experimentele eenheden. Twee van de met blokken verstrengelde contrasten zijn gegeven, nl. ABC en ADE.

a) Geef aan hoe de waarnemingen over de 4 blokken zijn verdeeld.

b) Hoe wordt de kwadratensom voor blokken berekend en hoeveel vrijheidsgraden behoren daarbij?

c) Als interacties van 3 en meer factoren te verwaarlozen zijn, hoeveel vrijheidsgraden zijn dan beschikbaar voor de restvariantie?

6.7. Een 2^5-experiment wordt uitgevoerd in 4 blokken van 8 niveaucombinaties. De met blokken verstrengelde contrasten zijn ABD, ACE en BCDE.

a) Hoe zijn de waarnemingen over de blokken verdeeld?

b) Geef voor elk van de contrasten ABD, ACE en BCDE aan met welk orthogonaal contrast tussen blokken zij verstrengeld zijn.

c) Hoeveel vrijheidsgraden zijn beschikbaar voor het schatten van hoofdeffecten, interacties, blokeffect en restvariantie, als interacties van 3 en meer factoren verwaarloosbaar zijn?

6.8. Onderstaande tabel geeft het resultaat van een ¼ fractie van een 2^6 proefopzet, in de verlote volgorde waarin de waarnemingen werden uitgevoerd.

Waarn. nr.	Comb.	Y	Waarn. nr.	Comb.	Y
1	df	78	9	(1)	74
2	abcdef	129	10	bde	86
3	cde	110	11	ae	111
4	adef	118	12	bc	69
5	cef	78	13	abd	131
6	acf	116	14	bcdf	86
7	acd	138	15	abf	101
8	bef	89	16	abce	98

a) Wat zijn de definiërende contrasten?

b) Analyseer deze gegevens, ervan uitgaande dat alle interacties te verwaarlozen zijn.

c) Vat het resultaat van deze proef kort samen.

6.9. a) Geef aan hoe de waarnemingen in 2 blokken zijn gedeeld voor een $\frac{1}{4}$ fractie van een 2^5-experiment met als definiërende contrasten I = ABC = - CDE = = - ABDE, terwijl ABCDE met blokken is verstrengeld.

b) Hieronder volgen de gegevens van een halve fractie van een 2^5-experiment in verlote volgorde:

(1)	11	ac	11	acd	18	abce	14
cde	15	d	19	ce	17	ab	17
ae	14	abd	19	bcd	18	ade	14
bc	20	be	21	abcde	18	bde	20

(i) Bepaal het definiërende contrast.

(ii) Toets welke effecten significant zijn, indien interacties van 3 en meer factoren te verwaarlozen zijn. Geef daarbij aan welke effecten verstrengeld zijn.

6.10. Voor een 2^{7-3}-proefopzet zijn de definiërende contrasten:

I = ABEG = ACFG = BCEF = ABCD = CDEG = BDFG = ADEF .

Bovendien is EG verstrengeld met blokken.

a) Bij welke niveaucombinaties moeten waarnemingen worden gedaan en hoe zijn de niveaucombinaties verdeeld over de twee blokken?

b) Hoeveel 2-factoren interacties kunnen worden bepaald? (Een effect kan worden bepaald als het alleen verstrengeld is met interacties van 3 of meer factoren.)

6.11. De uitkomsten van een fractionele 2^4-proef uitgevoerd in twee blokken zijn:

Blok 1		Blok 2	
cd	15.8	bd	8.9
(1)	9.9	ad	12.1
abcd	13.5	ac	15.6
ab	7.9	bc	13.0

a) Wat is het definiërend contrast?

b) Welk effect is verstrengeld met het blokeffect?

c) Zet de niveaucombinaties in een standaardvolgorde en voer een Yates-analyse uit.
Geef daarbij aan welke effecten verstrengeld zijn.

d) Toets welke van de hoofdeffecten significant zijn ($\alpha = 0.05$).
Neem daarbij aan dat er geen interacties zijn.

6.12. Bij een 2^{5-2}-experiment werden de volgende (gecodeerde) waarnemingen gevonden.

(1)	= −2	de	= −4
ab	= 18	abde	= 24
bcd	= 10	acd	= 2
ace	= 4	bce	= 12 .

Interacties tussen de factoren zijn te verwaarlozen.

a) Wat zijn de definiërende contrasten?

b) Voer een Yates-analyse uit.

c) Ga na welke hoofdeffecten significant zijn ($\alpha = 0.05$).

6.13. Een proef werd opgezet om de invloed van een drietal factoren op de vereiste kracht (in p.k's) bij het draaien van een werkstuk te bepalen. De factoren waren: T = type beitel (2 types), M = manier van draaien (2 manieren: continu en onderbroken) en H = hoek van de beiteltop met het werkstuk (2 hoeken: 15° en 30°). Men besloot 3 metingen te doen bij iedere niveaucombinatie en het experiment werd volledig verloot.

De resultaten waren: (reeds gecodeerd)

		T = type beitel			
		1		2	
		H = hoek beitel		hoek	
		15°	30°	15°	30°
M = manier van draaien	Continu	5	4	4	6
		2	4	4	8
		6	6	4	5
	Onderbroken	4	3	0	4
		1	5	1	4
		2	4	4	3

a) Voer de Yates-analyse uit.

b) Welk(e) effect(en) is (zijn) significant? ($\alpha = 0.05$)

c) Geef het aangepaste model (met weglating van eventuele niet-significante effecten).

d) Wat is de geschatte responsie bij continu draaien onder een hoek van 25°?

6.14. In zekere situatie (5 factoren, 2 niveaus) wenst men een experiment in 4 blokken à 4 waarnemingen uit te voeren. De 3e orde effecten (ABC etc.) kunnen bij voorbaat afwezig worden geacht. Ten aanzien van de keuze der met contrasten tussen blokken te verstrengelen effecten bestaat geen voorkeur. Als zodanig zijn aangewezen AB en AC.

De waarnemingen zijn:

blok 1		blok 2		blok 3		blok 4		
(1)	45	abde	71	acde	85	bc	51	
abcd	85	ce	59	be	45	ad	47	
abce	73	cd	59	bd	41	ae	63	
de	45	ab	67	ac	69	bcde	55	
$\Sigma \sim = 248$		$\Sigma \sim = 256$		$\Sigma \sim = 240$		$\Sigma \sim = 216$		$\Sigma \sim = 960$

a) Welk(e) definiërend(e) contrast(en) is (zijn) gekozen?

b) Laat zien dat bij het in deze opzet resulterende model geen restkwadratensom ter beschikking staat.
Uit vroegere waarnemingen is echter een schatting van $\underline{GK}_{residu}$ bekend, nl. $\hat{\sigma}_0^2 = 15$ met $\nu = 8$.

c) Voer vervolgens het algoritme van Yates uit en beoordeel de diverse effecten op hun significantie.

d) Bereken rechtstreeks de KS die behoort bij verschillen tussen blokken en vergelijk deze KS met de juiste combinatie van KS'en, verkregen via het Yates-algoritme.

e) Geef het aangepaste model met de bijbehorende parameterschattingen.

6.15. De (gecodeerde) waarnemingen van een 2^4-experiment in 4 blokken zijn:

blok 1		blok 2		blok 3		blok 4	
(1)	2	d	2	b	1	bd	1
cd	8	c	7	bcd	6	bc	5
ab	0	abd	4	a	4	ad	4
abcd	6	abc	4	acd	10	ac	8

a) Bereken met behulp van het Yates-algorithme de effecten en kwadratensommen.

b) Welke effecten zijn verstrengeld met het blokeffect.

c) Het is bekend dat de met het blokeffect verstrengelde effecten en bovendien alle interacties van 3 en meer factoren verwaarloosbaar klein zijn. Geef de variantie-analysetabel en toets welke effecten significant zijn ($\alpha = 0.05$).

6.16. De meetresultaten van de helft van een 2^4-experiment zijn

bcd	4.9	acd	5.2
c	5.0	abd	4.7
d	4.8	a	5.0
b	4.8	abc	5.4

a) Bereken met behulp van het Yates-algorithme de effecten en bijbehorende kwadratensommen.
Geef daarbij aan hoe de effecten verstrengeld zijn.

b) Uit vroegere metingen is een schatting bekend van σ_0^2:

$$\hat{\sigma}_0^2 = 0.01 \text{ met } \nu = 10 .$$

Toets welke effecten significant zijn (α = 0,05).

c) Als alleen rekening wordt gehouden met de significante effecten, wat zijn dan de schattingen van de hoogste en van de laagste responsie?

6.17. Van een 2^6-proefopzet is een kwart uitgevoerd in twee blokken. Het betrof de volgende niveau-combinaties:

blok 1	blok 2
(1)	ab
bdf	adf
acf	bcf
abcd	cd
ade	bde
cdef	ace
bce	ef
abef	abcdef

a) Wat zijn de definiërende contrasten en welk contrast is verstrengeld met blokken?

b) Zet de waarnemingen in een standaardvolgorde. Geef vervolgens hierbij de groepen van verstrengelde effecten in de juiste volgorde.

c) Hoeveel vrijheidsgraden zijn beschikbaar voor hoofdeffecten, interacties, blokeffect en restvariantie?
(Opmerking: Interacties van 3 en meer factoren worden verondersteld van geen betekenis te zijn.)

6.18. Op 3 achtereenvolgende dagen is een halve fractie van een 2^4-proefopzet uitgevoerd. Elke dag is de volgorde der waarnemingen opnieuw door loting vastgesteld.

De resultaten van de proef waren:

niveau- combinatie	1^e^ dag	2^e^ dag	3^e^ dag
a	6	11	7
abd	5	10	15
acd	-1	2	-1
d	-2	0	2
b	18	24	12
c	9	10	5
bcd	11	9	10
abc	10	6	14
	56	72	64

$$\sum_{i=1}^{24} x_i^2 = 2454.$$

a) Wat is het definiërend contrast?

b) Ga na welke effecten significant zijn ($\alpha = 0{,}05$).

c) Geef, uitgaande van het gereduceerde model, een 95%-betrouwbaarheidsinterval voor de responsie bij niveaucombinatie bd.
(Gebruik de variantieschatting die verkregen wordt na samenvoegen van de niet-significante effecten).

6.19. Een proef met 6 factoren, ieder op 2 niveau's, wenst men te beperken tot 16 waarnemingen.

a) Hoe moet de proef worden ingericht, opdat geen enkel hoofdeffect met een 2-factor interactie samenvalt?

b) Geef een groep van 16 combinaties voor een dergelijk experiment in een standaardvolgorde en geef hierbij ook de juiste volgorde aan van de groepen verstrengelde effecten.

6.20. i) Een fractionele 2^6-proef wordt uitgevoerd in blokken.
De definiërende contrasten zijn: I = -ABCD = - CDEF = ABEF.
Het contrast ADF is met blokken verstrengeld.
Bij welke niveaucombinaties moeten waarnemingen worden gedaan en hoe zijn deze waarnemingen over de blokken verdeeld?

ii) Om het effect van een kwalitatieve factor A en twee kwantitatieve factoren B en C op de responsie y na te gaan werd een 2^3-proef uitgevoerd. De volgende effecten bleken significant te zijn: A = 5, B = 3, C = 4 en AB = 6.
Het gemiddelde van alle 8 waarnemingen was: $\bar{y} = 40$. De meetfout werd uit voorafgaande proefmetingen bepaald: $\hat{\sigma}_0^2 = 14.4$, $\nu = 20$.
Geef een 95 %-betrouwbaarheidsinterval voor de verwachting van de geschatte responsie voor A op het hoge, B en C op het lage niveau.

6.21. De gemiddelde maandtemperatuur op een willekeurige plaats in Nederland hangt af van verschillende factoren, waarvan we de volgende drie factoren het meest relevant achten:

A ligging van de plaats al dan niet aan zee,
B ligging in noord-, midden- of zuid-Nederland,
C de maand waarin de temperatuur gemeten wordt.

We onderzoeken de invloed van deze factoren elk op twee niveau's:

A: wel - niet aan zee,
B: Noord-Nederland - Zuid-Nederland,
C: januari-juli.

De waarnemingen van de jaren 1974 en 1975 (blokken!) staan in de volgende tabel uitgedrukt in °C.

	aan zee		niet aan zee		
noord	Den Helder		Eelde		
	2.5	3.0	0.5	1.5	januari
	16.0	16.5	17.5	18.5	juli
	Vlissingen		Eindhoven		
zuid	4.0	4.0	1.5	2.0	januari
	17.0	18.0	18.0	19.5	juli
	1974	1975	1974	1975	

a) Geef het volledig lineaire model (met de kodering!).

b) Bepaal de kleinste kwadraten schattingen van de modelparameters en geef een schatting van de restvariantie.

c) Toets welke effecten significant zijn ($\alpha = 0.05$).

d) Geef de schattingen van de verwachte gemiddelde maandtemperaturen voor de 8 niveaucombinaties van de factoren A, B en C volgens het gereduceerde model.

6.22. De volgende tabel geeft 8 waarnemingen van een roteerbare proefopzet van 2 variabelen.

x_2 \ x_1	-2	-1	0	1	2
$-\sqrt{3}$		y_2		y_3	
-1					
0	y_1		y_7, y_8		y_4
1					
$\sqrt{3}$		y_6		y_5	

We beschouwen het model

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \underline{e} \ .$$

a) Laat zien dat var($\hat{\underline{y}}$) constant is in punten op gelijke afstand van het midden van het experimenteergebied.

b) Wat is de maximale waarde van var($\hat{\underline{y}}$) in het experimentele gebied en waar wordt dat maximum bereikt?

c) Wat is de minimale waarde van var($\hat{\underline{y}}$) en waar wordt dit minimum bereikt?

6.23. Gegeven is het lineaire model

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + \underline{e}\,.$$

We beschouwen een proefopzet met 12 waarnemingen, 4 in het punt (0,0) en 8 op de hoekpunten van een regelmatige 8-hoek:

	1	2	3	4	5	6	7	8
x_1	$\sqrt{2}$	1	0	-1	$-\sqrt{2}$	-1	0	1
x_2	0	1	$\sqrt{2}$	1	0	-1	$-\sqrt{2}$	-1

De fouten $\underline{e}$ zijn *onafhankelijk* en *normaal* verdeeld met bekende variantie σ_0^2.

a) Bepaal de variantie-covariantie *matrix* van de schatter $\underline{b}$ van de parametervector.

Hierbij kan de volgende matrix-inversie worden gebruikt:

$$\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}^{-1} = \frac{1}{8}\begin{pmatrix} 8 & -4 & -4 \\ -4 & 5 & 1 \\ -4 & 1 & 5 \end{pmatrix}.$$

b) Bepaal var($\hat{\underline{y}}$), de variantie van de geschatte verwachting van $\underline{y}$ in de volgende drie punten:

	1	2	3
x_1	0	1	1
x_2	0	0	1

c) Laat zien dat var($\hat{\underline{y}}$) constant is op cirkels om de oorsprong:

$$x_1^2 + x_2^2 = r^2.$$

7. Gemengde modellen met gekruiste en geneste factoren

7.1. In zeker land met 30 provincies is een onderzoek uitgevoerd naar variaties in de tarwe-opbrengst. Voor elk der provincies is in 5 gemeenten, telkens op 10 boerderijen per gemeente, door loting één hectare aangewezen, waarvan de opbrengst werd bepaald. De rekenaar heeft berekend KS_{tot} met 1499, en verder 7 kwadratensommen met resp. 29, 4, 9, 116, 261, 36 en 1044 vrijheidsgraden.

a) Stel het model op; geef daarbij ook aan, aan welke factoren U stochastische niveaus toekent en geef de bijbehorende modelvoorwaarden.

b) Stel de variantie-analysetabel op (waarbij U beschikt over de 8 bovengenoemde kwadratensommen).

c) Geef voor twee situaties aan hoe getoetst dient te worden:
I de provincies omvatten een groot aantal gemeenten;
II de provincies omvatten elk slechts vijf gemeenten.

7.2. Uit 3 partijen (P) van een grondstof worden telkens 3 monsters (M) genomen en aan ieder monster werd een analyse in duplo (H) uitgevoerd. We onderstellen dat H een hoofdeffect geeft, maar dat interacties waar H in vóórkomt te verwaarlozen zijn.
De uitkomsten zijn

	P1			P2			P3		
	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
H_1	20.2	26.2	23.8	22.0	22.6	22.9	23.1	22.9	21.8
H_2	24.1	26.9	24.9	23.5	24.6	25.0	22.9	23.7	23.5

a) Geef het model waar U vanuit gaat.

b) Geef de variantie componenten, die bij dit model horen.

c) Hoe luidt de variantie-analyse-tabel?
(Vermeld ook de tussenresultaten.)

d) Wat zijn de verwachtingen der GK-en?
Wat onderstelt U daarbij omtrent vaste en stochastische niveau's?

e) Toets welke effecten significant zijn en geef Uw conclusies samenvattend weer.

7.3. De uitzettingscoëfficient van glas wordt in de practijk bepaald door een plaatje van het te onderzoeken glas en een plaatje standaardglas aan elkaar te smelten en onder een polarisatiemicroscoop de spanning op het scheidingsvlak te meten.

De foutenbronnen zijn o.a.:

1. de nauwkeurigheid van het aan elkaar smelten;
2. De nauwkeurigheid van de waarneming.

Een onderzoek naar de grootte der fouten werd uitgevoerd als volgt:

Vier glasblazers vervaardigen ieder 3 plaatjes en deze werden alle door 3 verschillende waarnemers gemeten.

De resultaten zijn gegeven in de volgende tabel.

glasblazer	plaatje	waarnemer I	II	III
A	1	60	56	56
	2	58	55	53
	3	58	56	55
B	4	58	57	55
	5	62	65	60
	6	58	61	59
C	7	55	52	52
	8	56	59	60
	9	62	62	55
D	10	57	57	52
	11	63	63	64
	12	53	48	48

a) Stel het model op (beschouw blazers en waarnemers als vaste niveaus).

b) Voer de nodige berekeningen overzichtelijk uit (denk aan coderen!).

c) Geef de variantie-analysetabel, onderzoek hoe moet worden getoetst en toets. Geef schattingen van de modelparameters.

7.4. De variantie-analyse-tabel bij opgave 4.4 (het probleem van schimmelgroei op houten afwerkingen van buitenmuren van huizen) luidt:

Bron	KS	ν	GK
V	539.6	2	269.8
P	377.1	2	188.5
Z	550.0	1	550.0
VP	108.6	4	27.2
VZ	55.0	2	27.5
PZ	55.4	2	27.7
VPZ	35.0	4	8.7
Rest	61.25	54	1.13
Totaal	1781.9	71	

a) Bereken de verwachtingen van de GK-en als P een factor is met stochastische niveau's en V en Z factoren zijn met vaste niveau's.

b) Toets welke effecten significant zijn en vergelijk de konklusies met die van opgave 4.4, waar alle factoren vaste niveau's hadden.

7.5. Met 4 ongeijkte thermometers werd door 3 analisten het smeltpunt T bepaald van hydrochinon aan monsters uit een homogene partij. Iedere bepaling werd tweemaal uitgevoerd en wel op verschillende dagen.
De gecodeerde waarnemingen zijn:

Analist	Dag	Thermometer A	Thermometer B	Thermometer C	Thermometer D
I	-	4.0	3.5	1.5	3.5
	-	3.5	3.5	2.5	3.5
II	-	3.0	2.0	1.0	1.0
	-	3.0	3.5	2.0	2.0
III	-	3.5	3.0	3.0	2.5
	-	3.0	3.5	3.0	3.0

De variantie-analyse-tabel hierbij is

Bron	KS	ν	GK
Thermometers, Th	5.20	3	1.73
Analisten , A	4.75	2	2.38
Interactie , Th × A	2.83	6	0.47
Dagen , D	3.13	12	0.26
Totaal	15.91	23	

a) Al naar de wijze waarop de waarnemingen worden georganiseerd moeten dagen, D, worden opgevat als

1. ondergeschikt aan Th en A,
2. ondergeschikt aan A, doch nevengeschikt aan Th,
3. nevengeschikt aan Th en A beide.

Beschrijf de experimentele omstandigheden waaronder deze 3 modellen van toepassing zijn.
Op welke van deze modellen slaat de variantie-analyse als boven gegeven?

b) Stel A heeft vaste niveau's en Th heeft stochastische niveau's.
Is D een factor met vaste dan wel met stochastische niveau's? Motiveer Uw keuze!
Geef bij bovenstaande variantie-analyse de verwachtingen van de GK-en.

c) De KS tussen dagen met ν = 12 omvat de effecten D·A en DT·A met resp. 3 en 9 vrijheidsgraden. Bereken de KS voor D·A en voer de splitsing uit. Toets of D·A significant groter is dan DT·A. Geeft dit U reden aan het toegepaste model te twijfelen?

d) Vat de belangrijkste resultaten samen.

7.6. Uit 4 verschillende balen kunstmest zijn telkens aselect 3 monsters genomen. Aan ieder monster zijn twee bepalingen uitgevoerd van het gehalte ammoniumsulfaat (x_{ijk} in %: $i = 1,2,3,4$; $j = 1,2,3$; $k = 1,2$)

monsters		1	2	3
	1	7.30	7.41	7.37
		7.30	7.35	7.37
	2	7.25	7.30	7.29
baal		7.25	7.24	7.32
	3	7.40	7.43	7.31
		7.38	7.48	7.40
	4	7.26	7.37	7.24
		7.26	7.30	7.34

Als we de waarnemingen coderen $y_{ijk} = 100\ (x_{ijk} - 7.20)$ krijgen we de volgende kwadratensommen:

$$\Sigma y_{ijk}^2 = 5066 \qquad \Sigma y_{..k}^2 = 48690 \qquad \Sigma y_{.jk}^2 = 17142$$

$$\Sigma y_{i..}^2 = 27774 \qquad \Sigma y_{ij.}^2 = 9792 \qquad y_{...}^2 = 97344$$

$$\Sigma y_{.j.}^2 = 33600 \qquad \Sigma y_{i.k}^2 = 13986$$

a) Geef het bijbehorende statistische model (met onderstellingen).

b) Geef de variantieanalysetabel met de verwachte gemiddelde kwadraten en toets of er verschil is tussen de balen en tussen de monsters ($\alpha = 0.05$).

c) Geef puntschattingen van de statistisch relevante modelparameters (algemeen gemiddelde, variantiecomponenten).

7.7. Een fabricant heeft uit ieder van 4 partijen grondstoffen 3 monsters genomen en deze door 3 analisten laten analyseren.

De reeds gecodeerde waarnemingen zijn:

		Partij			
		1	2	3	4
Analist	A	4	0	8	2
	B	3	5	10	4
	C	5	1	9	3

We gaan ervan uit dat er geen interactie is tussen partijen en analisten.

a) Geef het bijbehorende model als

I. het alleen om deze 4 partijen gaat

II. de 4 partijen worden opgevat als een steekproef uit een veel grotere populatie van partijen.

b) Voer een variantie-analyse uit voor deze gegevens en toets welke effecten significant zijn ($\alpha = 0{,}05$).

c) Geef (eventueel na vereenvoudiging van het model) een schatting van de variantie van de toevallige fouten in de waarnemingen en (alleen voor model II) van de variantie tussen partijen.

d) Bereken de schatting voor de variantie van het gemiddelde van de 12 waarnemingen:

(i) voor model I

(ii) voor model II.

8. Gemengde opgaven

8.1. a) Wat is het doel van verloting? (Antwoord in maximaal 4 regels.)

b) Wat is het doel van blokvorming? (Antwoord in maximaal 4 regels.)

c) Waarom is het zinvol onderscheid te maken tussen kwalitatieve en kwantitatieve factoren?
In welke situatie is dit onderscheid niet relevant?

8.2. a) Welke criteria spelen een rol bij de keuze tussen een proefopzet met volledige verloting (zonder blokken) en een met blokvorming (met verloting binnen blokken)?

b) Wat zijn de onderstellingen bij toepassing van de F-toets in de variantie-analyse?

c) Wat zijn lineaire contrasten?

d) Wat is interactie?

8.3. a) Wat zijn de onderstellingen bij toepassing van de t-toets?

b) Bij iedere niveaukombinatie van een 2^n-proef worden m metingen gedaan. De restvariantie σ_0^2 wordt enkel geschat uit deze herhalingen (dus niet uit mogelijk te verwaarlozen hogere orde interacties).
Wat is het aantal vrijheidsgraden bij deze schatting?

c) Wanneer heet een proefopzet orthogonaal?
Wat zijn de gevolgen van orthogonaliteit?

8.4. a) i) Onder welke voorwaarde(n) mag een Grieks-Latijns vierkant worden toegepast?

ii) Beschouw een model met 2 factoren (A en B).
Geef een voorbeeld waarbij de tweede factor genest is binnen de eerste (B:A) en een voorbeeld waarbij A en B nevengeschikt zijn.
Hoe toetst men of het effect B:A significant is?

b) De helft van een 2^5-proefschema wordt uitgevoerd in twee blokken. Het definiërende contrast is I = +ABCDE en verstrengeld met blokken is BDE.
Geef aan hoe de waarnemingen over de twee blokken verdeeld zijn.

8.5. Vier patiënten (1), (2), (3) en (4) krijgen in twee opvolgende perioden I en II de behandelingen a en b, twee patiënten in de volgorde ab en twee patiënten in de volgorde ba.

De response is weergegeven in de volgende tabel:

Periode	Patienten							
	(1)		(2)		(3)		(4)	
I	a	24	b	9	a	33	b	14
II	b	21	a	12	b	28	a	27

a) Formuleer een lineair model zonder interacties, met bijbehorende coderingen
Aanwijzing: gebruik voor de factoren op 2 niveau's de codering (-1,+1).

b) Bereken schattingen van de parameters en van σ_0^2.
U kunt desgewenst gebruik maken van de volgende matrixinversie:

$$\begin{pmatrix} 8 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 2 & 1 & 1 \\ -1 & 1 & 2 & 1 \\ -1 & 1 & 1 & 2 \end{pmatrix}.$$

c) Toets de hypothese dat de behandelingen niet verschillen.

8.6. Wij hebben de 3 volgende groepen waarnemingen:

1.

x	3	2	2	5
y	2	3	5	10

2.

x	2	4	4	6
y	4	9	11	16

3.

x	2	1	1	4
y	6	8	9	13

a) Het model luidt:

$$\underline{y} = \alpha_i + \beta_i x + \underline{e} \ ,$$

waarbij i = 1,2,3 al naargelang het groep 1, 2 of 3 betreft. De toevallige fouten $\underline{e}$ zijn onafhankelijk normaal verdeeld met gemiddelde 0 en (onbekende) variantie σ^2.

Toets de hypothese $H_0 : \beta_1 = \beta_2 = \beta_3 \quad (\alpha = 0{,}05)$.

b) Beschouw nu het model

$$\underline{y} = \alpha_i + \beta x + \underline{e} \qquad (i = 1,2,3) \ .$$

De veronderstellingen over de verdeling van $\underline{e}$ zijn dezelfde als onder a).
Toets de hypothese $H_0 : \beta = 0 \quad (\alpha = 0{,}05)$.

8.7. Bij een factoriëel experiment zijn de volgende reeds gecodeerde waarnemingen verkregen.

x_2 \ x_1	-1	0	1
-1	1		3
0		4;6	
1	1		9

(Bij de niveaucombinatie $x_1 = 0$, $x_2 = 0$ zijn dus 2 waarnemingen gedaan).
Voor de analyse van deze gegevens gaan we uit van het model

$$\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \underline{e} \ .$$

Hierin is $x_3 = x_1 x_2$ en $x_4 = 3x_1^2 - 2$.

a) Bereken $(X'X)^{-1}$, $X'y$, b, $y'y$, $b'X'y$, KS_r en $\hat{\sigma}_0^2$.

b) Uit vroegere metingen is een schatting bekend van σ_0^2, nl. $\hat{\sigma}_0^2 = 4/3$, $\nu = 9$. Bereken uit deze schatting en uit de onder a) verkregen schatting een nieuwe schatting van σ_0^2.

c) Toets dan welke effecten significant zijn ($\alpha = 0.05$) en geef het aangepaste model na weglating van de niet-significante effecten.

d) Geef een 95%-betrouwbaarheidsinterval voor de verwachte responsie bij de niveaucombinatie $x_1 = 0$, $x_2 = 1$ (gebruik het gereduceerde model).

8.8. a) De opbrengst van een proces hangt vermoedelijk af van omwentelingssnelheid (S), concentratie (C) en temperatuur (T). Om dit na te gaan wordt een experiment uitgevoerd, waarbij voor elke factor twee niveaus worden genomen.

De resultaten (gecodeerd) zijn:

S in omw/min	200		400	
C in %	2	4	2	4
T in °C				
30	4	14	4	10
40	20	18	12	14

Aangenomen mag worden dat er tussen de 3 factoren geen belangrijke interacties te verwachten zijn.

i) Ga na welke effecten significant zijn ($\alpha = 0.05$)

ii) Als er geen interacties zijn tussen de factoren kan men om de effecten te schatten volstaan met 4 waarnemingen. Bij welke niveaucombinaties zou U dan de opbrengst willen bepalen?

b) i) Wat is de kleinste fractie van een 2^5-experiment waarbij de hoofdeffecten nog geschat kunnen worden, als interacties te verwaarlozen zijn?

ii) Geef definiërende contrasten voor zo'n fractie.

iii) Uit welke niveaucombinaties bestaat zo'n fractie?

8.9. Drie patiënten (1), (2) en (3) krijgen elk gedurende de perioden I en II twee van de drie verschillende behandelingen a, b en c volgens het schema:

	Patiënt		
Periode	(1)	(2)	(3)
I	a	b	c
II	b	c	a

Het model voor de response $\underline{y}$ luidt:

$$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \underline{e}_{ijk} \qquad \begin{array}{l} i = 1,2 \text{ (periodes)}, \\ j = 1,2,3 \text{ (patiënten)}, \\ k = 1,2,3 \text{ (behandelingen)}, \end{array}$$

met als bijvoorwaarden $\sum\alpha_i = \sum\beta_j = \sum\gamma_k = 0$.

a) Laat zien dat, als $\mathcal{E}\underline{e}_{ijk} = 0$,

$$\tfrac{1}{3}(\underline{y}_{111} - \underline{y}_{133} - \underline{y}_{212} + \underline{y}_{231}),$$

een zuivere schatter is voor γ_1 en geef overeenkomstige schatters voor γ_2 en γ_3.

b) Zij verder gegeven dat de $\underline{e}_{ijk}$ onafhankelijk normaal verdeeld zijn met variantie $\sigma_0^2 = 4$ en onderstel dat de volgende waarnemingen zijn verricht:

	Patiënten		
Periode	(1)	(2)	(3)
I	19	21	14
II	27	26	13

Toets de hypothese $H_0 : \gamma_1 = 0 \quad (\alpha = 0{,}05)$.

c) Toets de hypothese $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0 \quad (\alpha = 0{,}05)$.

Aanwijzing: Denk eraan dat $\dfrac{KS_{r0}}{\sigma_0^2}$ onder H_0 een χ^2-verdeling heeft en dat σ_0^2 gegeven is.

ANTWOORDEN

Antwoorden

1. Inleiding.

1.1. b) slijtage van de boor, inhomogeniteit van de materialen.

c) bij slechts van 5 van de 30 paren waarnemingen is de eerste kleiner dan de tweede. Vergelijk
$P(\underline{x} \le 5 | \underline{x} \sim BN\ (p = \frac{1}{2},\ n = 30)) = 0{,}0003$ met $\frac{1}{2}\alpha$.

1.2. a) basiseenheid waaraan of waarover waarnemingen worden gedaan.

b) te onderzoeken kwantiteit of kwaliteit, die mogelijk oorzaak van variaties kan zijn.

c) instelling (kwantitatieve faktor) of kwaliteit (kwalitatieve faktor).

d) stimulus die wordt toegepast met het doel het effect op de onderzoeksituatie waar te nemen of het effect te vergelijken met die van andere behandelingen.

e) het aselect toewijzen van niveaucombinaties cq behandelingen aan de experimentele eenheden.

f) het verdelen van experimentele eenheden in homogene groepen.

g) het aantal waarnemingen dat binnen de specificaties van het systeem cq. model vrij kan worden gekozen.

h) lineaire functie met bekende constante coëfficiënten, waarvan de som nul is.

i) het deel van de variantie van een serie waarnemingen dat overblijft na aftrek van het effect van systematische elementen.

k) model, waarbij de afhankelijk variabele een lineaire combinatie is van de verklarende variabelen.

1.3. a) 1° het opnemen van alle variaties als gevolg van onbekende, onvermijdbare en niet-meetbare factoren in de restvariantie: men voorkomt verstrengeling.

2° het isoleren van bronnen van heterogeniteit van de experimentele eenheden: onbeheerste variaties in het waarnemingsmateriaal worden buiten de restvariantie gehouden.

b) $\underline{y}_i = \Sigma\beta_i x_i + \underline{e}_i$ of $\underline{y} = X\beta + \underline{e}$

$\mathcal{E}\underline{e} = 0 \qquad \mathcal{E}\underline{ee}' = \sigma^2 I$

1.4. a) 1 vr. gr. voor het gemiddelde
2 " " " verschillen tussen methoden
21 " " " verschillen tussen personen binnen methoden.

b) 1 vr. gr. voor het gemiddelde
2 " " " verschillen tussen methoden
7 " " " verschillen tussen personen
14 " " " verschillen binnen methoden en binnen personen.

1.5. a) 1 vr. gr. voor het gemiddelde
3 " " " verschillen tussen methoden
32 " " " verschillen binnen methoden.

b) 1 vr. gr. voor het gemiddelde
3 " " " verschillen tussen methoden
8 " " " verschillen tussen monsters
24 " " " verschillen binnen methoden en binnen monsters.

1.6. a) I $\underline{y} = \beta_0 + \beta_1 x + \underline{e} \qquad \underline{e} \sim N(0, \sigma^2 I)$

$x = 0$ bij A of $x = -1$ bij A
$x = 1$ bij B $\quad$ $x = 1$ bij B

II $\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_9 x_9 + \underline{e} \qquad \underline{e} \sim N(0, \sigma^2 I)$

$$\begin{array}{ll} x_1 = 0 \text{ bij A} & \left.\begin{array}{l} x_i = 1 \text{ voor monster } i \\ x_i = 0 \text{ voor andere monsters} \end{array}\right\} i = 2, \ldots, 9 \\ x_1 = 1 \text{ bij B} & \end{array}$$

b) II ; de methoden worden binnen elk monster vergeleken.

c) I : $t_{16} = -0.69$ of $u = -0.68$

II : $t_8 = -3.16$ of $u = -3.04$

d) de waarnemingen zijn onafhankelijke aselecte steekproeven uit normale verdelingen met dezelfde (onbekende) variantie.

1.7. a) $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1}_{\text{stimuli}} + \underbrace{\beta_2 x_2 + \ldots\ldots + \beta_9 x_9}_{\text{personen}} + \underline{e}$

met geschikte codering en de nodige onderstellingen.

b) $t_8 = -1.64$ of $u = -1.70$

c) $(-2.45,\ 0.41)$ of $(-2.20,\ 0.16)$

d) De factor stimulus is verstrengeld met de factor tijd; beter is het de volgorde van toedienen te verloten.

1.8. a) Om verschillen tussen personen buiten de beoordelingsmaatstaf (restvariantie) te houden, verdient een proefopzet met blokvorming (en verloting binnen de blokken) de voorkeur.

b) $t_7 = -2.57$ of $u = -2.53$

1.9. a) $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1}_{\text{machines}} + \underbrace{\beta_2 x_2 + \ldots\ldots + \beta_6 x_6}_{\text{rekenaars}} + \underline{e}\quad , \quad \underline{e} \sim N(0, \sigma^2\ I)$

$$x_1 = 0 \text{ of } -1 \text{ voor A} \qquad \left.\begin{array}{l} x_i = 1 \text{ voor rekenaar } i \\ x_i = 0 \text{ voor andere rekenaars} \end{array}\right\} i = 2,\ldots,6$$

$$x_1 = 1 \quad \text{voor B}$$

b) $t_5 = -2.16$ of $u = -1.94$

c) $(-15.3,\ 1.3)$ of $(-14.1,\ 0.1)$

2. Lineaire modellen

2.1. a) $b' = \frac{1}{84}\,(120,\ 42,\ 10,\ -42)$

b) (0.003, 0.110)

2.2. a) $b' = \frac{1}{84}\,(-60,\ -12,\ 12)$

b) $\hat{\sigma}_0^{\,2} = \frac{1}{7}$, $\nu = 4$

c) (0.23, 1.08)

2.3. a) $a' = \frac{1}{84}\,(-12,\ -12,\ 12)$; $\hat{\sigma}_0^{\,2} = \frac{1}{7}$, $\nu = 4$

b) X'X is een diagonaalmatrix.

c) $\beta_0 = \alpha_0 - 4\alpha_2$; $\beta_1 = \alpha_1$; $\beta_2 = \alpha_2$

2.4. a) $\hat{y} = 2.1 - 0.6x$

b) $\hat{e}' = (0.3,\ -0.1,\ -0.5,\ 0.1,\ 0.2)$ $\quad X'\hat{e} = \begin{pmatrix}0\\0\end{pmatrix}$

c) $y'y = 15.25$, $\nu = 5$; $(Xb)'Xb = b'X'y = 14.85$, $\nu = 2$; $\hat{e}'\hat{e} = 0.40$, $\nu = 3$

d) X'X diagonaalmatrix.

2.5. a) i) $\hat{\sigma}^2 = 1048.58$, $\nu = 6$

ii) $t_6 = -0.28$

iii) $\underline{y} = \beta_0 + \beta_1 x_1 + \underline{e}$, $\underline{e} \sim N(0,\ \sigma_0^{\,2}\, I)$

$x_1 = 1$ voor eerste serie

$x_1 = -1$ voor tweede serie

iv) $b' = \frac{1}{8}\,(316,\ -26)$; $\hat{\sigma}_0^{\,2} = 1048.58$, $\nu = 6$

b) i) $\hat{\sigma}^2 = 116.83$, $\nu = 3$

ii) $t_3 = -0.85$

iii) $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1}_{\text{series}} + \underbrace{\beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4}_{\text{paren}} + \underline{e}$, $\underline{e} \sim N(0, \sigma_0^2\, I)$

$x_1 = 1$, 1^e serie $x_2 = 1$, 1^e paar $x_3 = 1$, 3^e paar
$x_1 = -1$, 2^e serie $x_2 = -1$, 2^e paar $x_3 = -1$, 4^e paar
$x_2 = 0$, 3^e en 4^e paar $x_3 = 0$, 1^e en 2^e paar

$x_4 = 1$, 1^e en 2^e paar
$x_4 = -1$, 3^e en 4^e paar

iv) $b' = \frac{1}{8}(316, -26, -232, 140, 104)$; $\hat{\sigma}_0^2 = 116.83$, $\nu = 3$

2.6. Matrix van waarnemingen rijgewijs lezen.

a),b) I $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2}_{\text{methoden}} + \underline{e}$, $\underline{e} \sim N(0, \sigma_0^2\, I)$

$x_1 = 1$ voor B $x_2 = 1$ voor C
$x_1 = 0$ anders $x_2 = 0$ anders

$$X' = \begin{pmatrix} 1111 & 1111 & 1111 \\ 0000 & 1111 & 0000 \\ 0000 & 0000 & 1111 \end{pmatrix}$$

II $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2}_{\text{methoden}} + \underbrace{\beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5}_{\text{kolommen}} + \underline{e}$, $\underline{e} \sim N(0, \sigma_0^2\, I)$

x_1 en x_2 coderen als bij I

$x_3 = 1$, 2^e kolom $x_4 = 1$, 3^e kolom $x_5 = 1$, 4^ekolom
$x_3 = 0$, anders $x_4 = 0$, anders $x_5 = 0$, anders

$$X' = \begin{pmatrix} 1111 & 1111 & 1111 \\ 0000 & 1111 & 0000 \\ 0000 & 0000 & 1111 \\ 0100 & 0100 & 0100 \\ 0010 & 0010 & 0010 \\ 0001 & 0001 & 0001 \end{pmatrix}$$

c) I $b' = \frac{1}{4}\,(1160,\ 92,\ 21)$

$\hat{\mu} = 299.42$

$b_0 = \hat{\mu} + \hat{\alpha}_1 \qquad \hat{\alpha}_1 = -9.42$

$b_1 = \hat{\alpha}_2 - \hat{\alpha}_1 \qquad \hat{\alpha}_2 = 13.58$

$b_2 = \hat{\alpha}_3 - \hat{\alpha}_1 \qquad \hat{\alpha}_3 = -4.17$

II $b' = \frac{1}{12}\,(3655,\ 276,\ 63,\ -312,\ -380,\ -8)$

d) I 1 vr. gr. voor gemiddelde

2 " " " methoden

9 " " " rest

II 1 vr. gr. voor gemiddelde

2 " " " methoden

3 " " " kolommen

6 " " " rest

2.7. a) $b' = \frac{1}{4}\,(249,\ -27,\ 17)$; $\hat{\sigma}_0^{\,2} = 6\frac{1}{4}$, $\nu = 1$

b) $r' = \left(-\frac{5}{4}\ ,\ \frac{5}{4}\ ,\ \frac{5}{4}\ ,\ -\frac{5}{4}\right)$

c) $F_5^1 = \frac{25}{16}$

d) $(-9.32,\ -4.18)$

2.8. a) $x_1 = \frac{1}{10}\,(T_1 - 60)$; $x_2 = \frac{1}{5}\,(P - 15)$; $x_4 = \frac{1}{50}\,(T_2 - 150)$;

$x_3 = -1$ voor K_1 , $x_3 = 1$ voor K_2

b) $b' = (20.5000,\ -2.5125,\ -1.8375,\ -0.7875,\ 0.6500)$

$\hat{\sigma}_0^{\,2} = 7.5284$, $\nu = 11$

d) $r'r = 82.8125$

e) $t_{11} = -3.66$; $(-4.02,\ -1.00)$

f) $\hat{\underline{y}} = \underline{b}_0 - \underline{b}_1 + \underline{b}_2 - \underline{b}_3 + \underline{b}_4$; $\hat{y} = 22.6125$; $\widehat{\mathrm{var}}\ \hat{\underline{y}} = \frac{5}{16}\,\hat{\sigma}_0^{\,2} = 2.353$

$\mathcal{E}\hat{\underline{y}} \in (19.85,\ 25.37)$

2.9. a) $b' = \frac{1}{24}$ (7186, 63, -163) ; $\hat{\sigma}_0 = 17.97$, $\nu = 9$

b) (12.4, 32.9)

c) GK = 581.08 , $\nu = 2$

d) $F_9^2 = 1.80$

2.10. a) $b' = \frac{1}{24}$ (7186, 63, -163, -8, -68, 342)

$\hat{\sigma}_0{}^2 = 69.86$, $\nu = 6$

b) Methoden: GK = 581.08 , $\nu = 2$; $F_6^2 = 8.32$

Blokken : GK = 828.53 , $\nu = 3$; $F_6^3 = 11.9$

2.11. Codering: $x_1 = \frac{1}{25}(T - 200)$

$x_2 = \frac{1}{4}(C - 8)$

a) $b' = (75, 11, 5)$; $\hat{\sigma}_0{}^2 = 4$, $\nu = 3$

b) $t_3 = 4.33$ of $F_3^1 = 18.75$

c) De waarnemingen zijn onderling onafhankelijk en normaal verdeeld met dezelfde variantie.

d) $x_1 = 0$, $x_2 = \frac{1}{2}$; vâr $\hat{\underline{y}} = \frac{1}{4}\hat{\sigma}_0{}^2 = 1$

$\mathcal{E}\hat{\underline{y}} \in (74.32, 80.68)$

2.12. I a) $x_1 = -1$, A_1 $x_2 = \frac{1}{10}(t - 30)$

$x_2 = 1$, A_2

$b' = \frac{1}{20}$ (211.4, 24.6, 16.1) ; $\hat{\sigma}_0{}^2 = 0.22$, $\nu = 7$

b) $\hat{y}_A = 9.34 + 0.805\, x_2$

$\hat{y}_B = 11.80 + 0.805\, x_2$

II a) $b' = \frac{1}{20}$ (211.4, 24.6, 16.1, 4.1)

$\hat{\sigma}_0{}^2 = 0.122$, $\nu = 6$

b) $\hat{y}_A = 9.34 + 0.6\, x_2$

$\hat{y}_B = 11.80 + 1.01\, x_2$

2.13. a) $x_1 = \frac{x - 2.7}{0.1}$

$b' = \frac{1}{168}(2793,\ 95)$; $\hat{\sigma}_0^{\,2} = 6.026$, $\nu = 6$

$\hat{\text{var}}\ \underline{b} = \begin{pmatrix} 0.753 & 0 \\ 0 & 0.036 \end{pmatrix}$

$\hat{\sigma}_0^{\,2} = \hat{\sigma}_{\underline{\bar{x}}}^{\,2} = 6.026 \Rightarrow \hat{\sigma}_{\underline{x}}^{\,2} = 10 * 6.026 = 60.26$

b) $\hat{\text{var}}\ \underline{a} = \begin{pmatrix} 27.0 & -9.7 \\ -9.7 & 3.6 \end{pmatrix}$; $\alpha_1 \in (0.99,\ 10.31)$

c) $p'p = 36.155$

d) Het systematische tekenverloop van de residuën wijst erop dat een kwadratische term in het model ontbreekt.

2.14. a) $b' = \frac{1}{32}(2800,\ -36 - 8\sqrt{2},\ 20 - 4\sqrt{2},\ -76,\ -100,\ -152)$

$\hat{\sigma}_0^{\,2} = 4.57$, $\nu = 6$

b) $F_6^3 = 12.64$

2.15. a) i) Als de A-metingen de eerste dag en de B-metingen de tweede dag zijn uitgevoerd, dan is de factor "omstandigheden" verstrengeld met de factor "dagen".

ii) De toetsprocedure is niet correct: $\nu = 9$

2.16. a) $GK_A = 21.23$, $\nu = 2$

$GK_B = 4.06$, $\nu = 3$

$GK_{rest} = 0.20$, $\nu = 6$

b) A: $F_6^2 = 5.14$; B: $F_6^3 = 4.76$

c) $b' = (9.05,\ 2.30,\ -0.075,\ 0.447,\ 0.033)$; $\hat{\sigma}_0^{\,2} = 0.196$, $\nu = 7$

$$\hat{\text{var}}\ \underline{b} = \begin{pmatrix} 0.016 & 0 & 0 & 0 & 0 \\ 0 & 0.0245 & 0 & 0 & 0 \\ 0 & 0 & 0.0082 & 0 & 0 \\ 0 & 0 & 0 & 0.00327 & 0 \\ 0 & 0 & 0 & 0 & 0.016 \end{pmatrix}$$

d) $F_7^1 < 1$

e) $F_7^2 = 30.5$

f) $\hat{y}_{A1} = 4.7 + 0.06\ t \qquad \hat{\sigma}_0^{\,2} = 0.196\ , \quad \nu = 7$

$\hat{y}_{A2} = 7.2 + 0.06\ t$

$\hat{y}_{A3} = 9.3 + 0.06\ t$

2.17. a) $\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underline{e}\ , \quad \underline{e} \sim N(0, \sigma_0^{\,2}\ I)$

x_1 = begingewicht - 10

$x_2 = \begin{cases} -1, & A \\ 1, & B \end{cases}$

b) $b' = (11.057,\ 0.952,\ -0.825)$

$\hat{y}_A = 2.362 + 0.952\ x \qquad$ (x = begingewicht)

$\hat{y}_B = 0.712 + 0.952\ x$

c) $\hat{\sigma}_0^{\,2} = 0.7268\ , \quad \nu = 5$

$\beta_2 \in (-1.607,\ -0.043)$

d) $t_5 = 5.86$

e) $\hat{\text{var}}\ \hat{y} = \frac{135}{440}\hat{\sigma}_0^{\,2} = 0.2230$

$\mathcal{E}\underline{\hat{y}} \in (11.18,\ 13.09)$

f) $\underline{y} = \alpha_0 + \alpha_1 x_1 + \underline{e}\ , \quad \underline{e} \sim N(0, \sigma_0^{\,2}\ I)$

$x_1 = \begin{cases} -1, & A \\ 1, & B \end{cases}$

$a' = (13.9125,\ -0.5875)\ ; \quad \hat{\sigma}_0^{\,2} = 4.7583\ , \quad \nu = 6$

$t_6 = -0.76$

g) $\underline{z} = \gamma_0 + \gamma_1 x_1 + \underline{e}\ , \quad \underline{e} \sim N(0, \sigma_0^{\,2}\ I)$

$x_1 = \begin{cases} -1, & A \\ 1, & B \end{cases} \qquad z_i$ = gewichtstoename = responsie - begingewicht

$c' = (0.9125,\ -0.8375)\ ; \quad \hat{\sigma}_0^{\,2} = 0.6162\ , \quad \nu = 6$

$t_6 = -3.02$

2.18. a) $a' = (1.6,\ 0.8)$; $\hat{\sigma}_0^{\ 2} = 1.2$, $\nu = 3$

b) $b' = (2.4,\ 2.2)$; $\hat{\sigma}_0^{\ 2} = 1.2$, $\nu = 3$

c) $t_6 = 2.86$

2.19. a) $\hat{y} = 60 + 2x_1 + x_2$; $x_1 = \frac{1}{10}$ (temp. - 80)

$x_2 = \frac{1}{5}$ (dosis - 40)

$\hat{\sigma}_0^{\ 2} = 1.5$, $\nu = 4$

b) i) $t_4 = \sqrt{2}$ of $F_4^1 = 2$

ii) $F_4^2 = 9.33$

2.20. a) $b' = (23,\ 2,\ -2,\ 2)$; $\hat{\sigma}_0^{\ 2} = 4.4$, $\nu = 5$

b) $t_5 = -1.17$

c) $t_5 = -2.34$

2.21. a) proefopzet I: $\sum_i \text{var}\ \underline{b}_i = \frac{25}{36}\,\sigma_0^{\ 2}$

proefopzet II: $\sum_i \text{var}\ \underline{b}_i = \frac{35}{72}\,\sigma_0^{\ 2}$

b) proefopzet II, want X'X is singulier.

3. Latijnse vierkanten

3.1. a) $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{dagen}} + \underbrace{\beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6}_{\text{perioden}} + \underbrace{\beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9}_{\text{omstandigheden}} + \underline{e}$

met geschikte codering!!

of $\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \underline{e}_{ijk}$, $\Sigma\alpha_i = \Sigma\beta_j = \Sigma\gamma_k = 0$

b)

Bron	ν	GK	F
Omstandigheden	3	$172\frac{1}{6}$	2.62
Dagen	3	$30\frac{2}{3}$	<1
Perioden	3	$26\frac{2}{3}$	<1
Rest	6	$65\frac{2}{3}$	

c) $\hat{y} = 115.5$; $\hat{\sigma}_0^2 = 65\frac{2}{3}$, $\nu = 6$

d) H_0: $\beta_1 = \ldots\ldots = \beta_9 = 0$ of $a_i = b_j = c_k = 0$, $\forall_{i,j,k}$

$F_6^9 = \dfrac{688/9}{65\frac{2}{3}} = 1.17$

$\hat{y} = \hat{\beta}_0 = \hat{\mu} = \bar{y} = 115.5$

3.2. a) Op elke auto van iedere soort één band.

$\underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{auto's}} + \underbrace{\beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6}_{\text{banden}} + \underline{e}$

b) $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{auto's}} + \underbrace{\beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6}_{\text{banden}} + \underbrace{\beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9}_{\text{posities}} + \underline{e}$

Voorwaarde: geen interacties

Voordeel: orthogonale proefopzet met 16 waarnemingen i.p.v. 64.

c)

Bron	ν	GK	F
auto's	3	7.73	7.29
banden	3	9.23	8.71
posities	3	2.56	2.42
rest	6	1.06	

Er is verschil tussen auto's en tussen banden.

3.3. a)

Bron	ν	GK	F
Posities	4	0.142	2.09
Strip	4	0.078	1.15
Electrode	4	1.072	15.8
Rest	12	0.068	

De elektroden verschillen significant.

b) Geen interacties.

3.4.

Bron	ν	GK	F
Rijen	3	$5\frac{1}{3}$	8
Kolommen	3	$3\frac{1}{3}$	5
Meststoffen	3	$1\frac{1}{3}$	2
Tarwerassen	3	$11\frac{1}{3}$	17
Rest	3	$\frac{2}{3}$	

Er is verschil tussen tarwerassen.

3.5. a) $$\underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{dagen}} + \underbrace{\beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6}_{\text{perioden}} + \underbrace{\beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9}_{\text{omstandigheden}} + \underline{e}$$

met geschikte codering!

of $\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \underline{e}_{ijk}$, $\Sigma\alpha_i = \Sigma\beta_j = \Sigma\gamma_k = 0$

b)

Bron	ν	GK	F
Dagen	3	$5\frac{1}{3}$	4
Perioden	3	$2\frac{2}{3}$	2
Omstandigheden	3	8	6
Rest	6	$1\frac{1}{3}$	

c) Omstandigheden verschillen significant.

3.6. a)

Bron	ν	GK	F
Tijd	3	$25\frac{2}{3}$	11.8
Vernis	3	$8\frac{1}{6}$	3.8
Aceton	3	$9\frac{1}{6}$	4.2
Acryl	3	$\frac{1}{2}$	< 1
Rest	3	$2\frac{1}{6}$	

Alleen dompeltijd is significant.

b) Er zijn geen interacties.

c) De waarnemingen zijn onderling onafhankelijk en normaal verdeeld met dezelfde variantie.

d) Er zijn geen vrijheidsgraden beschikbaar voor de restvariantie.

4. Faktoriële proefopzetten

4.1. a) $\underline{y}_{ij} = \mu + \alpha_i + \beta_j + \underline{e}_{ij}$, $\Sigma\alpha_i = \Sigma\beta_j = 0$ $\quad \underline{e}_{ij} \sim N(0, \sigma_0^2)$
$\underline{e}_{ij}$ o.o.

of $\quad \underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2}_{\text{waarnemers}} + \underbrace{\beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5}_{\text{dieren}} + \underline{e}$, $\quad \underline{e} \sim N(0, \sigma_0^2\, I)$

$$x_1 = \begin{cases} -1, & A \\ 1, & B \\ 0, & C \end{cases} \qquad x_2 = \begin{cases} 1, & A \\ 1, & B \\ -2, & C \end{cases}$$

$$x_3 = \begin{cases} -1, & 1^e \text{ dier} \\ 1, & 2^e \text{ dier} \\ 0, & 3^e \text{ dier} \\ 0, & 4^e \text{ dier} \end{cases} \qquad x_4 = \begin{cases} 0, & 1^e \text{ dier} \\ 0, & 2^e \text{ dier} \\ -1, & 3^e \text{ dier} \\ 1, & 4^e \text{ dier} \end{cases} \qquad x_5 = \begin{cases} -1, & 1^e \text{ dier} \\ -1, & 2^e \text{ dier} \\ 1, & 3^e \text{ dier} \\ 1, & 4^e \text{ dier} \end{cases}$$

b)

Bron	ν	GK	F
waarnemers	2	5.25	6.52
dieren	3	1902.97	
rest	6	0.805	

H_0: $\beta_1 = \beta_2 = 0$ of $\alpha_i = 0$, $\forall_i$ wordt verworpen

4.2. a)

Bron	ν	GK
fosfor	2	$24\frac{8}{9}$
glas	1	$533\frac{5}{9}$
fosfor * glas	2	$\frac{8}{9}$
rest	12	$2\frac{7}{9}$

b) Fosfor en glas significant.

4.3.

Bron	ν	GK	F
stad	2	2.521	3.66
bedrijfstak	2	27.485	39.9
interactie	4	20.301	29.5
rest	27	0.688	

De gemiddelde lonen per stad en per bedrijfstak zijn:

	H	I	D
SF	3.77	8.72	5.80
C	4.98	9.38	4.83
NY	9.24	6.49	5.27

De standaardafwijking van deze gemiddelden is 0.41 (nl. $\sqrt{\frac{0.688}{4}}$)

4.4. a)

Bron	ν	GK	F
V	2	269.79	238
P	2	188.54	166
Z	1	550.01	485
VP	4	27.15	23.9
VZ	2	27.51	24.3
PZ	2	27.68	24.4
VPZ	4	8.74	7.71
Rest	54	1.13	

Opmerking: zie ook opgave 7.4.

b) $\underline{d} = \bar{\underline{y}}_{p_i v_j} - \bar{\underline{y}}_{p_{i'} j'}$

$\text{var } \underline{d} = \text{var } \bar{\underline{y}}_{p_i v_j} + \text{var } \bar{\underline{y}}_{p_{i'} j'} = \frac{1}{8}\sigma_0^2 + \frac{1}{8}\sigma_0^2 = \frac{1}{4}\sigma_0^2$

c) $$\left.\begin{array}{l} GK_{\text{Blokken}} = 0.72\ , \quad \nu = 3 \\ GK_{\text{Rest}} = \hat{\sigma}_0^2 = 1.16\ , \quad \nu = 51 \end{array}\right\} \rightarrow F^3_{51} < 1$$

4.5. a) $\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{e}_{ijk}\ , \sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{2} \beta_j = \sum_{i=1}^{3} (\alpha\beta)_{ij} = \sum_{j=1}^{2} (\alpha\beta)_{ij} = 0$

of $\quad \underline{y} = \beta_0 + \underbrace{\beta_1 x_1 + \beta_2 x_2}_{\text{snelheid}} + \underbrace{\beta_3 x_3}_{\text{metaal}} + \underbrace{\beta_4 x_1 x_3 + \beta_5 x_2 x_3}_{\text{interactie}} + \underline{e}$

met geschikte codering!!

b)

Bron	ν	GK	F
snelheid	2	30.5	20.3
metaal	1	18.0	12
interactie	2	1.5	1
rest	12	1.5	

c) Snelheid en metaal zijn significant.

d) $\hat{\sigma}_0^2 = 1.5$, $\nu = 12$

	laag	matig	hoog
I	9	14	12
II	12	16	13

De standaardafwijking van de gemiddelde levensduren in bovenstaande tabel is 0.7 .

4.6. a) $\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \underline{e}_{ijk}$

methoden $i = 1, 2, 3$
lijm $j = 1, 2$
herhalingen $k = 1, 2$

$\sum_i \alpha_i = \sum_i \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$

$\underline{e}_{ijk} \sim N(0, \sigma_0^2)$ $\underline{e}_{ijk}$ o.o.

b)

Bron	ν	GK	F
methoden	2	9	1.38
lijm	1	48	7.38
interactie	2	3	< 1
rest	6	6½	

c) De lijmsoorten verschillen significant.

5. Codering en orthogonale polynomen

5.1. a) Bv. $x_1' = (\ 0,\ 0,\ 0,\ 0,\ 2,\ 2,\ 2,\ -1,\ -1,\ -1,\ -1,\ -1,\ -1)$

$x_2' = (-9,\ -9,\ -9,\ -9,\ 4,\ 4,\ 4,\ 4,\ 4,\ 4,\ 4,\ 4,\ 4)$

b) $$b = \begin{pmatrix} \bar{y}_{..} \\ \frac{1}{3}\ (\bar{y}_{2.} - \bar{y}_{3.}) \\ \frac{1}{13}\ (\overline{y_{2.} + y_{3.}} - \bar{y}_{1.}) \end{pmatrix} = \begin{pmatrix} 23.23 \\ 2.94 \\ -0{,}28 \end{pmatrix}$$

$$b'X'y = 7015.69 + \underline{156.06 + 36.67}$$
$$192.73 = KS_{tussen}$$

c) $\hat{\sigma}_0^{\ 2} = 42.558$, $\nu = 10$

d) $KS_{totaal} = 618.31$; $KS_{tussen} = 192.73$

e)

Bron	ν	GK	F
tussen typen	2	96.36	2.26
rest	10	42.56	

5.2. a) $\hat{y} = 0.7310 + 0.0254\ x - 0.000267\ x^2$

of $50(\hat{y} - 1.20) = 4.75 + 0.345\ \varphi_1 - 0.333\ \varphi_2$

$\hat{\sigma}_0^{\ 2} = 0.000385$, $\nu = 5$

b) $(1.5 * 10^{-4},\ 23.1 * 10^{-4})$

5.3. a) $\hat{y} = 196 + 34.1\ \varphi_1 + 5.86\ \varphi_2$; $\hat{\sigma}_0^{\ 2} = 2.02$, $\nu = 3$

b) $\left|\frac{x - \bar{x}}{10}\right| = \sqrt{\frac{111}{60}} \Rightarrow x_1 = 31.4$, $x_2 = 58.6$; voor $x = \bar{x}$ relatief maximum.

5.4. a) $\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \underline{e}$

$x_1 = \frac{1}{15}$ (druk - 65) $\qquad \underline{e} \sim N(0,\ \sigma_0^{\ 2}\ I)$

$x_2 = \frac{1}{100}$ (temp. - 200)

$x_3 = 3x_2^{\ 2} - 2$

b) $b' = (23.83,\ 0.50,\ 2.75,\ 0.42)$; $\hat{\sigma}_0^{\ 2} = 0.5$, $\nu = 2$

c) $\hat{y} = 23.83 + 2.75\ x_2$; $\hat{\sigma}_0^{\ 2} = 0.5$, $\nu = 2$ of $\hat{\sigma}_0^{\ 2} = 1.15$, $\nu = 4$

d) $\mathcal{E}\underline{\hat{y}} \in (21.00,\ 23.92)$ $(\nu = 2)$ of $\mathcal{E}\underline{\hat{y}} \in (21.04,\ 23.88)$ $(\nu = 4)$

5.5. a) $x_1 = \frac{C - 0.65}{0.05}$; $x_2 = \frac{T - 200}{20}$; $x_3 = x_1 x_2$; $x_4 = \frac{x_1^2 - 5}{4}$

b) $b' = (10, 4, 5, 3, 2)$; $\hat{\sigma}_0^2 = 14\frac{2}{3}$, $\nu = 3$

c) $\hat{\sigma}_0^2 = 13.5$, $\nu = 10$; $t_{10} = 1.54$ of $F_{10}^1 = 2.37$

d) $\mathcal{E}\underline{\hat{y}} \in (16.64, 27.36)$

5.6. a) $x_1 = \frac{C - 6}{2}$; $x_2 = \frac{T - 200}{25}$; $x_3 = 3x_2^2 - 2$

b) $b' = (17, 2, 12, -1)$; $\hat{\sigma}_0^2 = 4$, $\nu = 8$

c) Alle nulhypothesen worden verworpen.

d) $\mathcal{E}\underline{\hat{y}} \in (15.6, 20.4)$

6. 2^n-proeven

6.1. a)

	I	A	B	AB	C	AC	BC	ABC
Effect	$24\frac{1}{2}$	6	$3\frac{1}{2}$	1	4	$\frac{1}{2}$	0	$\frac{1}{2}$
KS	4802	288	98	8	128	2	0	2

b) $\hat{\sigma}_0^2 = 3$, $\nu = 4$; significant als KS > 23.1 ($\alpha = 0.05$) dus A, B en C.

c) (26.6, 33.4)

6.2. $A = T_1$, $B = P$, $C = K$, $D = T_2$

a)

	I	A	B	AB	C	AC	BC	ABC
Effect	20.50	−2.51	−1.84	−0.60	−0.79	−1.00	−0.08	0.66
KS	6724.0	101.0	54.0	5.8	9.9	16.0	0.1	7.0

	D	AD	BD	ABD	CD	ACD	BCD	ABCD
Effect	0.65	0.19	0.41	0.85	−1.16	0.38	−0.50	0.84
KS	6.8	0.6	2.7	11.6	21.6	2.2	4.0	11.2

b) $\hat{\sigma}_0^2 = 7.21$, $\nu = 5$; significant als KS > 47.7 ($\alpha = 0.05$) dus T_1 en P.

6.3. a) $\underline{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3 +$
$+ \beta_8 x_4 + \underline{e}$

$x_1 = \begin{cases} -1, & \text{A laag} \\ +1, & \text{A hoog} \end{cases}$ $\quad x_2 = \begin{cases} -1, & \text{B laag} \\ +1, & \text{B hoog} \end{cases}$ $\quad x_3 = \begin{cases} -1, & \text{C laag} \\ +1, & \text{C hoog} \end{cases}$

$x_4 = \begin{cases} -1, & \text{blok I} \\ +1, & \text{blok II} \end{cases}$

$\underline{e} \sim N(0, \sigma_0^2 I)$

b)

	I	A	B	AB	C	AC	BC	ABC	Blokken
Effect	$36\frac{1}{2}$	$-5\frac{7}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	3	$-\frac{5}{8}$	$\frac{3}{8}$	0	$1\frac{1}{8}$
KS	21316	$552\frac{1}{4}$	$\frac{1}{4}$	4	144	$6\frac{1}{4}$	$2\frac{1}{4}$	0	$20\frac{1}{4}$

c) $\hat{\sigma}_0^2 = \frac{11}{32}$, $\nu = 8$; significant als KS > 1.83 ($\alpha = 0.05$) dus B, ABC niet.

d) $\hat{y} = 36\frac{1}{2} - 5\frac{7}{8} x_1 + 3x_3 + \frac{1}{2} x_1 x_2 - \frac{5}{8} x_1 x_3 + \frac{3}{8} x_2 x_3 - 1\frac{1}{8} x_4$

e) $\widehat{\text{var}}(\underline{\hat{\beta}}_i) = \frac{1}{16} \hat{\sigma}_0^2 = 0.021$, $\nu = 8$

6.4. a)

	I	A	B	AB	C	AC	BC	ABC
Effect	$100\frac{1}{8}$	$17\frac{5}{8}$	$-1\frac{1}{2}$	$-1\frac{1}{2}$	$1\frac{7}{8}$	$1\frac{1}{8}$	$-4\frac{1}{2}$	$1\frac{1}{4}$
KS	$160400\frac{1}{4}$	$4970\frac{1}{4}$	36	36	$56\frac{1}{4}$	$20\frac{1}{4}$	324	25

$\hat{\sigma}_0^{\,2} = 200\frac{1}{4}$, $\nu = 8$; A significant.

b) $KS_H = 1056\frac{1}{4}$, $\nu = 1$; $\hat{\sigma}_0^{\,2} = 77.96$, $\nu = 7$

A en H significant.

c) Volledige 2^4-analyse: $\hat{\sigma}_0^{\,2} = 35.85$, $\nu = 5$; A, H en BC significant.

d) Kies model volgens b)

e) $\hat{\sigma}_0^{\,2} = 78$, $\nu = 7$; A en H significant.

f) $\hat{y} = 100\frac{1}{8} + 17\frac{5}{8}\,x_A + 8\frac{1}{8}\,x_H$

$$x_A = \begin{cases} -1, & \text{A laag} \\ +1, & \text{A hoog} \end{cases} \qquad x_H = \begin{cases} -1, & H_1 \\ +1, & H_2 \end{cases}$$

$\widehat{\text{var}}\,(\hat{\underline{\beta}}_A) = \frac{1}{16}\,\hat{\sigma}_0^{\,2} = 4.9$

6.5. a) $\nu = 8$

b) $t_8 = 1.2$

c) (2.42, 3.58)

d) i) $\nu = 7k - 7$

ii) $k = 13$

6.6. a) hoofdblok: (1), bc, de, bcde, abd, acd, abe, ace

blok 2 : a, abc, ade, abcde, bd, cd, be, ce

blok 3 : b, c, bde, cde, ad, abcd, ae, abce

blok 4 : d, bcd, e, bce, ab, ac, abde, acde

b) $KS_{b1} = KS_{ABC} + KS_{ADE} + KS_{BCDE}$, $\nu = 3$

c) $\nu = 13$

6.7. a) hoofdblok : (1), bd, ce, bcde, abc, ade, acd, abe

blok 2 : a, abd, ace, abcde, bc, de, cd, be

blok 3 : b, d, bce, cde, ac, abde, abcd, ae

blok 4 : c, bcd, e, bde, ab, acde, ad, abce b)

	ABD	ACE	BCDE
hoofdblok	-	-	+
blok 2	+	+	+
blok 3	+	-	-
blok 4	-	+	-

c)

<u>Bron</u>	<u>ν</u>
hoofdeffecten	5
interacties	10
blokken	3
rest	13

6.8. a) I = ABCE = BCDF = ADEF

b)

stand. volgorde	waar-neming	effect	KS	verstrengelde effecten
(1)	(1)	100.75	162409	I (= ABCE = BCDF = ADEF)
a	ae	17.00	4624	A (= BCE = ABCDF = DEF)
b	bef	-2.125	72.25	B (= ACE = CDF = ABDEF)
ab	abf	-0.875	12.25	AB = CE (= ACDF = BDEF)
c	cef	2.25	81	C (= ABE = BDF = ACDEF)
ac	acf	0.25	1	AC = BE (= ABDF = CDEF)
bc	bc	-5.375	462.25	BC = AE = DF (= ABCDEF)
abc	abce	1.625	42.25	E (= ABC = BCDEF = ADF)
d	df	8.75	1225	D (= ABCDE = BCF = AEF)
ad	adef	2.5	100	AD = EF (= BCDE = ABCF)
bd	bde	0.625	6.25	BD = CF (= ACDE = ABEF)
abd	abd	3.375	182.25	(ABD = CDE = ACF = BEF)
cd	cde	4	256	CD = BF (= ABDE = ACEF)
acd	acd	-2	64	(ACD = BDE = ABF = CEF)
bcd	bcdf	-1.375	30.25	F (= ABCEF = BCD = ADE)
abcd	abcdef	-0.375	2.25	DE = AF (= ABCD = BCEF)

Samenvoegen van interacties geeft $\hat{\sigma}_0^2 = 121$, $\nu = 9$

A en D zijn significant.

c) $\hat{y} = 100\frac{3}{4} + 17x_A + 8\frac{3}{4}\, x_D$

$$x_A = \begin{cases} -1, & \text{A laag} \\ +1, & \text{A hoog} \end{cases} \qquad x_D = \begin{cases} -1, & \text{D laag} \\ +1, & \text{D hoog} \end{cases}$$

$\hat{\sigma}_0^{\;2} = 121$, $\nu = 9$

6.9. a) blok 1: a, b, ade, bde

blok 2: cd, abcd, ce, abce

b) i) I = ABCE

ii)

stand. volgorde	waarneming	KS	verstrengelde effecten
(1)	(1)	$4422\frac{1}{4}$	I (= ABCE)
a	ae	16	A (= BCE)
b	be	49	B (= ACE)
ab	ab	$2\frac{1}{4}$	AB = CE
c	ce	1	C (= ABE)
ac	ac	$\frac{1}{4}$	AC = BE
bc	bc	$6\frac{1}{4}$	BC = AE
abc	abce	0	E (= ABC)
d	d	16	D (= ABCDE)
ad	ade	$6\frac{1}{4}$	AD (= BCDE)
bd	bde	$6\frac{1}{4}$	BD (= ACDE)
abd	abd	4	(ABD = CDE)
cd	cde	$\frac{1}{4}$	CD (= ABDE)
acd	acd	25	(ACD = BDE)
bcd	bcd	1	(BCD = ADE)
abcd	abcde	$12\frac{1}{4}$	DE (= ABCD)

$\hat{\sigma}_0^{\;2} = 10$, $\nu = 3$; geen significante effecten.

6.10. a) blok 1: (1), abcd, abeg, cdeg, abf, cdf, efg, abcdefg

blok 2: bcef, adef, acfg, bdfg, adg, bcg, bde, ace.

b) Geen.

6.11. a) I = ABCD

b) AB = CD = Blokken

c)

stand. volgorde	waarneming	KS	verstrengelde effecten
(1)	(1)	1168.86	I (= ABCD)
a	ab	0.28	A (= BCD)
c	bc	45.60	C (= ABD)
ac	ac	0.10	AC = BD
d	bd	1.90	D (= ABC)
ad	ad	0.01	AD = BC
cd	cd	0.78	CD = AB = Blokken
acd	abcd	12.75	B (= ACD)

d) $\hat{\sigma}_0^2 = 0.056$, $\nu = 2$; B, C en D significant.

6.12. a) I = -ABC = -CDE = ABDE

b)

stand. volgorde	waarneming	KS	verstrengelde effecten
(1)	(1)	512	I (= -ABC = -CDE = ABDE)
a	ace	128	A (= -BC = -ACDE = BDE)
b	bce	512	B (= -AC = -BCDE = ADE)
ab	ab	8	-C (= AB = -ABCDE = DE)
d	de	0	D (= -ABCD = -CE = ABE)
ad	acd	8	(AD = -BCD = -ACE = BE)
bd	bcd	8	(BD = -ACD = -BCE = AE)
abd	abde	8	E (= ABD = -CD = -ABCE)

c) $\hat{\sigma}_0^2 = 8$, $\nu = 2$; B is significant, A bijna significant.

6.13. a)

	I	M	T	MT	H	MH	TH	MTH
KS	$360\frac{3}{8}$	$22\frac{1}{24}$	$\frac{1}{24}$	$2\frac{1}{24}$	$15\frac{1}{24}$	$\frac{3}{8}$	$2\frac{1}{24}$	$1\frac{1}{24}$

$\hat{\sigma}_0^2 = 2$, $\nu = 16$

b) M en H zijn significant (KS > 8.98)

c) $\hat{y} = 3\frac{7}{8} - \frac{23}{24} x_M + \frac{19}{24} x_H$

$$x_M = \begin{cases} -1, & \text{continu} \\ +1, & \text{onderbroken} \end{cases} \qquad x_H = \frac{\alpha - 22.5}{7.5}$$

d) $\hat{y} = 5.10$

6.14. a) I = ABCDE

b)

hoofdeffecten	5	vr. gr.
interacties	7	vr. gr.
blokeffect	3	vr. gr.
totaal	15	vr. gr.

Er zijn dus geen vrijheidsgraden over voor de restvariantie.

c) A, C, CD en Blokken zijn significant.

d) $KS_{bl} = KS_{AB} + KS_{BC} + KS_{AC} = 224$

e) $\hat{y} = 60 + 10x_A + 7x_C + 3x_{CD}$

6.15. b) AB, CD en ABCD

c)

Bron	ν	GK
A	1	4
B	1	$20\frac{1}{4}$
C	1	81
D	1	$6\frac{1}{4}$
AC	1	1
BC	1	$2\frac{1}{4}$
AD	1	$2\frac{1}{4}$
BD	1	1
Blokken	3	$1\frac{1}{6}$
Rest	4	$\frac{5}{8}$

B, C en D zijn significant

6.16. a)

stand. volgorde	waarneming	$KS \cdot 10^2$	verstrengelde effecten
(1)	d	$60\frac{1}{2}$	I (= -ABCD)
a	a	8	A (= -BCD)
b	b	$\frac{1}{2}$	B (= -ACD)
ab	abd	0	AB = -CD
c	c	18	C (= -ABD)
ac	acd	$4\frac{1}{2}$	AC = -BD
bc	bcd	2	BC = -AD
abc	abc	$4\frac{1}{2}$	-D (= ABC)

bij codering $y = (x - 4.7).10$

b) A en C zijn significant (KS > 0.0496)

c) $\hat{y}_{max} = 5.225$; $\hat{y}_{min} = 4.725$

6.17. a) I = ABCD = CDEF = ABEF

ACE (= BDE = ADF = BCF) is met blokken verstrengeld.

c)

hoofdeffecten	6	vr. gr.
interacties	7	vr. gr.
blokken	1	vr. gr.
rest	1	vr. gr.
totaal	15	vr. gr.

6.18. a) I = - ABCD

b) B en D

c) $\hat{\sigma}_0^2 = 15.14$, $\nu = 21$; $\mathcal{E}(\widehat{\underline{bd}}) \in (6.1,\ 11.9)$

6.19. a) Bv. I = ABCD = ABEF = CDEF

6.20. i) blok 1: c, abd, abcef, def, be, bcdf, af, acde

blok 2: abc, d, cef, abdef, ae, acdf, bf, bcde

ii) $\mathcal{E}\underline{\hat{y}} \in (25.7,\ 38.3)$

6.21. a) $\underline{y} = \beta_0 + \beta_1 x_A + \beta_2 x_B + \beta_3 x_C + \beta_4 x_A x_B + \beta_5 x_A x_C + \beta_6 x_B x_C +$

$+ \beta_7 x_A x_B x_C + \beta_8 x_H + \underline{e}$, $\qquad \underline{e} \sim N(0, \sigma_0^2 I)$

$x_A = \begin{cases} -1, & \text{aan zee} \\ +1, & \text{niet aan zee} \end{cases} \qquad x_B = \begin{cases} -1, & \text{Noord} \\ +1, & \text{Zuid} \end{cases}$

$x_C = \begin{cases} -1, & \text{januari} \\ +1, & \text{juli} \end{cases} \qquad x_H = \begin{cases} -1, & 1974 \\ +1, & 1975 \end{cases}$

b) $b' = (10, -\frac{1}{8}, \frac{1}{2}, 7\frac{5}{8}, -\frac{1}{8}, \frac{7}{8}, 0, 0, \frac{3}{8})$;

$\hat{\sigma}_0^2 = \frac{3}{28}$, $\nu = 7$

c) B, C, AC en H zijn significant (KS > 0.60)

d)

		aan zee	niet aan zee
Noord	jan.	$2\frac{3}{4}$	1
	juli	$16\frac{1}{4}$	18
Zuid	jan.	$3\frac{3}{4}$	2
	juli	$17\frac{1}{4}$	19

6.22. a)

$$\text{var } \underline{b} = \frac{1}{48}\begin{pmatrix} 24 & 0 & 0 & 0 & -6 & -6 \\ 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ -6 & 0 & 0 & 0 & 3 & 1 \\ -6 & 0 & 0 & 0 & 1 & 3 \end{pmatrix}\sigma_0^{\,2}$$

In een punt (x_1, x_2) geldt:

$$\text{var } \hat{\underline{y}} = \text{var}\left(\underline{b}_0 + \underline{b}_1 x_1 + \underline{b}_2 x_2 + \underline{b}_3 x_1 x_2 + \underline{b}_4 x_1^{\,2} + \underline{b}_5 x_2^{\,2}\right)$$

$$= \frac{1}{48}\left\{ 24 - 8\left(x_1^{\,2} + x_2^{\,2}\right) + 3\left(x_1^{\,2} + x_2^{\,2}\right)^2 \right\}\sigma_0^{\,2}$$ is constant voor punten op gelijke afstand van het midden.

b) $\min(\text{var } \hat{\underline{y}}) = \frac{7}{18}\sigma_0^{\,2}$ voor $x_1^{\,2} + x_2^{\,2} = \frac{4}{3}$

c) $\max(\text{var } \hat{\underline{y}}) = \frac{5}{6}\sigma_0^{\,2}$ voor $x_1^{\,2} + x_2^{\,2} = 4$

6.23. a)

$$\text{var } \underline{b} = \frac{1}{32}\begin{pmatrix} 8 & 0 & 0 & -4 & 0 & -4 \\ 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ -4 & 0 & 0 & 5 & 0 & 1 \\ 0 & 0 & 0 & 0 & 8 & 0 \\ -4 & 0 & 0 & 1 & 0 & 5 \end{pmatrix}\sigma_0^{\,2}$$

b) $\text{var } \hat{\underline{y}}_{(0,0)} = \frac{1}{4}\sigma_0^{\,2}$; $\text{var } \hat{\underline{y}}_{(1,0)} = \frac{9}{32}\sigma_0^{\,2}$; $\text{var } \hat{\underline{y}}_{(1,1)} = \frac{5}{8}\sigma_0^{\,2}$

c) $\text{var } \hat{\underline{y}}_{(x_1,x_2)} = \frac{1}{32}\left(8 - 4r^2 + 5r^4\right)\sigma_0^{\,2}$, waarbij $r^2 = x_1^{\,2} + x_2^{\,2}$

7. Gemengde modellen met gekruiste en geneste factoren

7.1. a)

provincie	A	$a = 30$
gemeente	B:A	$b = 5$
boerderij	C:AB	$c = 10$

$$\underline{y}_{ijk} = \mu + \alpha_i + \underline{\beta}_{j:i} + \underline{\gamma}_{k:ij} + \underline{e}_{ijk}$$

$$\Sigma\alpha_i = \mathcal{E}\underline{\beta}_{j:i} = \mathcal{E}\underline{\gamma}_{k:ij} = \mathcal{E}\underline{e}_{ijk} = 0$$

$$\text{var } \underline{\beta}_{j:i} = \sigma^2_{B:A}, \quad \text{var } \underline{\gamma}_{k:ij} = \sigma^2_{C:AB}, \quad \text{var } \underline{e}_{ijk} = \sigma_0^2$$

Alle stochasten onderling onafhankelijk.

b)

Bron	ν
A	$a - 1 = 29$
B:A	$a(b-1) = 120$
C:AB	$ab(c-1) = 1350$

c) I

$$\mathcal{E}(\underline{GK}_A) = 50\,\sigma_A^2 + 10\,\sigma^2_{B:A} + \sigma^2_{C:AB} + \sigma_0^2$$

$$\mathcal{E}(\underline{GK}_{B:A}) = 10\,\sigma^2_{B:A} + \sigma^2_{C:AB} + \sigma_0^2$$

$$\mathcal{E}(\underline{GK}_{C:AB}) = \sigma^2_{C:AB} + \sigma_0^2$$

$b^* = c^* = 1$

II

$$\mathcal{E}(\underline{GK}_A) = 50\,\sigma_A^2 + \sigma^2_{C:AB} + \sigma_0^2$$

$$\mathcal{E}(\underline{GK}_{B:A}) = 10\,\sigma^2_{B:A} + \sigma^2_{C:AB} + \sigma_0^2$$

$$\mathcal{E}(\underline{GK}_{C:AB}) = \sigma^2_{C:AB} + \sigma_0^2$$

$b^* = 0, \; c^* = 1$

7.2. a)

$$\underline{y}_{ijk} = \mu + \alpha_i + \underline{\beta}_j + \underline{\gamma}_{k:j} + \underline{e}_{ijk}$$

of

$$\underline{y}_{ijk} = \mu + \underline{\alpha}_i + \underline{\beta}_j + \underline{\gamma}_{k:j} + \underline{e}_{ijk}$$

$i = 1, 2$	(H)	$a = 2$	$a^* = 0$ of 1
$j = 1, 2, 3$	(P)	$b = 3$	$b^* = 1$
$k = 1, 2, 3$	(M)	$c = 3$	$c^* = 1$

onderstellingen: $\mathcal{E}\underline{e}_{ijk} = 0$, var $\underline{e}_{ijk} = \sigma_0^2$, alle stochasten 0.0.

$(\Sigma\alpha_i = 0 \; , \; \frac{\Sigma\alpha_i^2}{a-1} = \sigma_A^2)$ of $(\mathcal{E}\underline{\alpha}_i = 0 \; , \; \text{var } \underline{\alpha}_i = \sigma_A^2)$

$\mathcal{E}\underline{\beta}_j = 0$, var $\underline{\beta}_j = \sigma_B^2$

$\mathcal{E}\underline{\gamma}_{k:j} = 0$, var $\underline{\gamma}_{k:j} = \sigma_{C:B}^2$

b) σ_A^2 , σ_B^2 , $\sigma_{C:B}^2$ en σ_0^2

c)

Bron	ν	$GK.10^2$
A	1	1027.6
B	2	291.0
C:B	6	355.1
Rest	8	66.2

d)

	σ_A^2	σ_B^2	$\sigma_{C:B}^2$	σ_0^2
A	9	0	0	1
B		6	2	1
C:B			2	1

e) A : $F_8^1 = 15.52$

B : $F_6^2 < 1$

C:B : $F_8^6 = 5.36$

Conclusies: analysefout $\hat{\sigma}_0^2 = 0.66$, $\nu = 8$

monsternemingsfout $\hat{\sigma}_{C:B}^2 = 1.44$

$a^* = 0 \quad H_1 = 22.83 \quad H_2 = 24.34$

of $a^* = 1 \quad \hat{\sigma}_A^2 = 1.07$

7.3. a) glasblazer A , $a = 4$, $a^* = 0$

waarnemer B , $b = 3$, $b^* = 0$

plaatje C:A , $c = 3$, $c^* = 1$

$\underline{y}_{ijk} = \mu + \alpha_i + \beta_j + \underline{\gamma}_{k:i} + (\alpha\beta)_{ij} + \underline{e}_{ijk}$

met de nodige onderstellingen.

c)

Bron	ν	GK	$\mathcal{E}$(GK)	Toets
A	3	21.04	$9\,\sigma_A^2 + 3\,\sigma_{C:A}^2 + \sigma_0^2$	$F_8^3 < 1$
B	2	21.20	$12\,\sigma_B^2 + \sigma_0^2$	$F_{16}^2 = 4.83$
C:A	8	51.47	$3\,\sigma_{C:A}^2 + \sigma_0^2$	$F_{16}^8 = 11.7$
AB	6	3.12	$3\,\sigma_{AB}^2 + \sigma_0^2$	$F_{16}^6 < 1$
Rest	16	4.39	σ_0^2	

$\hat{\mu} = 57.22 \qquad \hat{\sigma}_0^2 = 4.39\ , \quad \nu = 16$

$\hat{\beta}_1 = 1.11 \qquad \hat{\sigma}_{C:A}^2 = 15.7$

$\hat{\beta}_2 = 0.36$

$\hat{\beta}_3 = -1.47$

7.4. a) $v = p = 3\ , \quad z = 2\ , \ v^* = z^* = 0\ , \quad p^* = 1$

$\mathcal{E}(\underline{GK}(V)) = 24\,\sigma_V^2 + 8\,\sigma_{VP}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(P)) = 24\,\sigma_P^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(Z)) = 36\,\sigma_Z^2 + 12\,\sigma_{PZ}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(VP)) = 8\,\sigma_{VP}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(VZ)) = 12\,\sigma_{VZ}^2 + 4\,\sigma_{VPZ}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(PZ)) = 12\,\sigma_{PZ}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(VPZ)) = 4\,\sigma_{VPZ}^2 + \sigma_0^2$

b)

V tegen VP	F_4^2	= 9.92
P tegen σ_0^2	F_{54}^2	= 166.8
Z tegen PZ	F_2^1	= 19.9
VP tegen σ_0^2	F_{54}^4	= 24.1
VZ tegen VPZ	F_4^2	= 3.16
PZ tegen σ_0^2	F_{54}^2	= 24.5
VPZ tegen σ_0^2	F_{54}^4	= 7.70

7.5. thermometer T, $t = 4$
analisten A, $a = 3$
dagen D, $d = 2$

a) 1° D:TA ; de waarnemingen zijn uitgevoerd op 24 verschillende dagen
2° D:A ; de waarnemingen zijn uitgevoerd op 6 verschillende dagen
3° D ; de waarnemingen zijn uitgevoerd op 2 verschillende dagen

De variantie-analyse slaat op de eerste situatie.

b) $t^* = 1$, $a^* = 0$, $d^* = 1$

$\mathcal{E}(\underline{GK}(T)) = 6\,\sigma_T^2 + \sigma_{D:TA}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(A)) = 8\,\sigma_A^2 + 2\,\sigma_{TA}^2 + \sigma_{D:TA}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(TA)) = 2\,\sigma_{TA}^2 + \sigma_{D:TA}^2 + \sigma_0^2$

$\mathcal{E}(\underline{GK}(D:TA)) = \sigma_{D:TA}^2 + \sigma_0^2$

c) $KS_{(D:A)} = 1.59$, $\nu = 3$; $KS_{(DT:A)} = 1.54$, $\nu = 9$

$F_9^3 = 3.10$

d) analist I : $T = 3.19$
analist II : $T = 2.19$
analist III: $T = 3.06$

Ook vanuit practisch standpunt lijkt een interactie TA erg onwaarschijnlijk, zodat $GK_{(TA)}$ en $GK_{(D:TA)}$ samengevoegd kunnen worden.
De schatting voor de meetfout wordt dan: $\hat{\sigma}_0^2 = 0.33$, $\nu = 18$
De schatting voor de variantie tussen thermometers wordt: $\hat{\sigma}_T^2 = 0.23$

7.6. a) $\underline{y}_{ijk} = \mu + \underline{\alpha}_i + \underline{\beta}_{j:i} + \underline{e}_{ijk}$

$\mathcal{E}\underline{\alpha}_i = 0$ (B) var $\underline{\alpha}_i = \sigma_B^2$ var $\underline{e}_{ijk} = \sigma_0^2$

$\mathcal{E}\underline{\beta}_{j:i} = 0$ (M:B) var $\underline{\beta}_{j:i} = \sigma_{M:B}^2$ $\underline{e} \sim N(0, \sigma_0^2\, I)$ $\underline{\alpha}$, $\underline{\beta}$ en $\underline{e}$ o.o.

b)

<u>Bron</u>	<u>ν</u>	<u>GK</u>	$\mathcal{E}(\underline{GK})$
B	3	191	$6\,\sigma_B^2 + 2\,\sigma_{M:B}^2 + \sigma_0^2$
M:B	8	33.4	$2\,\sigma_{M:B}^2 + \sigma_0^2$
H:MB	12	14.2	σ_0^2

Alleen B is significant.

c) $\hat{\mu} = 7.33$; $\hat{\sigma}_0^2 = 14.2 * 10^{-4}$, $\nu = 12$; $\hat{\sigma}_B^2 = 26.3 * 10^{-4}$.

7.7. a) I $\underline{y}_{ij} = \mu + \alpha_i + \beta_j + \underline{e}_{ij}$; $i = 1, 2, 3$; $j = 1, 2, 3, 4$

A: analisten $\Sigma\alpha_i = \Sigma\underline{\beta}_j = 0$, $\underline{e} \sim N(0, \sigma_0^2 I)$

B: partijen

II $\underline{y}_{ij} = \mu + \alpha_i + \underline{\beta}_j + \underline{e}_{ij}$; $i = 1, 2, 3$; $j = 1, 2, 3, 4$

A: analisten $\Sigma\alpha_i = \mathcal{E}\underline{\beta}_j = 0$, $\underline{e} \sim N(0, \sigma_0^2 I)$

B: partijen var $\underline{\beta}_j = \sigma_B^2$

b)

Bron	ν	GK	$\mathcal{E}$(GK)
Analisten A	2	4	$\sigma_0^2 + 4\,\sigma_A^2$
Partijen B	3	29	$\sigma_0^2 + 3\,\sigma_B^2$
Rest	6	2	σ_0^2

B is significant.

c) $\hat{\sigma}_0^2 = 2$, $\nu = 6$; $\hat{\sigma}_B^2 = 9$

d) I : $\hat{\text{var}}\ \bar{\underline{y}} = \frac{1}{12}\hat{\sigma}_0^2 = \frac{1}{6}$

II: $\hat{\text{var}}\ \bar{\underline{y}} = \frac{1}{12}\hat{\sigma}_0^2 + \frac{1}{4}\hat{\sigma}_B^2 = 2\frac{5}{12}$

8. Gemengde opgaven

8.1. a) Doel is te bereiken dat de eventuele effecten van niet-interessante of onbeheerste factoren optreden als toevallige fouten in de waarnemingen.

b) Eventuele verschillen tussen blokken worden verwijderd uit de restvariantie om een scherper criterium te krijgen bij de beoordeling van de interessante factoren.

c) Alleen bij kwantitatieve factoren is het mogelijk het effect van de factor te splitsen in lineaire, kwadratische, etc., bijdragen (bij voldoende niveau's). Hierin ligt opgesloten dat bij 2 niveau's het onderscheid wegvalt.

8.2. a) Als de experimentele eenheden niet (voldoende) homogeen zijn tengevolge van onvermijdelijke factoren, wordt een proefopzet met blokvorming gekozen. Het afsplitsen van blokeffecten leidt tot een scherpere beoordeling van de interessante factoren.

b) De waarnemingen vormen onderling onafhankelijke aselecte steekproeven uit normaal verdeelde populaties met gelijke varianties.

c) Een lineair contrast is een lineaire combinatie van waarnemingen, zo dat de som der coefficiënten nul is. Keuze van orthogonale contrasten leidt tot gemakkelijke splitsing van de totale kwadratensom, waarbij tevens de schatters der modelparameters onderling onafhankelijk zijn.

d) Interactie van 2 (of meer) factoren: het effect van één der factoren is afhankelijk van het niveau van de andere factor (resp. niveaucombinatie der andere factoren).

8.3. a) De waarnemingen vormen onderling onafhankelijke aselecte steekproeven uit normaal verdeelde populaties met gelijke (onbekende) varianties.

b) $2^n(m - 1)$

c) Als voor ieder paar factoren geldt dat de parameterschatters van de ene factor ongecorreleerd zijn met de parameterschatters van de andere factor. Gevolg: voor het toetsen of een effect van nul verschilt is het niet nodig een tweede lineair model aan te passen. De totale kwadratensom kan worden gesplitst in een aantal kwadratensommen, elk behorend bij één der factoren of factorencombinaties, en een restkwadratensom.

8.4. a) i) Als er geen interacties zijn. Daarnaast gelden uiteraard voorwaarden als normaliteit, enz.

ii) A: werkmethoden ; B: arbeiders

B is gekruist met A (A, B, AB) als iedere arbeider volgens elke methode werkt.

B is genest (A, B:A) als voor iedere werkmethode andere arbeiders worden genomen.

B : A moet worden getoetst tegen σ_0^2.

b) blok 1: a, c, abd, abe, ade, bcd, bce, cde
blok 2: b, d, e, abc, acd, ace, bde, abcde

8.5. a) $\underline{y} = \beta_0 + \underbrace{\beta_1 x_1}_{\text{per.}} + \underbrace{\beta_2 x_2}_{\text{beh.}} + \underbrace{\beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5}_{\text{pat.}} + \underline{e}$

$x_1 = \begin{cases} -1, & \text{periode I} \\ +1, & \text{periode II} \end{cases} \qquad x_2 = \begin{cases} -1, & \text{behandeling a} \\ +1, & \text{behandeling b} \end{cases}$

x_3 = 1 voor patiënt (2), anders 0
x_4 = 1 voor patiënt (3), anders 0
x_5 = 1 voor patiënt (4), anders 0

$\underline{e} \sim N(0, \sigma_0^2 I)$

b) $b' = (22\frac{1}{2}, 1, -3, -12, 8, -2)$; $\hat{\sigma}_0^2 = 13$, $\nu = 2$

c) $t_2 = -2{,}35$ of $F_2^1 = 5\,\frac{7}{13}$

8.6. a) $F_6^2 = \frac{33}{38}$

b) $F_8^1 = 22.0$

8.7. a) $b' = (4, 2\frac{1}{2}, 1\frac{1}{2}, 1\frac{1}{2}, -\frac{1}{2})$; $\hat{\sigma}_0^2 = 2$, $\nu = 1$

b) $\hat{\sigma}_0^2 = 1.4$, $\nu = 10$

c) $\hat{y} = 4 + 2\frac{1}{2}\, x_1 + 1\frac{1}{2}\, x_2 + 1\frac{1}{2}\, x_1 x_2$

d) $\mathcal{E}\underline{\hat{y}} \in (3.8, 7.2)$

8.8. a)

	I	C	S	CS	T	CT	ST	CST
KS:	1152	32	32	0	128	32	8	8

i) T is significant (KS > 92.52)

ii) I = CST : c, s, t, cst
of I = -CST : (1), cs, ct, st

b) i) $\frac{1}{4}$

ii) Bv. I = ABCD = -ABE = -CDE

iii) Bij de definiërende contrasten van ii) horen de volgende waarnemingen: (1), ab, cd, abcd, ace, bce, ade, bde.

8.9. a) $\mathcal{E}\underline{\hat{\gamma}}_1 = \mathcal{E}\{\frac{1}{3}(\underline{y}_{111} - \underline{y}_{133} - \underline{y}_{212} + \underline{y}_{231})\} = \gamma_1$

$\underline{\hat{\gamma}}_2 = \frac{1}{3}(\underline{y}_{122} - \underline{y}_{111} - \underline{y}_{223} + \underline{y}_{212})$

$\underline{\hat{\gamma}}_3 = \frac{1}{3}(\underline{y}_{223} - \underline{y}_{122} - \underline{y}_{231} + \underline{y}_{133})$

b) var $\underline{\hat{\gamma}}_1 = \frac{16}{9}$, $u = -2\frac{1}{4}$

c) $X^2 = 5\frac{1}{4}$, $\nu = 2$

AANVULLING BIJ COLLEGEDICTAAT STATISTISCHE THEORIE VAN PROEFOPZETTEN, Najaar 1981.

Alternatieve uitwerking van voorbeeld 3.2.1, pag.16 e.v.

$$\underline{y} = 50 - 2x + 0{,}04x^2 + \underline{e} \quad , \quad \underline{e} \quad N(0,4) \; .$$

Codering: $x_0 = 1$, $x_1 = (x-35)/5$, $x_2 = x_1^2$, $z = y - 20$.

$$\underline{z} = 9 + 4x_1 + x_2 + \underline{e} \; .$$

De waarnemingen zijn:

x_1	-3	-2	-1	0	1	2	3
z	5,8	8,1	4,6	12,4	16,8	21,3	26,0

$$X = \begin{pmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \qquad X'X = \begin{pmatrix} 7 & 0 & 28 \\ 0 & 28 & 0 \\ 28 & 0 & 196 \end{pmatrix} = 7 \begin{pmatrix} 1 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 28 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{84} \begin{pmatrix} 28 & 0 & -4 \\ 0 & 3 & 0 \\ -4 & 0 & 1 \end{pmatrix}$$

$$X'z = \begin{pmatrix} 95{,}0 \\ 99{,}2 \\ 425{,}2 \end{pmatrix} \qquad b = (x'X)^{-1}X'z = \frac{1}{84} \begin{pmatrix} 959{,}2 \\ 297{,}6 \\ 45{,}2 \end{pmatrix} = \begin{pmatrix} 111{,}42 \\ 3{,}54 \\ 0{,}538 \end{pmatrix} .$$

De geschatte relatie is dus

$$z = 11{,}42 + 3{,}54x_1 + 0{,}538x_1^2 \; .$$

De overeenstemming met de ware vergelijking is tamelijk slecht. Dit wordt veroorzaakt door de vrij grote negatieve waarde van e_7, nl. -4,0. Als we $\sigma_0 = 2$ bekend veronderstellen, kunnen we, op grond van de geschatte waarden, betrouwbaarheidsintervallen voor de β's opstellen.

$$\beta_0 = 11,42 \pm 1,96\sqrt{\frac{28}{84}} \ . \ 2 = 11,42 \pm 2.26 \ (\beta_0 = 9) \ .$$

$$\beta_1 = 3,54 \pm 1,96\sqrt{\frac{3}{84}} \ . \ 2 = 3,54 \pm 0,74 \ (\beta_1 = 4) \ .$$

$$\beta_2 = 0.538 \pm 1,96\sqrt{\frac{1}{84}} \ . \ 2 = 0,538 \pm 0.428 \ (\beta_2 = 1) \ .$$

Alleen het interval voor β_1 bevat dus de ware waarde.

Uit $(X'X)^{-1}$ kunnen we ook de correlatie tussen de schatters $\underline{b}_0$ en $\underline{b}_2$ afleiden. Deze is

$$\rho(\underline{b}_0, \underline{b}_2) = \frac{-4}{\sqrt{28}} = -0,76 \ .$$

Hiermee in overeenstemming is het feit dat β_0 te hoog en β_2 te laag wordt geschat.

<u>Pag.22</u>

De restkwadratensom wordt als volgt berekend:

$$z'z = \sum z_i^2 = 1686,10 \ ,$$

$$b'X'z = \frac{1}{84}(959,2 \times 95,0 + 297,6 \times 99,2 + 45,2 \times 425,2) = 1665,06 \ .$$

(Voor het verkrijgen van een nauwkeurigere uitkomst is de factor $\frac{1}{84}$ apart gehouden.)

Dus

$$KS_r = 1686,10 - 1665,06 = 21,04 \ .$$

Het aantal vrijheidsgraden is n - (k+1) = 7 - 3 = 4, dus de schatting van $\hat{\sigma}_0^2$ is

$$\hat{\sigma}_0^2 = \frac{21,04}{4} = 5,26 \quad ; \quad \hat{\sigma}_0 = 2,29 \ .$$

Pag.24

Hieruit volgt:

$$\hat{\sigma}^2(\underline{b}_0) = \frac{1}{84} \times 28 \times 5,26 = 1,75 \quad , \ \hat{\sigma}(\underline{b}_0) = 1,32 \quad ,$$

$$\hat{\sigma}^2(\underline{b}_1) = \frac{1}{84} \times 3 \times 5,26 = 0,188 \quad , \ \hat{\sigma}(\underline{b}_1) = 0,433 \ ,$$

$$\sigma^2(\underline{b}_2) = \frac{1}{84} \times 5,26 = 0,0626 \quad , \quad \hat{\sigma}(\underline{b}_2) = 0,250 \ .$$

Het betrouwbaarheidsinterval met 95% betrouwbaarheid voor β_0 is ($t_4(0,05) = 2,78$):

$$11,42 - 2,78 \times 1,32 < \beta_0 < 11,42 + 2,78 \times 1,32$$

of

$$7,75 < \beta_0 < 15,09 \quad \text{(ware waarde: } \beta_0 = 9) \ .$$

Evenzo:

$$2,34 < \beta_1 < 4,74 \quad \text{(ware waarde: } \beta_1 = 4) \ ,$$

$$-0,16 < \beta_2 < 1,23 \quad \text{(ware waarde: } \beta_2 = 1) \ .$$

Alle drie de betrouwbaarheidsintervallen bevatten dus de ware waarde, maar ze zijn tamelijk ruim. Om nauwere betrouwbaarheidsintervallen te krijgen moeten meerdere waarnemingen worden uitgevoerd, of waarnemingen bij x-waarden die verder uit elkaar liggen.

Pag.26 .

Als illustratie nemen we voorbeeld 3.2.1.
Het aangepaste model is:

$$\underline{\hat{z}} = \underline{b}_0 + \underline{b}_1 x_1 + \underline{b}_2 x_1^2 \ .$$

Wat is de variantie van $\underline{\hat{z}}$ in het punt $x_1 = 1$?
We beschouwen dus

$$\ell = 1.\beta_0 + 1.\beta_1 + 1.\beta_2 \ .$$

De schatter $\underline{\hat{z}}$ heeft als variantie:

$$[1 \quad 1 \quad 1]\ (X'X)^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \sigma_0^2 =$$

$$[1 \quad 1 \quad 1]\ \frac{1}{84} \begin{pmatrix} 28 & 0 & -4 \\ -4 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \sigma_0^2 = \frac{24}{84}\,\sigma_0^2 = \frac{2}{7}\,\sigma_0^2 \ .$$