Variantie-analyse

3.1 Het twee-steekproevenprobleem

In Statistiek & kansrekening zijn vragen aan de orde geweest zoals “heeft invoering van nieuwe veiligheidsmaatregelen geleid tot een vermindering van verloren gegane werktijd door ongelukken?”. Twee steekproeven (betrekking hebbende op de situatie na resp. vóór invoering van de nieuwe maatregelen) werden met elkaar vergeleken en de hypothese van het gelijk zijn van de beide verwachtingen werd getoetst. Met het oog op generalisatie naar meer dan twee steekproeven, waarbij we statistisch willen onderzoeken of de optredende verwachtingen gelijk zijn, herhalen we kort de theorie en bewerken deze voor om tot generalisatie over te kunnen gaan.

Het kansmodel bij het twee-steekproevenprobleem luidt als volgt. De s.v.-en $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ zijn o.o. met $X_i \sim N(\mu, \sigma^2)$ $(i = 1, \ldots, m)$ en $Y_j \sim N(\nu, \sigma^2)$ $(j = 1, \ldots, n)$ waarbij μ, ν en σ^2 onbekende parameters zijn. De eerste steekproef is $(X_1, \ldots, X_m)$ en de tweede steekproef is $(Y_1, \ldots, Y_n)$. Het gaat hier en in het vervolg om een drietal veronderstellingen die aan de variantie-analyse ten grondslag liggen. We vermelden ze daarom nog eens expliciet.

1. De waarnemingen zijn alle **onafhankelijk** van elkaar.
2. De waarnemingen zijn **normaal** verdeeld.
3. De varianties in de verschillende steekproeven zijn **gelijk**.

Dit laatste kan eventueel tot stand gekomen zijn na een transformatie voor stabiele spreiding (zie hoofdstuk 2).

Om de nulhypothese

$$H_0 : \mu = \nu$$

te toetsen tegen

$$H_1 : \mu \neq \nu$$

nemen we als toetsingsgrootheid

$$T = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \text{ met } \bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i,\ \bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

$$\text{en } S = \sqrt{S^2} \text{ met } S^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2},$$

$$\text{waarbij } S_X^2 = \frac{\sum_{i=1}^{m}(X_i - \bar{X})^2}{m-1}, S_Y^2 = \frac{\sum_{j=1}^{n}(Y_j - \bar{Y})^2}{n-1}.$$

We verwerpen H_0 voor grote waarden van $|T|$, of, equivalent, voor grote waarden van

$$T^2 = \frac{(\bar{X} - \bar{Y})^2 / \left(\frac{1}{m} + \frac{1}{n}\right)}{S^2}.$$

In de teller van T^2 vergelijken we het gemiddelde van de ene steekproef met het gemiddelde van de andere steekproef.

In het voorbeeld uit Statistiek & kansrekening is in de eerste steekproef, die betrekking heeft op de periode nadat de nieuwe veiligheidsmaatregelen waren ingesteld, gemiddeld 91 uur werktijd verloren gegaan door ongelukken, terwijl in de periode voorafgaand aan de nieuwe veiligheidsmaatregelen (de tweede steekproef) een gemiddelde van 108 uur is waargenomen.

De tweede steekproef valt dus gemiddeld hoger uit. Bij deze vergelijking van de steekproeven beschouwen we beide steekproeven in hun geheel, gerepresenteerd door hun steekproefgemiddeldes. We spreken van de **variatie tussen de steekproeven**.

Ieder van beide steekproeven bestaat in dit voorbeeld uit 50 getallen, corresponderend met 50 maanden na, respectievelijk 50 maanden vóór invoering van de nieuwe veiligheidsmaatregelen. De getallen in de eerste steekproef zijn niet steeds gelijk aan de gemiddelde waarde 91. Er is variatie rond de 91, die we terugvinden in de steekproefvariantie 14.1^2 van de eerste steekproef. Ook binnen de tweede steekproef is er sprake van variatie rond, in dit geval, 108, tot uitdrukking komend in de steekproefvariantie 14.3^2 van de tweede steekproef. We spreken dan van de **variatie binnen de steekproeven**. Deze variatie binnen de steekproef vinden we terug in de noemer van de toetsingsgrootheid T^2, waar we S^2 tegenkomen, die is opgebouwd m.b.v. de steekproefvarianties S_X^2 en S_Y^2.

Hoe groter de variatie **tussen** de steekproeven, hoe groter, naar we mogen verwachten, de afwijking tussen de verwachtingen van beide steekproeven. Maar is een verschil van 17 tussen beide gemiddelden, zoals we hier aantreffen, groot? Hoe kleiner de variatie **binnen** de steekproeven, des te scherper tekent een verschil van 17 **tussen** de steekproeven zich af.

> We moeten de variatie **tussen** de steekproeven dus relateren aan de variatie **binnen** de steekproeven.

Dat is precies wat in de toetsingsgrootheid T^2 gebeurt! De variatie tussen de steekproeven $(\bar{X} - \bar{Y})^2$ wordt gedeeld door de variatie binnen de steekproeven S^2. De afwijking 17 wordt a.h.w. in S-eenheden gemeten.

We gaan nog een stapje verder. Er is nog meer aan de hand. De noemer S^2 is een (zuivere) schatter van de gemeenschappelijke variantie σ^2, ongeacht of de verwachtingen van beide steekproeven gelijk zijn of niet. (Immers S^2 maakt gebruik van $X_i - \bar{X}$ en $Y_j - \bar{Y}$.) Zijn de verwachtingen van beide steekproeven gelijk dan is ook de teller $(\bar{X} - \bar{Y})^2/(\frac{1}{m} + \frac{1}{n})$ een zuivere schatter van σ^2 en moeten teller en noemer dus ongeveer dezelfde waarde opleveren. Daarentegen als de verwachtingen van beide steekproeven ongelijk zijn, dan zal de teller als regel grotere waarden aannemen.

> Dit is het basis idee bij **variantie-analyse**. We vergelijken twee schatters van σ^2, twee bronnen van variaties. Als de nulhypothese juist is, zijn beide ongeveer gelijk. Is de alternatieve hypothese juist dan heeft de teller een verwachting groter dan σ^2, terwijl de noemer ook dan nog verwachting σ^2 heeft. Dit verschaft ons een meetinstrument (toetsingsgrootheid) om uit te maken of we de nulhypothese al dan niet moeten verwerpen.

Ter voorbereiding op generalisatie van twee steekproeven naar k-steekproeven gaan we over op de volgende herformulering. Definieer voor $i = 1, 2$ en $j = 1, \ldots, n_i$

$$X_{ij} = \mu + \alpha_i + U_{ij}$$

met

$$U_{11}, U_{12}, \ldots, U_{1n_1}, U_{21}, \ldots, U_{2n_2}$$

o.o. s.v.-en, ieder met een $N(0, \sigma^2)$-verdeling, en met

$$\frac{n_1}{n_1 + n_2}\alpha_1 + \frac{n_2}{n_1 + n_2}\alpha_2 = 0.$$

De “vertaling” naar de eerdere beschrijving is als volgt.

oude notatie	nieuwe notatie
m	n_1
n	n_2
$X_1, \ldots, X_m$	$X_{11}, \ldots, X_{1n_1}$
$Y_1, \ldots, Y_n$	$X_{21}, \ldots, X_{2n_2}$
μ	$\mu + \alpha_1$
ν	$\mu + \alpha_2$.

De eerste steekproef wordt dus weergegeven door $X_{11}, X_{12}, \ldots, X_{1n_1}$, de tweede steekproef door $X_{21}, X_{22}, \ldots, X_{2n_2}$. Een toe te voegen derde steekproef ter grootte n_3 krijgt dan de notatie $X_{31}, X_{32}, \ldots, X_{3n_3}$, etc.

De verwachting in de eerste steekproef bedraagt $\mu + \alpha_1$, in de tweede $\mu + \alpha_2$ met $\dfrac{n_1}{n_1+n_2}\alpha_1 + \dfrac{n_2}{n_1+n_2}\alpha_2 = 0$. De interpretatie hiervan is dat μ het gemeenschappelijke deel vertegenwoordigt en α_1 aangeeft hoeveel de eerste steekproef daarboven of onder zit, terwijl α_2 dit voor de tweede steekproef aanduidt. Omdat α_1 en α_2 relatieve waarden zijn t.o.v. de gemeenschappelijke μ, geldt $\dfrac{n_1}{n_1+n_2}\alpha_1 + \dfrac{n_2}{n_1+n_2}\alpha_2 = 0$. Immers, hiermee krijgen we dat de verwachting van het gemiddelde van alle waarnemingen, $E((X_{11}+\ldots+X_{1n_1}+X_{21}+\ldots X_{2n_2}/(n_1+n_2))$, gelijk wordt aan $\{n_1(\mu+\alpha_1)+n_2(\mu+\alpha_2)\}/(n_1+n_2) = \mu + (n_1\alpha_1+n_2\alpha_2)/(n_1+n_2) = \mu$, zodat μ inderdaad de gemeenschappelijke waarde is. Komt er een derde steekproef bij, dan krijgt die als verwachting $\mu + \alpha_3$. Er geldt in dat geval $\dfrac{n_1}{n_1+n_2+n_3}\alpha_1 + \dfrac{n_2}{n_1+n_2+n_3}\alpha_2 + \dfrac{n_3}{n_1+n_2+n_3}\alpha_3 = 0$.

De nulhypothese vermeldt dat beide steekproeven gelijke verwachting hebben. In de nieuwe notatie betekent dat dus

$$H_0 : \alpha_1 = \alpha_2 = 0,$$

want als $\alpha_1 = \alpha_2 = 0$, dan hebben beide steekproeven dezelfde verwachting, weergegeven in de nieuwe notatie met μ. Deze herformulering nodigt uit tot generalisatie naar k-steekproeven.

3.2 Het *k*-steekproevenprobleem

3.2.1 De *k*-steekproeventoets

Voorbeeld 3.2.1 Bij een softwarehouse wil men onderzoeken of er op grond van het opleidingsniveau van de medewerkers verschil te merken is in voor het werk relevante kennis. Er worden 15 medewerkers aselect uitgekozen en onderworpen aan een examen. Dit zijn de resultaten.

opleidings-niveau						
A	81	84	69	85	84	95
B	94	83	86	81	78	
C	88	89	78	85		

Het steekproefgemiddelde van de eerste steekproef (niveau A) is 83.0, van de tweede steekproef (niveau B) 84.4 en van de derde (niveau C) 85.0. Zijn deze verschillen significant bij toetsing met onbetrouwbaarheidsdrempel $\alpha = 0.05$? □

Om een vraag als in voorbeeld 3.2.1 te beantwoorden ontwerpen we eerst een kansmodel, waarbinnen we deze vraag kunnen stellen en oplossen. Dit **kansmodel** ziet er aldus uit. Definieer voor $i = 1, 2, \ldots, k$ en $j = 1, \ldots, n_i$

$$X_{ij} = \mu + \alpha_i + U_{ij}$$

met

$$U_{11}, U_{12}, \ldots, U_{1n_1}, U_{21}, \ldots, U_{2n_2}, \ldots, U_{k1}, \ldots, U_{kn_k}$$

o.o. s.v.-en, ieder met een $N(0, \sigma^2)$-verdeling, en met

$$\frac{n_1}{n}\alpha_1 + \frac{n_2}{n}\alpha_2 + \ldots + \frac{n_k}{n}\alpha_k = 0, \text{ waarbij } n = \sum_{i=1}^{k} n_i.$$

In voorbeeld 3.2.1 is $k = 3$, is 81 de realisatie van X_{11}, 84 de realisatie van X_{12} en bijv. 89 de realisatie van X_{32}. Dus X_{ij} is het examenresultaat van de j^e medewerker bij opleidingsniveau i, waarbij $i = 1$ opleidingsniveau A is, $i = 2$ opleidingsniveau B en $i = 3$ opleidingsniveau C.

Alleen de X_{ij}'s nemen we waar, de overige grootheden in het model zijn niet waarneembaar: de U_{ij}'s zijn niet waarneembaar en de parameters μ, $\alpha_1, \ldots, \alpha_k$ en σ^2 zijn onbekend.

De verwachting in de i^e steekproef is dus $\mu + \alpha_i$. Soms schrijven we voor $\mu + \alpha_i$ ook wel μ_i. De vraag: "zijn er geen verschillen in opleidingsniveau te merken?" vertaalt zich in het kansmodel tot de vraag of $\alpha_1, \ldots, \alpha_k$ alle 0 zijn. Immers, in dat geval hebben alle X_{ij}'s dezelfde verwachting, genoteerd met μ. Het antwoord op deze vraag in het kansmodel komt tot stand door de nulhypothese

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$$

te toetsen tegen het alternatief dat niet alle α's 0 zijn, d.w.z. tegen

$$H_1 : \alpha_i \neq 0 \text{ voor zekere } i.$$

Net als in het twee-steekproevenprobleem maken we een toetsingsgrootheid, waarbij we de **variatie tussen de steekproeven** vergelijken met de **variatie binnen de steekproeven**. We noteren het totale steekproefgemiddelde met

$$X_{\bullet\bullet} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}}{n} \text{ met } n = \sum_{i=1}^{k} n_i,$$

de totale steekproefomvang. De $\bullet$-notatie geeft steeds aan dat we een gemiddelde nemen, in dit geval zowel t.a.v. i als j.

In voorbeeld 3.2.1 is

$$x_{\bullet\bullet} = \frac{(81 + 84 + \ldots + 95) + (94 + \ldots + 78) + (88 + \ldots + 85)}{15} = 84.0$$

Het verschil tussen X_{ij} en $X_{\bullet\bullet}$ laat zich nu splitsen in twee stukken: het verschil van X_{ij} t.o.v. $X_{i\bullet}$ en het verschil van $X_{i\bullet}$ t.o.v. $X_{\bullet\bullet}$. Hierbij is

$$X_{i\bullet} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}.$$

In formule krijgen we

$$X_{ij} - X_{\bullet\bullet} = (X_{ij} - X_{i\bullet}) + (X_{i\bullet} - X_{\bullet\bullet}).$$

We kwadrateren linker- en rechterlid en sommeren over i en j.

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - X_{\bullet\bullet})^2 =$$

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left\{ (X_{ij} - X_{i\bullet})^2 + 2(X_{ij} - X_{i\bullet})(X_{i\bullet} - X_{\bullet\bullet}) + (X_{i\bullet} - X_{\bullet\bullet})^2 \right\}.$$

We bekijken de "dubbele-product-term". Er geldt

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} 2(X_{ij} - X_{i\bullet})(X_{i\bullet} - X_{\bullet\bullet}) = \sum_{i=1}^{k} 2(X_{i\bullet} - X_{\bullet\bullet}) \sum_{j=1}^{n_i} (X_{ij} - X_{i\bullet}) = 0,$$

want

$$\sum_{j=1}^{n_i} (X_{ij} - X_{i\bullet}) = \sum_{j=1}^{n_i} X_{ij} - n_i X_{i\bullet} = n_i X_{i\bullet} - n_i X_{i\bullet} = 0.$$

De "dubbele-product-term" valt dus weg, resulterend in

(3.2.1)
$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-X_{\bullet\bullet})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-X_{i\bullet})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i\bullet}-X_{\bullet\bullet})^2$$
$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-X_{i\bullet})^2 + \sum_{i=1}^{k} n_i(X_{i\bullet}-X_{\bullet\bullet})^2.$$

De **totale variatie** (van X_{ij} t.o.v. $X_{\bullet\bullet}$)

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-X_{\bullet\bullet})^2$$

is hiermee opgesplitst in de **variatie binnen de steekproeven** (van X_{ij} t.o.v. $X_{i\bullet}$)

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-X_{i\bullet})^2$$

en de **variatie tussen de steekproeven** (van $X_{i\bullet}$ t.o.v. $X_{\bullet\bullet}$)

$$\sum_{i=1}^{k} n_i(X_{i\bullet}-X_{\bullet\bullet})^2.$$

Bij de variatie binnen de steekproeven kijken we hoe de X_{ij}'s variëren rond $X_{i\bullet}$, bij de variatie tussen de steekproeven kijken we hoe de steekproeven in hun geheel (gerepresenteerd door $X_{i\bullet}$) variëren rond de gemeenschappelijke waarde $X_{\bullet\bullet}$.

We introduceren enkele nieuwe namen. Het k-steekproevenmodel staat ook bekend als het **één-factor model (one-way analysis of variance) met *k* niveaus**. In voorbeeld 3.2.1 is "opleidingsniveau" de factor. Deze factor heeft 3 niveaus: A, B en C. De variatie tussen de steekproeven noemen we de kwadraatsom veroorzaakt door de factor. We gebruiken de Engelse afkorting SS(factor), Sum of Squares due to the factor. Dus

$$\text{SS(factor)} = \sum_{i=1}^{k} n_i(X_{i\bullet}-X_{\bullet\bullet})^2.$$

Behalve deze variatie die we kunnen toeschrijven aan de aanwezigheid van de factor, blijft van de totale variatie dan nog over de variatie binnen de steekproeven, die niet verklaard wordt door de factor. Deze rest-variatie noemen we SS(error). Dus

$$\text{SS(error)} \ = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - X_{i\bullet})^2.$$

In het k-steekproevenprobleem worden ook wel de namen SS(between) voor SS(factor) en SS(within) voor SS(error) gebruikt. De totale variatie noemen we SS(total), dus $SS(\text{total}) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - X_{\bullet\bullet})^2$. De totale variatie noemen we SS(total), dus

$$\text{SS(total)} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - X_{\bullet\bullet})^2.$$

De basisrelatie (3.2.1) laat zich dus zo opschrijven

$$\text{(3.2.2)} \qquad \boxed{\text{SS(total)} = \text{SS(factor)} + \text{SS(error)}.}$$

(Overigens is SS(factor) de **tweede** term en SS(error) de **eerste** term in het rechterlid van (3.2.1).)

Voorbeeld 3.2.2 (vervolg van voorbeeld 3.2.1) We berekenen SS(factor) en SS(error) voor de gegevens van voorbeeld 3.2.1. We hebben $x_{1\bullet} = 83.0$, $x_{2\bullet} = 84.4$, $x_{3\bullet} = 85.0$ en $x_{\bullet\bullet} = 84.0$. De realisatie van SS(factor) is dus

$$6 \times (83.0 - 84.0)^2 + 5 \times (84.4 - 84.0)^2 + 4 \times (85.0 - 84.0)^2 = 10.8.$$

Berekening van SS(total) levert op 584, zodat de waarde van SS(error) gelijk is aan $584 - 10.8 = 573.2$. □

De toetsingsgrootheid moet de variatie tussen de steekproeven vergelijken met de variatie binnen de steekproeven. Om dit goed te doen berekenen we de verwachtingen van SS(factor) en SS(error). Er geldt

$$\text{(3.2.3)} \quad E\{\text{SS(error)}\} = \sum_{i=1}^{k} E\left\{\sum_{j=1}^{n_i} (X_{ij} - X_{i\bullet})^2\right\} = \sum_{i=1}^{k} (n_i - 1)\sigma^2 = (n-k)\sigma^2,$$

want $\sum_{j=1}^{n_i} (X_{ij} - X_{i\bullet})^2$ is gelijk aan $(n_i - 1)$ maal de steekproefvariantie van de i^e-steekproef en de steekproefvariantie heeft verwachting σ^2 (zie Statistiek & kansrekening).

Verder is

$$X_{\bullet\bullet} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}}{n} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mu + \alpha_i + U_{ij})}{n} = \mu + \frac{\sum_{i=1}^{k} n_i \alpha_i}{n} + U_{\bullet\bullet}$$

$$= \mu + U_{\bullet\bullet}$$

en dus

$$X_{ij} - X_{\bullet\bullet} = \mu + \alpha_i + U_{ij} - X_{\bullet\bullet} = \alpha_i + U_{ij} - U_{\bullet\bullet}, i = 1, \ldots, k.$$

Er volgt

$$\begin{aligned}
\text{SS(total)} &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{\alpha_i + (U_{ij} - U_{\bullet\bullet})\}^2 \\
&= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{(\alpha_i)^2 + 2\alpha_i(U_{ij} - U_{\bullet\bullet}) + (U_{ij} - U_{\bullet\bullet})^2\} \\
&= \sum_{i=1}^{k} n_i(\alpha_i)^2 + \sum_{i=1}^{k} 2\alpha_i \sum_{j=1}^{n_i} (U_{ij} - U_{\bullet\bullet}) + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (U_{ij} - U_{\bullet\bullet})^2.
\end{aligned}$$

De laatste term, $\sum_{i=1}^{k} \sum_{j=1}^{n_i} (U_{ij} - U_{\bullet\bullet})^2$, is $(n-1)$ maal de steekproefvariantie van de U_{ij}'s en heeft dus verwachting $(n-1)\sigma^2$. De middelste term heeft verwachting 0, want $E(U_{ij}) = 0$ en dus ook $E(U_{\bullet\bullet}) = 0$, alsmede $E(U_{ij} - U_{\bullet\bullet}) = 0$. Derhalve geldt

$$E\{\text{SS(total)}\} = \sum_{i=1}^{k} n_i(\alpha_i)^2 + (n-1)\sigma^2.$$

M.b.v. (3.2.2) en (3.2.3) volgt dan

$$\begin{aligned}
(3.2.4) \qquad E\{\text{SS(factor)}\} &= E\{\text{SS(total)}\} - E\{\text{SS(error)}\} \\
&= \sum_{i=1}^{k} n_i(\alpha_i)^2 + (n-1)\sigma^2 - (n-k)\sigma^2 \\
&= \sum_{i=1}^{k} n_i(\alpha_i)^2 + (k-1)\sigma^2.
\end{aligned}$$

Dit betekent:

1. als H_0 juist is ($\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$), dan zijn SS(error)$/(n-k)$ en SS(factor)$/(k-1)$ beide zuivere schatters van σ^2 en nemen waarden aan die bij elkaar in de buurt liggen;
2. als H_0 niet juist is, is SS(error)$/(n-k)$ nog steeds een zuivere schatter van σ^2, maar SS(factor)$/(k-1)$ is in dat geval systematisch groter, want dan is $\sum_{i=1}^{k} n_i(\alpha_i)^2 > 0$.

Het getal $k-1$ heet het aantal **vrijheidsgraden** van SS(factor), terwijl $n-k$ het aantal vrijheidsgraden is van SS(error). Na deling van SS door het aantal vrijheidsgraden spreken we van MS, **Mean Square**. Dus

$$\begin{aligned} \text{MS(factor)} &= \text{SS(factor)}/(k-1) \\ \text{MS(error)} &= \text{SS(error)}/(n-k). \end{aligned}$$

Als toetsingsgrootheid hanteren we

$$F = \frac{\text{MS(factor)}}{\text{MS(error)}}\,.$$

De bovenvermelde conclusies luiden dan:

1. als H_0 juist is, neemt F waarden in de buurt van 1 aan;
2. als H_0 niet juist is, is F systematisch groter dan 1.

We verwerpen H_0 dus voor grote waarden van F. De kansverdeling van F onder H_0 blijkt een F-verdeling te zijn met $k-1$ vrijheidsgraden in de teller en $n-k$ vrijheidsgraden in de noemer. (We leiden dit resultaat niet af, maar vergelijk de F-toets bij Statistiek & kansrekening, waar de toetsingsgrootheid ook bestaat uit het quotiënt van twee schatters voor σ^2 onder H_0.) Een s.v. met een dergelijke verdeling duiden we aan met F^{k-1}_{n-k}.

We kunnen het k-steekproevenprobleem dus als volgt weergeven.

Kansmodel
$X_{ij} = \mu + \alpha_i + U_{ij} \quad i = 1, \ldots, k;\ j = 1, \ldots, n_i;$
$U_{11}, \ldots, U_{1n_1}, \ldots, U_{k1}, \ldots, U_{kn_k}$ o.o. $N(0, \sigma^2)$-verdeeld;
$\mu, \alpha_1, \ldots, \alpha_k, \sigma^2$ onbekende parameters,
$\frac{n_1}{n}\alpha_1 + \ldots + \frac{n_k}{n}\alpha_k = 0.$

Toetsingsprobleem
$H_0 : \alpha_1 = \ldots = \alpha_k = 0$
$H_1 : \alpha_i \neq 0$ voor zekere i

Toets
Verwerp H_0 als $F = \dfrac{\text{MS(factor)}}{\text{MS(error)}} \geq c$
met $P(F^{k-1}_{n-k} > c) = \alpha$ (c opzoeken in F-tabel)

De onbetrouwbaarheid van de toets is α.

3.2.2 ANOVA-tabel

In computerpakketten worden de benodigde grootheden om de toets in het k-steekproevenprobleem uit te voeren weergegeven in de variantie-analyse tabel (ANOVA table; ANOVA = ANalysis Of VAriance). Deze heeft de volgende vorm

Source	df	SS	MS	F
Factor	$k-1$	SS(factor)	MS(factor)	$\frac{\text{MS(factor)}}{\text{MS(error)}}$
Error	$n-k$	SS(error)	MS(error)	
Total	$n-1$	SS(total)		

df staat voor degrees of freedom. Soms worden andere namen gebruikt (in SPSS bijv. Between groups i.p.v. Factor, en Within groups i.p.v. Error), soms wordt ook de overschrijdingskans toegevoegd.

Voorbeeld 3.2.3 (vervolg van voorbeeld 3.2.1 en 3.2.2) De ANOVA-tabel voor de gegevens van voorbeeld 3.2.1 ziet er zo uit.

Source	df	SS	MS	F
Factor	2	10.8	5.4	0.11
Error	12	573.2	47.8	
Total	14	584.0		

Uit de tabel van de F-verdeling lezen we af bij $\alpha = 0.05$ dat de kritieke waarde c gelijk is aan 3.89. De nulhypothese wordt niet verworpen. Statistisch gezien is er dus niet voldoende bewijs om te beweren dat er een verschil is tussen de drie groepen. □

Als de F-test uitwijst dat de nulhypothese van gelijke verwachtingen voor de k-steekproeven niet verworpen wordt, dan zijn we in zekere zin klaar. Weliswaar hoeven de verwachtingen niet gelijk te zijn, maar de verschillen zijn dermate gering dat ze gemaskeerd worden door de aanwezige variabiliteit, misschien ook vanwege de kleine steekproefomvangen.

Wordt de nulhypothese echter wel verworpen, dan zijn er blijkbaar statistisch aantoonbare verschillen, en we zouden graag willen weten welke steekproeven dan wel van elkaar verschillen in verwachting.

Voorbeeld 3.2.4 Men onderzoekt de loodvergiftiging in 5 grote meren. Uit ieder meer worden 10 forellen gevangen en de loodconcentraties gemeten (in ppm = parts per million). Een samenvatting van de resultaten staat hieronder

	meer 1	meer 2	meer 3	meer 4	meer 5
steekproef-gemiddelde	4.1	3.7	2.4	4.6	3.4
steekproef-standaard-afwijking	0.82	1.06	0.68	1.44	0.93

De ANOVA-tabel ziet er als volgt uit

Source	df	SS	MS	F
Factor	4	27.32	6.83	6.6
Error	45	46.77	1.04	
Total	49	74.09		

Om te onderzoeken of er statistisch aantoonbare verschillen zijn tussen de loodconcentraties in de meren formuleren we een geschikt toetsingsprobleem en lossen dit op volgens de 8 stappen uit Statistiek & kansrekening (zie ook hoofdstuk 2). We noteren X_{ij} voor de loodconcentratie in het i^e meer van de j^e forel ($i = 1, \ldots, 5, j = 1, \ldots, 10$).

1. kansmodel: $X_{ij} = \mu + \alpha_i + U_{ij} \quad i = 1, \ldots, 5, j = 1, \ldots, 10$
   $U_{11}, \ldots, U_{1\,10}, \ldots, U_{51}, \ldots, U_{5\,10}$ o.o. $N(0, \sigma^2)$-verdeeld
   $\frac{10}{50}\alpha_1 + \frac{10}{50}\alpha_2 + \frac{10}{50}\alpha_3 + \frac{10}{50}\alpha_4 + \frac{10}{50}\alpha_5 = 0$
   $\mu, \alpha_1, \ldots, \alpha_5, \sigma^2$ onbekende parameters
2. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$
   $H_1 : \alpha_i \neq 0$ voor zekere i
3. $F = \dfrac{\text{MS(factor)}}{\text{MS(error)}}$ met $\text{MS(factor)} = \sum_{i=1}^{5} 10(X_{i\bullet} - X_{\bullet\bullet})^2/4$
   $$\text{MS(error)} = \sum_{i=1}^{5}\sum_{j=1}^{10}(X_{ij} - X_{i\bullet})^2/45$$
4. F-verdeling met 4 vrijheidsgraden in de teller en 45 vrijheidsgraden in de noemer
5. waarde van $F : 6.6$
6*. overschrijdingskans < 0.01
7. verwerp H_0, want de overschrijdingskans is (veel) kleiner dan $\alpha = 0.05$.
8. statistisch is aangetoond dat de verwachte loodconcentraties in de meren niet alle gelijk zijn. □

3.2.3 Multiple Comparison

Om te weten te komen welke verwachtingen van elkaar verschillen maken we gebruik van Tukey-Kramer's multiple comparison procedure. Deze procedure werkt als volgt. (We geven geen afleiding.)

1. Lees uit de tabel van de Studentized Range de kritieke waarde q af, die afhangt van k, df(error) en de onbetrouwbaarheidsdrempel α.
2. Bereken voor iedere i en j uit $\{1, \ldots, k\}$ met $i \neq j$
   $$d_{ij} = q\sqrt{\text{MS(error)} \times \tfrac{1}{2}\left(\tfrac{1}{n_i} + \tfrac{1}{n_j}\right)}$$
3. Als $|X_{i\bullet} - X_{j\bullet}| \geq d_{ij}$, dan concluderen we dat de i^e en j^e verwachting verschillend zijn, anders niet.

Opmerking Als de steekproefomvangen voor alle k-steekproeven gelijk zijn, hangt d_{ij} niet van i en j af en hoeven we deze slechts één keer uit te rekenen. □

Voorbeeld 3.2.5 (vervolg van voorbeeld 3.2.4) Voor de gegevens van voorbeeld 3.2.4 vinden we in de tabel bij onbetrouwbaarheidsdrempel $\alpha = 0.05$ de waarde $q = 4.03$. Omdat de steekproefomvang voor alle 5 steekproeven dezelfde is, nl. 10, hangt d niet van i en j af. We krijgen

$$d = 4.03\sqrt{1.04 \times \frac{1}{2}\left(\frac{1}{10} + \frac{1}{10}\right)} = 1.30$$

Om te kijken voor welke combinaties het verschil in steekproefgemiddelde tenminste 1.30 is, ordenen we de steekproefgemiddelden van hoog naar laag.

$$\begin{array}{lllll} x_{4\bullet} & x_{1\bullet} & x_{2\bullet} & x_{5\bullet} & x_{3\bullet} \\ 4.6 & 4.1 & 3.7 & 3.4 & 2.4. \end{array}$$

We beginnen bij de grootste, vergelijken deze met de kleinste, één na kleinste, enz. tot we onder d komen. Hier is dat al snel gebeurd, want $x_{4\bullet} - x_{3\bullet} = 2.2 \geq 1.30$, maar $x_{4\bullet} - x_{5\bullet} = 1.2 < 1.30$. We verklaren μ_4 significant verschillend van μ_3, maar houden μ_4, μ_1, μ_2 en μ_5 voor gelijk.

We kijken naar de één na grootste en passen hetzelfde procedé toe. We vinden μ_1 significant verschillend van μ_3, maar niet van μ_2 en μ_5. Ook μ_2 is significant verschillend van μ_3, maar niet van μ_5. Tenslotte zijn μ_5 en μ_3 niet significant verschillend.

Om deze bevindingen overzichtelijk op een rij te zetten, doen we het volgende. We geven de resultaten weer door de verwachtingen te ordenen overeenkomstig de bijbehorende steekproefgemiddelden en een streep te plaatsen onder die verwachtingen, die niet significant van elkaar verschillen.

$$\begin{array}{llllll} \mu_4 & \mu_1 & \mu_2 & \mu_5 & \mu_3 & \\ & & & & & 1^e \text{ stap} \\ & & & & & 2^e \text{ stap} \\ & & & & & 3^e \text{ stap} \\ & & & & & 4^e \text{ stap} \end{array}$$

De 2^e en 3^e stap zijn al weergegeven in de 1^e stap en kunnen dus weggelaten worden. We houden over

$$\begin{array}{lllll} \mu_4 & \mu_1 & \mu_2 & \mu_5 & \mu_3 \end{array}$$

Dit zegt dat μ_4, μ_1 en μ_2 significant groter zijn dan μ_3. □

3.2.4 Schatten, (simultane) betrouwbaarheidsintervallen

Over het schatten van de parameters μ, $\alpha_1, \ldots, \alpha_k$ hebben we nog niet veel gezegd. Uiteraard schatten we de verwachting $\mu + \alpha_i$ van de i^e-steekproef met $X_{i\bullet}$, en bijv. $\alpha_3 - \alpha_2$ met $X_{3\bullet} - X_{2\bullet}$. Constructie van betrouwbaarheidsintervallen voor bijv. $\alpha_3 - \alpha_2$ of $\mu + \alpha_i$ is ook mogelijk. Zo wordt het stochastisch betrouwbaarheidsinterval voor $\mu + \alpha_i$ met betrouwbaarheid $1 - \alpha$ gegeven door

$$\left(X_{i\bullet} - c\sqrt{\text{MS (error)}/n_i},\ X_{i\bullet} + c\sqrt{\text{MS (error)}/n_i}\right) \tag{3.2.5}$$

met $P(T_{n-k} \le c) = 1 - \frac{1}{2}\alpha$, waarbij T_{n-k} een Student-verdeling heeft met $n - k$ vrijheidsgraden. (De waarde c vinden we in de tabel van de Student-verdeling. We geven geen afleiding van dit resultaat.)

Voor de gegevens van voorbeeld 3.2.1 vinden we als 95%-betrouwbaarheidsinterval voor het verwachte resultaat bij opleidingsniveau B

$$\left(84.4 - 2.18\sqrt{47.8/5},\ 84.4 + 2.18\sqrt{47.8/5}\right) = (77.7,\ 91.1).$$

De structuur van dit soort betrouwbaarheidsintervallen is als volgt. Laat de variantie van de gebruikte schatter gelijk zijn aan $a\sigma^2$, dan wordt het betrouwbaarheidsinterval met betrouwbaarheid $1 - \alpha$ gegeven door

$$\left(\text{schatter } - c\sqrt{a\text{MS(error)}},\ \text{schatter } + c\sqrt{a\text{MS(error)}}\right) \tag{3.2.6}$$

met $P(T_{n-k} \le c) = 1 - \frac{1}{2}\alpha$.

De variantie van $X_{i\bullet}$ is σ^2/n_i en dus krijgen we uit (3.2.6) met $a = 1/n_i$ formule (3.2.5).

Bij het betrouwbaarheidsinterval voor μ nemen we als schatter $X_{\bullet\bullet}$. Deze schatter heeft als variantie σ^2/n met $n = \sum_{i=1}^k n_i$, de totale steekproefomvang. Het betrouwbaarheidsinterval voor μ met betrouwbaarheid $1 - \alpha$ wordt dus

$$\left(X_{\bullet\bullet} - c\sqrt{\text{MS(error)}/n},\ X_{\bullet\bullet} + c\sqrt{\text{MS(error)}/n}\right).$$

Voorbeeld 3.2.6 (vervolg van voorbeeld 3.2.4 en 3.2.5) We willen het 95%-betrouwbaarheidsinterval berekenen voor $\mu_2 - \mu_3$ ($=\alpha_2 - \alpha_3$) op grond van de gegevens van voorbeeld 3.2.4. Als schatter voor $\mu_2 - \mu_3$ hanteren we $X_{2\bullet} - X_{3\bullet}$. De variantie van $X_{2\bullet}$ is $\sigma^2/10$ en van $X_{3\bullet}$ ook $\sigma^2/10$. Vanwege de onafhankelijkheid is de variantie van $X_{2\bullet} - X_{3\bullet}$ gelijk aan $\sigma^2/10 + \sigma^2/10$. Het betrouwbaarheidsinterval wordt dus

$$\left(X_{2\bullet} - X_{3\bullet} - c\sqrt{\text{MS(error)}\{\tfrac{1}{10} + \tfrac{1}{10}\}},\ X_{2\bullet} - X_{3\bullet} + c\sqrt{\text{MS(error)}\{\tfrac{1}{10} + \tfrac{1}{10}\}}\right)$$

met c gegeven door $P(T_{45} \le c) = 0.975$. Uit de tabel van de Student-verdeling vinden we $c = 2.02$. Bij voorbeeld 3.2.4 lezen we af $x_{2\bullet} = 3.7$, $x_{3\bullet} = 2.4$ en de waarde van MS(error): 1.04. Invullen geeft als interval: $0.38 < \mu_2 - \mu_3 < 2.22$.□

Het is zelfs mogelijk **simultane** betrouwbaarheidsintervallen te maken. We spreken dan van multiple comparison (zie 3.2.3). De eerder vermelde Tukey-Kramer methode heeft een variant op het gebied van de betrouwbaarheidsintervallen. Dit gaat als volgt.

Schrijf

$$O_{ij} = X_{i\bullet} - X_{j\bullet} - q\sqrt{\text{MS(error)} \times \tfrac{1}{2}\left(\tfrac{1}{n_i} + \tfrac{1}{n_j}\right)}$$

$$B_{ij} = X_{i\bullet} - X_{j\bullet} + q\sqrt{\text{MS(error)} \times \tfrac{1}{2}\left(\tfrac{1}{n_i} + \tfrac{1}{n_j}\right)}$$

met q uit de tabel van de Studentized Range corresponderend met k, df(error) $=$ $n-$k en $\alpha(= 1-$betrouwbaarheid). Dan geldt

$$P(O_{ij} < \mu_i - \mu_j < B_{ij} \text{ voor alle } 1 \leq i, j \leq k \textbf{ tegelijk}) \geq 1 - \alpha.$$

De intervallen (O_{ij}, B_{ij}) zijn dus **simultane stochastische betrouwbaarheidsintervallen** voor $\mu_i - \mu_j$. (Daarop is ook de eerder beschreven toewijzing van significante verschillen gebaseerd.)

Voorbeeld 3.2.7 (vervolg van voorbeeld 3.2.4, 3.2.5 en 3.2.6) Voor de gegevens van voorbeeld 3.2.4. krijgen we (zie ook voorbeeld 3.2.5)
$o_{12} = 4.1 - 3.7 - 1.30 = -0.9,\ b_{12} = 4.1 - 3.7 + 1.30 = 1.7,$
$o_{13} = 4.1 - 2.4 - 1.3 = 0.4,$ enz. De **gehele** uitspraak

$$\begin{array}{lll} -0.9 < \mu_1 - \mu_2 < 1.7, & 0.4 < \mu_1 - \mu_3 < 3.0, & -1.8 < \mu_1 - \mu_4 < 0.8, \\ -0.6 < \mu_1 - \mu_5 < 2.0, & 0.0 < \mu_2 - \mu_3 < 2.6, & -2.2 < \mu_2 - \mu_4 < 0.4, \\ -1.0 < \mu_2 - \mu_5 < 1.6, & -3.5 < \mu_3 - \mu_4 < -0.9, & -2.3 < \mu_3 - \mu_5 < 0.3, \\ -0.1 < \mu_4 - \mu_5 < 2.5 & & \end{array}$$

heeft een betrouwbaarheid van tenminste 0.95. Omdat 0 niet in de intervallen betreffende $\mu_1 - \mu_3,\ \mu_2 - \mu_3,\ \mu_3 - \mu_4$ ligt, zijn $\mu_1,\ \mu_2$ en μ_4 significant verschillend van (groter dan) μ_3. Dit is een herhaling van het in voorbeeld 3.2.5 verkregen resultaat. In feite is bovengenoemde uitspraak een uitgebreidere versie van het in voorbeeld 3.2.5 verkregen resultaat. □

We zien dat het interval voor $\mu_2 - \mu_3$ in voorbeeld 3.2.6 **smaller** is dan het gedeelte betreffende $\mu_2 - \mu_3$ in voorbeeld 3.2.7. De verklaring hiervoor is eenvoudig. De foutenmarge van 0.05 kan in voorbeeld 3.2.6 volledig aan het interval voor $\mu_2 - \mu_3$ besteed worden. In voorbeeld 3.2.7 wordt deze foutenmarge uitgesmeerd over alle intervallen. Het interval betreffende $\mu_2 - \mu_3$ heeft bij de simultane betrouwbaarheidsintervallen van voorbeeld 3.2.7 daardoor dus een kleinere foutenmarge. Hiermee correspondeert een breder interval, want daarmee worden minder vaak uitspraken gedaan.

Het is ook mogelijk simultane betrouwbaarheidsintervallen op te stellen voor $\sum_{i=1}^{k} c_i\alpha_i$ met $\sum_{i=1}^{k} c_i = 0$. Voor meer details zie Miller (1986).

3.3 Tweevoudige tabellen

We bekijken tweevoudige tabellen, waarin twee factoren onafhankelijk van elkaar variëren en voor iedere combinatie van de **niveaus** van de twee factoren één waarneming beschikbaar is. Dit is het eenvoudigste geval van **two-way analysis of variance**.

Voorbeeld 3.3.1 In de volgende tabel staan de aantallen uren zonneschijn in 5 meetstations in de maand juli van 1976 − 1983.

jaar	De Kooy	Eelde	De Bilt	Vlissingen	Beek
1976	282.9	269.3	251.8	272.5	238.6
1977	191.2	189.8	183.8	192.0	160.5
1978	193.0	175.0	167.1	155.1	157.2
1979	169.4	134.7	144.4	181.9	117.2
1980	165.8	113.7	131.0	160.7	120.1
1981	210.0	160.7	167.8	179.0	137.8
1982	254.3	251.6	231.3	196.9	203.9
1983	250.4	251.2	252.5	215.3	263.4

De twee factoren hier zijn “jaar” met 8 niveaus en “meetstation” met 5 niveaus. De getallen binnen in de tabel zijn waargenomen waarden. □

We streven er nu naar structuur te ontdekken in de brij van getallen die zo'n tabel op het eerste gezicht is. Bij “structuur” denken we hier in de eerste plaats aan een **additief** model

$$X_{ij} = \mu + \alpha_i + \beta_j + U_{ij} \quad i = 1, \ldots, I, j = 1, \ldots, J. \tag{3.3.1}$$

Hierin is X_{ij} de waarneming bij rij i en kolom j (we zeggen wel X_{ij} is de waarneming in **cel** (i, j)). De parameter μ is een overall-waarde voor de hele tabel, niet afhankelijk van i en j, α_i is de parameter voor de bijdrage van niveau i van de rij-factor, β_j is de parameter voor de bijdrage van niveau j van de kolomfactor en U_{ij} tenslotte is de random fluctuatie t.o.v. het additieve model. We veronderstellen dat

$$U_{11}, \ldots, U_{1J}, U_{21}, \ldots, U_{2J}, \ldots, U_{I1}, \ldots, U_{IJ}$$

o.o. s.v.-en zijn, ieder met een $N(0, \sigma^2)$-verdeling. Hierbij is σ^2 een onbekende parameter, die niet van i en j afhangt.

De formulering van het model in (3.3.1) is echter nog niet voldoende gespecificeerd om μ, α_i en β_j vast te kunnen leggen. Immers, als we μ vervangen door $\mu + 1$ en α_i door $\alpha_i - 1$ blijft $\mu + \alpha_i$ hetzelfde. We moeten dus nog extra eisen aan de parameters opleggen om **identificatie** van de parameters te realiseren. We eisen, analoog aan het k-steekproeven model,

$$\sum_{i=1}^{I} \alpha_i = 0, \ \sum_{j=1}^{J} \beta_j = 0.$$

We noemen α_i het i^e **rij-effect** en β_j het j^e **kolom-effect**.

Voorbeeld 3.3.2 (vervolg van voorbeeld 3.3.1) Bij een beschrijving van de data van voorbeeld 3.3.1 m.b.v. het additieve model is de parameter μ de fictieve gemeenschappelijke waarde voor het aantal uren zonneschijn door alle jaren heen over alle 5 stations; α_i geeft het effect van het betreffende jaar over de 5 stations, en β_j beschrijft het effect van het meetstation over de jaren. Een "zonnig" jaar zal dus resulteren in een positieve α_i, een somber jaar in een negatieve α_i. Evenzo zal een meetstation met gewoonlijk meer zon dan de andere resulteren in een positieve β_j voor het betreffende station. □

3.3.1 Toetsen

We kunnen ons afvragen bij de gegevens van voorbeeld 3.3.1 of er verschil is tussen de jaren. In het kansmodel betekent het dat we willen toetsen

$$H_0 : \alpha_1 = \alpha_2 = \ldots \alpha_8 = 0$$

tegen het alternatief

$$H_1 : \alpha_i \neq 0 \text{ voor zekere } i.$$

We gaan op dezelfde manier te werk als in het k-steekproevenprobleem, d.w.z. dat we de variatie veroorzaakt door de factor "jaar" vergelijken met de niet aan de beide factoren toe te schrijven "rest"-variatie. Noteer

$$X_{i\bullet} = \frac{1}{J} \sum_{j=1}^{J} X_{ij}, \ X_{\bullet j} = \frac{1}{I} \sum_{i=1}^{I} X_{ij}, \text{ en } X_{\bullet\bullet} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}$$

voor het steekproefgemiddelde van respectievelijk de i^e rij, de j^e de kolom en alle waarnemingen. We splitsen het verschil tussen X_{ij} en $X_{\bullet\bullet}$ in drie stukken

$$X_{ij} - X_{\bullet\bullet} = (X_{i\bullet} - X_{\bullet\bullet}) + (X_{\bullet j} - X_{\bullet\bullet}) + (X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet}), \tag{3.3.2}$$

betrekking hebbend op de variatie van de i^e rij t.o.v. het geheel, de j^e kolom t.o.v. het geheel en wat overblijft. We kwadrateren linker- en rechterlid van (3.3.2) en sommeren over i en j. De dubbele producten vallen, net als bij het k-steekproeven probleem, weg. Bij wijze van voorbeeld bekijken we het dubbele product

$$2\sum_i\sum_j(X_{i\bullet}-X_{\bullet\bullet})(X_{ij}-X_{i\bullet}-X_{\bullet j}+X_{\bullet\bullet})=$$

$$2\sum_i(X_{i\bullet}-X_{\bullet\bullet})\sum_j(X_{ij}-X_{i\bullet}-X_{\bullet j}+X_{\bullet\bullet})=0$$

want voor iedere $i\in\{1,\dots,I\}$ geldt

$$\sum_j(X_{ij}-X_{i\bullet}-X_{\bullet j}+X_{\bullet\bullet})=$$

$$\sum_j X_{ij}-JX_{i\bullet}-\sum_j X_{\bullet j}+JX_{\bullet\bullet}=$$

$$JX_{i\bullet}-JX_{i\bullet}-JX_{\bullet\bullet}+JX_{\bullet\bullet}=0.$$

Bij het kwadrateren en sommeren in (3.3.2) krijgen we dus alleen de kwadraattermen en geen dubbele producten. Schrijf

$$\begin{aligned}\text{SS(total)}&=\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij}-X_{\bullet\bullet})^2\\ \text{SSA}&=J\sum_{i=1}^{I}(X_{i\bullet}-X_{\bullet\bullet})^2\\ \text{SSB}&=I\sum_{j=1}^{J}(X_{\bullet j}-X_{\bullet\bullet})^2\\ \text{SS(error)}&=\sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij}-X_{i\bullet}-X_{\bullet j}+X_{\bullet\bullet})^2,\end{aligned}\tag{3.3.3}$$

waarbij SSA betrekking heeft op de rijen en SSB op de kolommen. We hebben dus

$$\boxed{\text{SS(total)}=\text{SSA}+\text{SSB}+\text{SS(error)}}\tag{3.3.4}$$

waarmee de totale variatie in drie stukken gesplitst is, één betreffende factor A, één voor factor B en één voor het niet-verklaarde deel.

Voorbeeld 3.3.3 (vervolg van voorbeeld 3.3.1 en 3.3.2) Voor de gegevens van voorbeeld 3.3.1 krijgen we na enig rekenwerk

$$\begin{aligned}\text{SS(total)} &= 89708.9\\ \text{SSA} &= 74890.5\\ \text{SSB} &= 6401.9\\ \text{SS(error)} &= 8416.5.\end{aligned}$$ □

De toetsingsgrootheid voor het toetsen van $H_0 : \alpha_1 = \ldots = \alpha_I = 0$ is gebaseerd op SSA en SS(error). Voor het toetsen betreffende het kolomeffect zijn we aangewezen op SSB en SS(error). Dit wordt duidelijk als we de verwachtingen van SSA, SSB en SS(error) bepalen. Met behulp van $\sum_i \alpha_i = \sum_j \beta_j = 0$ volgt

$$\begin{aligned}E(\text{SSA}) &= E\left\{J\sum_i (X_{i\bullet} - X_{\bullet\bullet})^2\right\} = JE\left\{\sum_i \{\mu + \alpha_i + U_{i\bullet} - (\mu + U_{\bullet\bullet})\}^2\right\}\\ &= JE\left\{\sum_i \left\{\alpha_i^2 + 2\alpha_i(U_{i\bullet} - U_{\bullet\bullet}) + (U_{i\bullet} - U_{\bullet\bullet})^2\right\}\right\}\\ &= J\sum_i \alpha_i^2 + 0 + JE\left\{\sum_i (U_{i\bullet} - U_{\bullet\bullet})^2\right\} = J\sum_i \alpha_i^2 + J(I-1)\frac{\sigma^2}{J}\\ &= J\sum_i \alpha_i^2 + (I-1)\sigma^2,\end{aligned}$$

want $U_{1\bullet}, \ldots, U_{I\bullet}$ zijn o.o. s.v.-en met een $N(0, \frac{\sigma^2}{J})$-verdeling en $U_{\bullet\bullet}$ is het gemiddelde van deze s.v.-en, zodat

$$E(U_{i\bullet} - U_{\bullet\bullet}) = 0 \text{ en } E\left\{\sum_{i=1}^{I} (U_{i\bullet} - U_{\bullet\bullet})^2\right\} = (I-1)\text{var}(U_{i\bullet}) = (I-1)\frac{\sigma^2}{J}.$$

Evenzo geldt

$$\begin{aligned}E(\text{SSB}) &= I\sum_j \beta_j^2 + (J-1)\sigma^2\\ E\{\text{SS(total)}\} &= J\sum_i \alpha_i^2 + I\sum_j \beta_j^2 + (IJ-1)\sigma^2\end{aligned}$$

en dus is, zie (3.3.4),

$$\begin{aligned}&E\{\text{SS(error)}\} = E\{\text{SS(total)}\} - E(\text{SSA}) - E(\text{SSB}) =\\ &(IJ-1)\sigma^2 - (I-1)\sigma^2 - (J-1)\sigma^2 = (I-1)(J-1)\sigma^2.\end{aligned} \tag{3.3.5}$$

De mean squares worden nu gedefinieerd door

$$
\text{(3.3.6)} \qquad \begin{aligned}
\text{MSA} &= \text{SSA}/(I-1) \\
\text{MSB} &= \text{SSB}/(J-1) \\
\text{MS(error)} &= \frac{\text{SS (error)}}{(I-1)(J-1)}
\end{aligned}
$$

en

$I-1$ is het aantal vrijheidsgraden van SSA,

$J-1$ dat van SSB en $(I-1)(J-1)$ dat van SS(error).

We concluderen uit de berekeningen van de verwachtingen:

1. als H_0 juist is ($\alpha_1 = \ldots = \alpha_I = 0$) dan zijn MSA en MS(error) beide zuivere schatters van σ^2 en nemen waarden aan die bij elkaar in de buurt liggen;
2. als H_0 niet juist is, is MS(error) nog steeds een zuivere schatter van σ^2, maar MSA is in dat geval systematisch groter.

De toetsingsgrootheid wordt

$$F = \frac{\text{MSA}}{\text{MS(error)}},$$

en we concluderen

1. als H_0 juist is, neemt F waarden aan in de buurt van 1;
2. als H_0 niet juist is, is F systematisch groter dan 1.

We verwerpen H_0 voor grote waarden van F. De kansverdeling van F onder H_0 is een F-verdeling met $I-1$ vrijheidsgraden in de teller en $(I-1)(J-1)$ vrijheidsgraden in de noemer. (We bewijzen dit niet, maar vergelijk de F-toets bij Statistiek & kansrekening.)

Voorbeeld 3.3.4 (vervolg van voorbeeld 3.3.1, 3.3.2 en 3.3.3) We willen onderzoeken of er sprake is van een “plaats-effect” bij de gegevens over de aantallen uren zonneschijn, d.w.z. of er op de ene plaats significant meer of minder zon geweest is dan op een andere plaats. Vertaald in termen van het toetsingsprobleem betekent dit het toetsen van $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ tegen het alternatief $H_1 : \beta_j \neq 0$ voor zekere j. De toetsingsgrootheid is F = MSB/MS(error). Onder

H_0 heeft deze toetsingsgrootheid een F^4_{28}-verdeling. We nemen als onbetrouwbaarheidsdrempel $\alpha = 0.05$. De waarde van deze toetsingsgrootheid bedraagt hier (zie voorbeeld 3.3.3 en formule (3.3.6))

$$\frac{6401.9/4}{8416.5/28} = 5.32.$$

Uit de tabel van de F-verdeling lezen we af dat de kritieke waarde gelijk is aan 2.71. Omdat $5.32 > 2.71$, wordt H_0 verworpen. Statistisch is derhalve aangetoond dat er verschil is in aantallen uren zonneschijn tussen de genoemde plaatsen. □

We kunnen de besproken toetsen bij tweevoudige tabellen aldus weergeven.

Kansmodel

$X_{ij} = \mu + \alpha_i + \beta_j + U_{ij} \quad i = 1, \ldots, I, j = 1, \ldots, J$

$U_{11}, \ldots, U_{1J}, \ldots, U_{I1}, \ldots, U_{IJ}$ o.o. $N(0, \sigma^2)$-verdeeld

$\sum_{i=1}^{I} \alpha_i = 0, \ \sum_{j=1}^{J} \beta_j = 0,$

$\mu, \alpha_1, \ldots, \alpha_I, \beta_1, \ldots, \beta_J, \sigma^2$ onbekende parameters

Toetsingsprobleem

$H_0 : \alpha_1 = \ldots = \alpha_I = 0$

$H_1 : \alpha_i \neq 0$ voor zekere i

Toets

Verwerp H_0 als $F = \dfrac{\text{MSA}}{\text{MS(error)}} \geq c$

met $P\left(F^{I-1}_{(I-1)(J-1)} > c\right) = \alpha$ (c opzoeken in F-tabel)

Toetsingsprobleem

$H_0 : \beta_1 = \ldots = \beta_J = 0$

$H_1 : \beta_j \neq 0$ voor zekere j

Toets

Verwerp H_0 als $F = \dfrac{\text{MSB}}{\text{MS(error)}} \geq c$

met $P\left(F^{J-1}_{(I-1)(J-1)} > c\right) = \alpha$ (c opzoeken in F-tabel)

De onbetrouwbaarheid van de toetsen is α.

De ANOVA-tabel voor two-way analysis of variance ziet er zo uit.

Source	df	SS	MS	F
Factor A	$I-1$	SSA	MSA	$\dfrac{\text{MSA}}{\text{MS(error)}}$
Factor B	$J-1$	SSB	MSB	$\dfrac{\text{MSB}}{\text{MS(error)}}$
Error	$(I-1)(J-1)$	SS(error)	MS(error)	
Total	$IJ-1$	SS(total)		

Voorbeeld 3.3.5 (vervolg van voorbeeld 3.3.1, 3.3.2, 3.3.3 en 3.3.4) De ANOVA-tabel voor de gegevens van voorbeeld 3.3.1 ziet er zo uit

Source	df	SS	MS	F
jaar	7	74890.5	10698.6	35.59
meetstation	4	6401.9	1600.5	5.32
error	28	8416.5	300.6	
totaal	39	89708.9		

We beschrijven het toetsingsresultaat van voorbeeld 3.3.4 nog eens in termen van de 8 stappen uit Statistiek & kansrekening (zie ook hoofdstuk 2).

1. kansmodel: $X_{ij} = \mu + \alpha_i + \beta_j + U_{ij},\ i = 1, \ldots, 8, j = 1, \ldots, 5$ met $U_{11}, \ldots, U_{85}$ o.o. s.v.-en ieder met een $N(0, \sigma^2)$-verdeling; $\mu, \alpha_1, \ldots, \alpha_8, \beta_1, \ldots, \beta_5$ en σ^2 zijn onbekende parameters met $\alpha_1 + \ldots + \alpha_8 = 0$ en $\beta_1 + \ldots + \beta_5 = 0$;
2. $H_0 : \beta_1 = \ldots = \beta_5 = 0$
   $H_1 : \beta_j \neq 0$ voor zekere j;
3. $F =$ MSB/MS(error) met MSB $= 8 \sum_{j=1}^{5} (X_{\bullet j} - X_{\bullet\bullet})^2/4$ en
   MS(error) $= \sum_{i=1}^{8} \sum_{j=1}^{5} (X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet})^2/28$;
4. F-verdeling met 4 vrijheidsgraden in de teller en 28 vrijheidsgraden in de noemer;
5. waarde van F: 5.32;
6. kritieke waarde: 2.71; kritiek gebied $[2.71, \infty)$;
7. verwerp H_0, want $5.32 > 2.71$;
8. statistisch is aangetoond dat er verschil is in aantallen uren zonneschijn tussen de genoemde plaatsen. □

3.3.2 Schatten

Zoals we gezien hebben in voorbeeld 3.3.4 is statistisch aangetoond dat er verschil is in aantallen uren zonneschijn tussen de verschillende meetstations. We zouden dan graag willen weten hoe die verschillen zijn. Waar is er meer zon, en hoeveel meer? We kunnen dat aanpakken met behulp van de schattingstheorie. We schatten de parameter μ, die de overall-waarde voor de hele tabel representeert, met het totale gemiddelde van alle waarnemingen. De parameter α_i beschrijft de afwijking van niveau i t.o.v. μ. Deze schatten we met het verschil tussen het gemiddelde van de i^e rij en het totale gemiddelde. Op analoge wijze schatten we β_j. In formule:

(3.3.7) $\hat{\mu} = X_{\bullet\bullet}\,, \quad \hat{\alpha}_i = X_{i\bullet} - X_{\bullet\bullet}\,, \quad \hat{\beta}_j = X_{\bullet j} - X_{\bullet\bullet}\,.$

We kunnen nu schrijven (vgl. ook (3.3.2))

(3.3.8) $X_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + R_{ij}$

met

(3.3.9) $R_{ij} = X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet}\,.$

We noemen R_{ij} het **residu** behorend bij X_{ij}. We proberen namelijk X_{ij} te beschrijven met $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$. R_{ij} geeft dan aan wat er nog resteert. Formule (3.3.8) lijkt erg op formule (3.3.1). We zouden formule (3.3.8) kunnen opvatten als een neerslag van formule (3.3.1) in termen van waarneembare s.v.-en: $\hat{\mu}$ als schatter van μ, $\hat{\alpha}_i$ als schatter van α_i, $\hat{\beta}_j$ als schatter van β_j.

Er zijn echter conceptueel grote verschillen. De parameters μ, α_i en β_j zijn (en blijven) onbekende getallen, terwijl $\hat{\mu}$, $\hat{\alpha}_i$ en $\hat{\beta}_j$ s.v.-en zijn, waarvan we met behulp van de data de waardes kunnen berekenen. Ook R_{ij} is (via formule (3.3.9)) te berekenen, terwijl de random fluctuatie U_{ij} niet waarneembaar is (omdat μ, α_i en β_j onbekend zijn).

We kunnen formule (3.3.8) (of de bijbehorende realisaties van de betreffende s.v.-en) ook in tabelvorm presenteren. We noemen $\hat{\alpha}_i$ dan het geschatte rij-effect, $\hat{\beta}_j$ het geschatte kolom-effect, terwijl $\hat{\mu}$ rechts onderin genoteerd wordt. De residuen zijn de getallen binnen in de tabel.

Voorbeeld 3.3.6 (vervolg van voorbeeld 3.3.1, 3.3.2, 3.3.3, 3.3.4 en 3.3.5) Voor de gegevens van voorbeeld 3.3.1 krijgen we na enig rekenwerk

jaar	De Kooy	Eelde	De Bilt	Vlissingen	Beek	geschatte rij-effect
1976	–1.1	6.7	–8.8	8.9	–5.6	69.4
1977	–13.3	6.7	2.7	8.0	–4.2	–10.2
1978	2.5	5.9	0.0	–14.9	6.5	–24.1
1979	–1.1	–14.5	–2.7	31.8	–13.5	–44.1
1980	6.5	–24.2	–4.9	21.9	0.6	–55.4
1981	17.9	-10.0	-0.9	7.4	–14.5	–22.6
1982	5.7	24.4	6.1	–31.3	–4.9	34.0
1983	–17.2	5.0	8.3	–31.8	35.6	52.9
geschatte kolom-effect	21.0	–0.4	–2.4	0.6	–18.8	193.6

Uit de tabel lezen we de schattingen af:

$$\begin{aligned}
(3.3.10)\qquad & \hat{\mu} \;= x_{\bullet\bullet} = 193.6,\\
& \hat{\alpha}_1 = x_{1\bullet} - x_{\bullet\bullet} = 69.4,\ \hat{\alpha}_2 = -10.2,\ \hat{\alpha}_3 = -24.1,\ \hat{\alpha}_4 = -44.1\\
& \hat{\alpha}_5 = -55.4,\ \hat{\alpha}_6 = -22.6,\ \hat{\alpha}_7 = 34.0,\ \hat{\alpha}_8 = 52.9,\\
& \hat{\beta}_1 = x_{\bullet 1} - x_{\bullet\bullet} = 21.0,\ \hat{\beta}_2 = -0.4,\ \hat{\beta}_3 = -2.4,\ \hat{\beta}_4 = 0.6,\ \hat{\beta}_5 = -18.8,
\end{aligned}$$

terwijl bijv. de waarde r_{23} van het residu R_{23} gelijk is aan 2.7. We kunnen met de tabel en formule (3.3.8) de oorspronkelijke waarden x_{ij} weer makkelijk terugvinden. Zo is de waarde 181.9 van Vlissingen in 1979 (zie voorbeeld 3.3.1) af te lezen als $193.6 + (-44.1) + 0.6 + 31.8$.

Enkele voorlopige conclusies hieruit zijn:
- de jaren 1976 ($\hat{\alpha}_1$), 1982 ($\hat{\alpha}_7$) en 1983 ($\hat{\alpha}_8$) waren zonnig in vergelijking met de andere jaren, terwijl 1979 ($\hat{\alpha}_4$) en 1980 ($\hat{\alpha}_5$) weinig zon gaven;
- in De Kooy ($\hat{\beta}_1$) schijnt de zon vaker in vergelijking met de andere meetstations en Beek ($\hat{\beta}_5$) heeft relatief weinig zon;
- er komen nogal wat grote residuen voor (in absolute waarde). □

Opmerking De eerder gepresenteerde grootheden SSA, SSB en SS(error) kunnen ook als volgt weergegeven worden (zie (3.3.3), (3.3.7) en (3.3.9))

$$\begin{aligned}
\text{SSA} \quad &= J\sum_{i=1}^{I}\hat{\alpha}_i^2\\
\text{SSB} \quad &= I\sum_{j=1}^{J}\hat{\beta}_j^2\\
\text{SS(error)} &= \sum_{i=1}^{I}\sum_{j=1}^{J}R_{ij}^2
\end{aligned}$$

SSA kan dus berekend worden door alle gekwadrateerde geschatte rij-effecten bij elkaar op te tellen en het geheel te vermenigvuldigen met het aantal kolommen. Op analoge wijze krijgen we SSB, terwijl SS(error) de som is van alle gekwadrateerde residuen, dus de som van alle gekwadrateerde getallen binnen in de tabel. □

3.3.3 (Simultane) Betrouwbaarheidsintervallen

De aan het eind van voorbeeld 3.3.6 genoemde conclusies hebben een voorlopig karakter, omdat de grootte van de schatter afgezet moet worden tegen de variantie van de betreffende schatter. Zo is de variantie van $\hat{\alpha}_1$ gelijk aan $\sigma^2(I-1)/(IJ)$

(we geven geen afleiding van dit resultaat). We kunnen deze variantie niet berekenen, want σ^2 is onbekend. Uit (3.3.5) en (3.3.6) volgt dat we σ^2 zuiver kunnen schatten met MS(error). De manier om het geheel nu te presenteren is een betrouwbaarheidsinterval voor α_1.

Analoog aan de in Statistiek & kansrekening en de in 3.2.4 ontwikkelde betrouwbaarheidsintervallen (zie ook formule (3.2.6)) krijgen we als stochastisch betrouwbaarheidsinterval voor α_i met betrouwbaarheid $1-\alpha$:

$$\Big((X_{i\bullet} - X_{\bullet\bullet}) - c\sqrt{\text{MS(error)}(I-1)/(IJ)},$$
$$(X_{i\bullet} - X_{\bullet\bullet}) + c\sqrt{\text{MS(error)}(I-1)/(IJ)}\Big)$$

met c de waarde waarvoor $P(T_{(I-1)(J-1)} \leq c) = 1 - \frac{1}{2}\alpha$. Hierbij is $T_{(I-1)(J-1)}$ een s.v. met een Student-verdeling met $(I-1)(J-1)$ vrijheidsgraden. De waarde c vinden we in de tabel van de Student-verdeling.

Door verwisseling van de rol van i en j vinden we op analoge wijze een betrouwbaarheidsinterval voor β_j. Het stochastisch betrouwbaarheidsinterval voor μ met betrouwbaarheid $1-\alpha$ ziet er zo uit

$$\Big(X_{\bullet\bullet} - c\sqrt{\text{MS(error)}/(IJ)},\ X_{\bullet\bullet} + c\sqrt{\text{MS(error)}/(IJ)}\Big)$$

met $P(T_{(I-1)(J-1)} \leq c) = 1 - \frac{1}{2}\alpha$. Ook voor het verschil van twee parameters heeft het betrouwbaarheidsinterval deze structuur. Zo is, met c als hierboven,

$$\Big(X_{\bullet i} - X_{\bullet j} - c\sqrt{2\text{MS(error)}/I},\ X_{\bullet i} - X_{\bullet j} + c\sqrt{2\text{MS(error)}/I}\Big)$$

een stochastisch betrouwbaarheidsinterval voor $\beta_i - \beta_j$ met betrouwbaarheid $1-\alpha$.

Net als in 3.2.4 (zie ook formule (3.2.6)) hangen de factoren $(I-1)/(IJ), 1/(IJ)$, $2/I$ in bovenstaande betrouwbaarheidsintervallen samen met de varianties van $X_{\bullet i} - X_{\bullet\bullet}, X_{\bullet\bullet}, X_{\bullet i} - X_{\bullet j}$. Deze varianties zijn namelijk

$$\begin{aligned}
\text{var}(X_{i\bullet} - X_{\bullet\bullet}) &= \frac{\sigma^2(I-1)}{IJ} \\
\text{var}\, X_{\bullet\bullet} &= \frac{\sigma^2}{IJ} \\
\text{var}(X_{\bullet i} - X_{\bullet j}) &= \frac{2\sigma^2}{I}\,.
\end{aligned}$$

De onbekende parameter σ^2 wordt in alle drie gevallen geschat met MS(error).

Voorbeeld 3.3.7 (vervolg van voorbeeld 3.3.1, 3.3.2, 3.3.3, 3.3.4, 3.3.5 en 3.3.6) We berekenen voor de gegevens van voorbeeld 3.3.1 het numerieke betrouwbaarheidsinterval voor het kolom-effect van Beek, β_5, bij een betrouwbaarheid van 95% ($\alpha = 0.05$). We krijgen (zie (3.3.10) voor $\hat{\beta}_5 = x_{\bullet 5} - x_{\bullet\bullet}$ en voorbeeld 3.3.5 voor de waarde van MS(error))

$$(-18.8 - 2.05\sqrt{300.6 \times 4/40},\ -18.8 + 2.05\sqrt{300.6 \times 4/40}) = (-30.0, -7.6).$$

Met een betrouwbaarheid van 95% geldt dus dat het verwachte aantal uren zonneschijn in Beek tussen de 7.6 en 30 uur achterblijft bij het overall verwachte aantal uren zonneschijn. Daarmee is de voorlopige conclusie dat Beek relatief weinig zon heeft, gekwantificeerd. □

We hebben hier gekeken naar betrouwbaarheidsintervallen voor elke parameter apart of voor het verschil van twee parameters. Het is ook mogelijk **simultane** betrouwbaarheidsuitspraken te doen (**multiple comparison**) voor alle verschillen $\alpha_1 - \alpha_2, \alpha_1 - \alpha_3, \ldots, \alpha_{I-1} - \alpha_I$ tegelijk.

Er zijn verschillende methoden. Wij bespreken de methode van Tukey-Kramer, die vergelijkbaar is met de gelijknamige methode in het k-steekproevenprobleem (zie 3.2.4). Er geldt

$$P(O_{is} < \alpha_i - \alpha_s < B_{is} \text{ voor alle } 1 \le i, s \le I \textbf{ tegelijk}) \ge 1 - \alpha,$$

waarbij

$$O_{is} = X_{i\bullet} - X_{s\bullet} - q\sqrt{\text{MS(error)}/J}$$

$$B_{is} = X_{i\bullet} - X_{s\bullet} + q\sqrt{\text{MS(error)}/J}$$

en q de kritieke waarde uit de Studentized Range tabel is met voor k ingevuld I en die verder afhangt van df(error) $= (I-1)(J-1)$ en $\alpha (= 1-$ betrouwbaarheid).

Door verwisseling van de rol van rijen en kolommen vinden we op analoge wijze simultane betrouwbaarheidsintervallen voor $\beta_j - \beta_t$ voor alle paren j, t.

Voorbeeld 3.3.8 (vervolg van voorbeeld 3.3.1, 3.3.2, 3.3.3, 3.3.4, 3.3.5, 3.3.6 en 3.3.7) We berekenen voor de gegevens van voorbeeld 3.3.1 de simultane betrouwbaarheidsintervallen voor $\beta_j - \beta_t$ voor alle j, t bij een betrouwbaarheid van 95%. We vinden o.a. (voor $x_{\bullet 1} - x_{\bullet 2} = (x_{\bullet 1} - x_{\bullet\bullet}) - (x_{\bullet 2} - x_{\bullet\bullet}) = \hat{\beta}_1 - \hat{\beta}_2$ zie voorbeeld 3.3.6, voor de waarde van MS(error) zie voorbeeld 3.3.5)

$$o_{12} = 21.0 - (-0.4) - 4.12\sqrt{300.6/8} = -3.9,$$

$$b_{12} = 21.0 - (-0.4) + 4.12\sqrt{300.6/8} = 46.7.$$

Uiteindelijk krijgen we dat met betrouwbaarheid 95% **tegelijkertijd** geldt

$$-3.9 < \beta_1 - \beta_2 < 46.7; \quad -1.9 < \beta_1 - \beta_3 < 48.7; \quad -4.9 < \beta_1 - \beta_4 < 45.7;$$

$$14.5 < \beta_1 - \beta_5 < 65.1; \quad -23.3 < \beta_2 - \beta_3 < 27.3; \quad -26.3 < \beta_2 - \beta_4 < 24.3;$$

$$-6.9 < \beta_2 - \beta_5 < 43.7; \quad -28.3 < \beta_3 - \beta_4 < 22.3; \quad -8.9 < \beta_3 - \beta_5 < 41.7;$$

$$-5.9 < \beta_4 - \beta_5 < 44.7.$$

Wat opvalt is de breedte van de intervallen en dat 0 in bijna alle intervallen ligt. Alleen β_1 is duidelijk groter dan β_5. □

3.3.4 Randomized Block Design

Sterk verwant met de besproken eenvoudige vorm van two-way analysis of variance is **randomized block design**. In Statistiek & kansrekening hebben we het twee-steekproevenprobleem gezien en het probleem van **gepaarde waarnemingen**. Het twee-steekproevenprobleem is aan het begin van dit hoofdstuk gegeneraliseerd tot het k-steekproevenprobleem. Het randomized block design is een soortgelijke generalisatie van paren naar **blokken** waarnemingen, waarbij binnen een blok een zekere homogeniteit aanwezig is, net als binnen een paar. In plaats van dat er binnen een blok twee waarnemingen beschikbaar zijn (zoals bij gepaarde waarnemingen), zijn er nu meer waarnemingen beschikbaar.

Voorbeeld 3.3.9 Drie steden worden door 10 beoordelaars bekeken op de leefkwaliteit. Elk van de beoordelaars geeft een score tussen 0 en 100. Dit zijn de resultaten

	beoordelaar									
stad	1	2	3	4	5	6	7	8	9	10
1	68	40	82	56	70	80	47	55	78	53
2	72	43	89	60	75	91	58	68	77	65
3	65	42	84	50	68	86	50	52	75	60

Om te kijken of er sprake is van een verschil tussen de steden, moeten we rekening houden met het beoordelaarseffect. Als we dat niet zouden doen, zou een eventueel aanwezig verschil tussen de steden onopgemerkt kunnen blijven door de variatie tussen de beoordelaars. (Een eerste blik op de cijfers maakt duidelijk dat die variatie inderdaad vrij groot is en het eventuele stadseffect heel wel kan verdoezelen. Nog anders gezegd: we moeten de score 91 van beoordelaar 6 voor stad 2 vergelijken met de scores 80 en 86 voor stad 1 en 3 door dezelfde beoordelaar en niet met bijv. de scores 47, 58 en 50 van beoordelaar 7.) We hebben in dit voorbeeld dus 10 blokken en één factor op 3 niveaus. □

Het woord **randomized** in de naam randomized block design slaat op het feit dat we binnen een blok de verschillende niveaus op een "randomized" manier moeten aanbieden. Zo moeten we in voorbeeld 3.3.9 niet steeds eerst stad 1, dan stad 2 en dan stad 3 laten beoordelen. Er moet steeds geloot worden welke stad het eerste, het tweede en het laatste moet worden beoordeeld. Vandaar het woord randomized.

De analyse van het randomized block design is precies hetzelfde als de eerder besproken two-way analysis of variance. We beschouwen de blokken als de ene factor en de aanwezige factor (steden in voorbeeld 3.3.9) als de andere factor.

Voorbeeld 3.3.10 (vervolg van voorbeeld 3.3.9)
De ANOVA-tabel voor de gegevens van voorbeeld 3.3.9 ziet er zo uit.

Source	df	SS	MS	F
Factor (stad)	2	304.2	152.1	15.63
Blocks (beoordelaars)	9	5705.0	633.9	65.15
Error	18	175.1	9.73	
Totaal	29	6184.3		

In SPSS wordt dit aldus gepresenteerd.

ANALYSIS OF VARIANCE
Scores
by City
Person

Source of variation	Sum of squares	DF	Mean square	F	Signif of F
Main effects	6009.167	11	546.288	56.147	0.000
city	304.200	2	152.100	15.633	0.000
person	5704.967	9	633.885	65.150	0.000
Explained	6009.167	11	546.288	56.147	0.000
Residual	175.133	18	9.730		
Total	6184.300	29	213.252		

30 cases were processed, 0 cases (0.0 pct) were missing.

We willen de nulhypothese toetsen dat er geen verschil is tussen de steden bij onbetrouwbaarheidsdrempel $\alpha = 0.05$. We volgen de 8 stappen uit Statistiek & kansrekening (zie ook hoofdstuk 2).

1. kansmodel: $X_{ij} = \mu + \alpha_i + \beta_j + U_{ij} \quad i = 1,2,3, j = 1,\ldots,10$
   $U_{11},\ldots,U_{1\,10},U_{21},\ldots,U_{3\,10}$ o.o. $N(0,\sigma^2)$-verdeeld
   $\alpha_1 + \alpha_2 + \alpha_3 = 0, \sum_{j=1}^{10} \beta_j = 0,\ \mu,\alpha_1,\alpha_2,\alpha_3,\beta_1,\ldots,\beta_{10},\sigma^2$
   onbekende parameters
2. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$
   $H_1 : \alpha_i \neq 0$ voor zekere i
3. $F = \dfrac{\text{MSA}}{\text{MS(error)}}$ met $\text{MSA} = \dfrac{10\sum_{i=1}^{3}(X_{i\bullet} - X_{\bullet\bullet})^2}{2}$
   $$\text{MS(error)} = \frac{\sum_{i=1}^{3}\sum_{j=1}^{10}(X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet})^2}{18}$$
4. F-verdeling met 2 vrijheidsgraden in de teller en 18 vrijheidsgraden in de noemer
5. waarde van F : 15.63
6. kritieke waarde 3.55; kritieke gebied: $F \geq 3.55$
7. verwerp H_0, want $15.63 > 3.55$
8. statistisch is aangetoond dat er verschil in leefkwaliteit is tussen de steden. □

3.3.5 Interactie

In het tweevoudige variantie-analyse model dat we tot nu toe bekeken hebben, is er van uit gegaan dat beide factoren los van elkaar opereren. Vaak komt het voor dat beide factoren elkaar beïnvloeden. We spreken van **interactie**.

Om aanwezigheid van interactie te kunnen onderzoeken hebben we meer waarnemingen per cel nodig. We bespreken hier de eenvoudige situatie met evenveel waarnemingen per cel.

Voorbeeld 3.3.11 De fabrikant van een nieuw product wil het effect weten op de verkoop van verschillende soorten verpakking en van beschikbaarheid in verschillende winkeltypes. De prijs en hoeveelheid per verpakking is hetzelfde. Er zijn 5 soorten verpakking en 4 winkeltypes. Per combinatie zijn 3 waarnemingen beschikbaar. □

Het kansmodel (met interactie) dat we hanteren ziet er zo uit.

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + U_{ijk},\ i = 1,\ldots,I, j = 1,\ldots,J, k = 1,\ldots,K.$$

In verband met identificatie van de parameters geldt

$$\alpha_1 + \ldots + \alpha_I = 0, \quad \beta_1 + \ldots + \beta_J = 0$$
$$\gamma_{i1} + \ldots + \gamma_{iJ} = 0 \quad \text{voor } i = 1,\ldots,I$$
$$\gamma_{1j} + \ldots + \gamma_{Ij} = 0 \quad \text{voor } j = 1,\ldots,J.$$

Omdat $\gamma_{11}, \ldots, \gamma_{1J}$ relatieve afwijkingen zijn t.o.v. α_1 geldt $\gamma_{11} + \ldots + \gamma_{1J} = 0$, enz. Verder zijn de U_{ijk}'s o.o. s.v.-en, ieder met een $N(0, \sigma^2)$-verdeling. De parameter γ_{ij} heet het **interactie-effect** in cel (i, j). Analoog aan (3.3.4) geldt

$$\boxed{\text{SS(total)} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SS(error)}.}$$

Hierin is SSA de kwadraatsom betrekking hebbend op factor A (rijen), SSB die op factor B (kolommen), terwijl SSAB de kwadraatsom betreffende de interactie is. We geven geen verdere formules voor deze kwadraatsommen. De bijbehorende aantallen vrijheidsgraden zijn

term	SSA	SSB	SSAB	SS(error)
vrijheidsgraden	$I-1$	$J-1$	$(I-1)(J-1)$	$IJ(K-1)$

Dit leidt tot de volgende mean squares

$$\begin{aligned}
\text{MSA} &= \text{SSA}/(I-1), \\
\text{MSB} &= \text{SSB}/(J-1), \\
\text{MSAB} &= \text{SSAB}/\{(I-1)(J-1)\}, \\
\text{MS(error)} &= \text{SS(error)}/\{IJ(K-1)\}.
\end{aligned}$$

Bij het toetsen op interactie, toetsen we

$$H_0 : \gamma_{11} = \ldots = \gamma_{IJ} = 0$$

tegen

$$H_1 : \gamma_{ij} \neq 0 \text{ voor zekere } i, j.$$

Als toetsingsgrootheid hanteren we

$$F = \text{MSAB}/\text{MS(error)}.$$

Als er geen interactie is (alle γ's 0) dan neemt F waarden in de buurt van 1 aan, want dan zijn MSAB en MS(error) beide zuivere schatters van σ^2. Als H_0 niet juist is, is MS(error) nog steeds een zuivere schatter van σ^2, maar MSAB is in dat geval systematisch groter. We verwerpen H_0 dus voor grote waarden van F. De kansverdeling van F onder H_0 is een F-verdeling met $(I-1)(J-1)$ vrijheidsgraden in de teller en $IJ(K-1)$ vrijheidsgraden in de noemer.

Voorbeeld 3.3.12 (vervolg van voorbeeld 3.3.11) De resultaten van het onderzoek uit voorbeeld 3.3.11 zijn als volgt.

winkeltype (A)	type verpakking (B) 1	2	3	4	5
1	5	6	4	9	3
	6	8	3	10	7
	4	7	5	12	4
2	7	5	3	4	12
	8	5	6	4	14
	8	6	4	6	11
3	3	6	8	9	10
	2	6	9	8	7
	4	5	6	7	7
4	8	12	4	10	7
	9	11	7	12	8
	9	10	3	12	6

De ANOVA-tabel luidt:

Source	SS	df	MS	F
A	49.38	3	16.46	10.29
B	93.23	4	23.31	14.57
AB	268.37	12	22.36	13.98
Treatments	410.98	19		
Error	64.00	40	1.60	
Total	474.98	59		

Merk op dat $I = 4$ (aantal rijen = aantal winkeltypes), $J = 5$ (aantal kolommen = aantal types verpakkingen) en $K = 3$ (aantal waarnemingen per cel). De regel "Treatments" in de variantie analyse tabel is de sommatie van de regels A, B en AB, voor zover van toepassing. Deze regel wordt soms wel weggelaten.

We willen toetsen of er sprake is van interactie en voeren daartoe de 8 stappen uit. Als onbetrouwbaarheidsdrempel nemen we $\alpha = 0.05$.

1. kansmodel: $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + U_{ijk},\ i = 1, \ldots, 4, j = 1, \ldots, 5,$ $k = 1, 2, 3$ met $\alpha_1 + \ldots + \alpha_4 = 0,\ \beta_1 + \ldots + \beta_5 = 0,\ \gamma_{i1} + \ldots + \gamma_{i5} = 0$ voor $i = 1, \ldots, 4, \gamma_{1j} + \ldots + \gamma_{4j} = 0$ voor $j = 1, \ldots, 5$; $U_{111}, \ldots, U_{453}$ o.o. s.v.-en ieder met een $N(0, \sigma^2)$-verdeling; $\mu, \alpha_1, \ldots, \alpha_4, \beta_1, \ldots, \beta_5, \gamma_{11}, \ldots, \gamma_{45}$, σ^2 onbekende parameters.
2. $H_0 : \gamma_{11} = \ldots = \gamma_{45} = 0$
   $H_1 : \gamma_{ij} \neq 0$ voor zekere i, j;
3. $F = \text{MSAB}/\text{MS(error)}$;
4. F-verdeling met 12 vrijheidsgraden in de teller en 40 vrijheidsgraden in de noemer;
5. waarde van F : 13.98;

6. kritieke waarde: 2.00; kritieke gebied: $F \geq 2.00$;
7. verwerp H_0, want $13.98 > 2.00$;
8. statistisch is aangetoond dat er interactie is tussen het winkeltype en het type verpakking.

We kunnen ook toetsen of het winkeltype er toe doet. We voeren de 8 stappen uit en nemen als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

1. kansmodel: zie boven
2. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$
   $H_1 : \alpha_i \neq 0$ voor zekere i;
3. $F = \text{MSA}/\text{MS(error)}$;
4. F-verdeling met 3 vrijheidsgraden in de teller en 40 vrijheidsgraden in de noemer;
5. waarde van $F : 10.29$;
6. kritieke waarde: 2.84; kritieke gebied: $F \geq 2.84$;
7. verwerp H_0, want $10.29 > 2.84$;
8. statistisch is aangetoond dat het winkeltype van invloed is op de verkoop.

Evenzo kan statistisch aangetoond worden dat het type verpakking een significant effect heeft. $\square$

Ook het schatten en het bepalen van de betrouwbaarheidsintervallen worden op analoge wijze als in het model zonder interactie uitgevoerd. Als schatter voor γ_{ij} nemen we

$$\begin{aligned} (3.3.11) \qquad & X_{ij\bullet} - X_{\bullet\bullet\bullet} - (X_{i\bullet\bullet} - X_{\bullet\bullet\bullet}) - (X_{\bullet j\bullet} - X_{\bullet\bullet\bullet}) = \\ & X_{ij\bullet} - X_{i\bullet\bullet} - X_{\bullet j\bullet} + X_{\bullet\bullet\bullet}, \end{aligned}$$

want $X_{ij\bullet}$ is een zuivere schatter van $\mu + \alpha_i + \beta_j + \gamma_{ij}, X_{\bullet\bullet\bullet}$ een zuivere schatter van μ, $X_{i\bullet\bullet} - X_{\bullet\bullet\bullet}$ een zuivere schatter van α_i en $X_{\bullet j\bullet} - X_{\bullet\bullet\bullet}$ een zuivere schatter van β_j, zodat (3.3.11) een zuivere schatter van $\mu + \alpha_i + \beta_j + \gamma_{ij} - \mu - \alpha_i - \beta_j = \gamma_{ij}$ oplevert.

De variantie van deze schatter is $\sigma^2(I-1)(J-1)/(IJK)$. Dus wordt het betrouwbaarheidsinterval voor γ_{ij} met betrouwbaarheid $1 - \alpha$ gegeven door

$$\Big(X_{ij\bullet} - X_{i\bullet\bullet} - X_{\bullet j\bullet} + X_{\bullet\bullet\bullet} - c\sqrt{\text{MS(error)}(I-1)(J-1)/(IJK)}, \\ X_{ij\bullet} - X_{i\bullet\bullet} - X_{\bullet j\bullet} + X_{\bullet\bullet\bullet} + c\sqrt{\text{MS(error)}(I-1)(J-1)/(IJK)}\Big)$$

met $P(T_{IJ(K-1)} \leq c) = 1 - \frac{1}{2}\alpha$, waarbij $T_{IJ(K-1)}$ een Student-verdeling heeft met $IJ(K-1)$-vrijheidsgraden.

Voorbeeld 3.3.13 (vervolg van 3.3.11 en 3.3.12) We bepalen het 95%-betrouwbaarheidsinterval voor γ_{23}. Berekening geeft $x_{23\bullet} = 4.333$, $x_{2\bullet\bullet} = 6.87$, $x_{\bullet 3\bullet} = 5.17$ en $x_{\bullet\bullet\bullet} = 7.02$. Uit voorbeeld 3.3.12 lezen we af MS(error) = 1.60. De tabel van de Student-verdeling geeft $c = 2.02$. Invullen levert het volgende interval op: $(-0.687 - 2.02\sqrt{0.32}, -0.687 + 2.02\sqrt{0.32}) = (-1.830, 0.456)$. □

Wanneer interactie significant aanwezig is, moet men bij de interpretatie van hoofdeffecten grote voorzichtigheid betrachten, want iedere cel heeft nu een eigen verwachting en bij sterke interactie kun je vrijwel geen betekenis meer geven aan hoofdeffecten.

Voorbeeld 3.3.14 (vervolg van voorbeeld 3.3.11, 3.3.12 en 3.3.13) Het 95%-betrouwbaarheidsinterval voor $\alpha_4 - \alpha_3$ wordt gegeven door ($\mathrm{var}(X_{4\bullet\bullet} - X_{3\bullet\bullet}) = 2\sigma^2/(JK)$)

$$\Big(X_{4\bullet\bullet} - X_{3\bullet\bullet} - c\sqrt{\mathrm{MS(error)} \times 2/(JK)},$$
$$X_{4\bullet\bullet} - X_{3\bullet\bullet} + c\sqrt{\mathrm{MS(error)} \times 2/(JK)}\Big)$$

met $P(T_{IJ(K-1)} \leq c) = 1 - \frac{1}{2}\alpha$. Invullen van $x_{4\bullet\bullet} = 8.53$, $x_{3\bullet\bullet} = 6.47$, $c = 2.02$, MS(error) = 1.60 geeft $(1.127, 2.993)$. Omdat 0 niet in dit interval ligt, is α_4 significant groter dan α_3. Het feit dat winkeltype 4 tot significant hogere verkoop leidt dan winkeltype 3 ($\alpha_4 > \alpha_3$), wil nog niet zeggen dat deze conclusie ook bij (bijna) ieder type verpakking geldt: het gemiddelde bij twee van de vijf types verpakkingen (verpakking type 3 en 5) valt bij winkeltype 3 hoger uit dan bij winkeltype 4! Dit komt door de sterke interactie. □

3.4 Residuen

Residuen zijn die grootheden die overblijven na een aanpassing. In het algemeen geldt

$$\boxed{\text{residu} = \text{gegeven} - \text{aangepaste waarde.}}$$

Bij additiviteit in tweevoudige tabellen (zonder interactie) is dit

$$R_{ij} = X_{ij} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j).$$

Met het additieve model proberen we met $I + J - 1$ schatters (namelijk van μ, $\alpha_1, \ldots, \alpha_{I-1}, \beta_1, \ldots, \beta_{J-1}$; $\hat{\alpha}_I$ en $\hat{\beta}_J$ volgen uit $\hat{\alpha}_1 + \ldots + \hat{\alpha}_I = 0$, $\hat{\beta}_1 + \ldots + \hat{\beta}_J = 0$) de IJ waarden in de tabel zo goed mogelijk te beschrijven. Dat lukt natuurlijk niet perfect. Het restgedeelte: waarde − schatter, dus $X_{ij} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)$ noemen we het **residu** R_{ij}. We gebruiken de residuen om

– de kwaliteit van de aanpassing te onderzoeken
– de aanpassing stapsgewijs te verbeteren
– transformaties te vinden, die een betere aanpassing geven
– uitschieters te traceren.

De kwaliteit van de aanpassing wordt afgemeten aan de grootte van de residuen **t.o.v. de aangepaste waarden**. De grootte van de residuen zelf zegt nog niets. Immers, als we bijv. metingen X hebben in millimeters en we geven deze metingen vervolgens met Z weer in meters, dan worden alle van belang zijnde grootheden, waaronder de residuen, 1000 maal zo klein. De kwaliteit van de aanpassing is uiteraard door deze andere weergave niet verbeterd, terwijl de residuen wel 1000 maal zo klein zijn geworden.

Een eerste indruk van de residuen krijgen we door er samenvattingen van te maken (zie hoofdstuk 1), eventueel gevolgd door Q-Q plots en/of toetsen voor aanpassing. Veel (klassieke) modellen hebben een storingsterm, die verondersteld is normaal verdeeld te zijn. De residuen kunnen ons vertellen of deze veronderstelling redelijk is (zie ook hoofdstuk 2). Verder is uit deze samenvattingen gewoonlijk makkelijk te achterhalen of er uitschieters zijn.

Uitschieters in de residuen
We spreken van uitschieters als er enkele uitzonderlijk grote residuen zijn (in absolute waarde) vergeleken met de rest. Preciezer: noem het laagste steekproefkwartiel van de residuen L en het hoogste steekproefkwartiel H, dan is de steekproefkwartielafstand $H - L$ en is R_{ij} een uitschieter als $R_{ij} \leq L - 1.5 \times (H - L)$ of als $R_{ij} \geq H + 1.5 \times (H - L)$. Zijn de uitschieters getraceerd, dan moet men trachten ze te verklaren en er mee rekening houden tijdens de verdere analyse.

Schaakbordpatronen
Het herkennen van dit soort patronen gaat als volgt. Rangschik de rijen en kolommen zó dat de rij- en kolomeffecten in numerieke volgorde staan van klein naar groot. We spreken van een schaakbordpatroon als één van onderstaande patronen optreedt.

+	−		−
−	+		+
−	+		

of

−	+	−
+	−	+
−	+	

In een dergelijk geval is het verstandig te zoeken naar een transformatie van de data, die dit patroon opheft. Daarmee wordt dan een structureel effect in de data opgespoord en in het model verwerkt. Het zoeken naar een geschikte transformatie wordt vereenvoudigd door het toepassen van de diagnostische plot. Deze wordt besproken in 3.5.

3.5 Opgaven

Pas bij het uitvoeren van toetsen steeds de 8 stappen toe uit Statistiek & kansrekening (zie ook hoofdstuk 2)

1. Een onderzoek naar de kwaliteit van de bossen leverde o.a. de volgende gegevens op over de hoogte van aselect gekozen bomen.

	Bos 1	Bos 2	Bos 3
gemiddelde hoogte in meters	20.3	23.1	22.4
steekproef-standaardafwijking	3.41	5.79	4.72
steekproef-grootte	34	76	29

   a. Bereken het totale steekproefgemiddelde $x_{\bullet\bullet}$.

   b. Bereken SS(factor).

   c. Bereken SS(error).

   d. Geef de ANOVA-tabel.

   e. Toets de nulhypothese van geen verschil in verwachte hoogte tussen de bossen bij onbetrouwbaarheidsdrempel $\alpha = 0.05$.

   f. Ga m.b.v. Tukey-Kramer's methode na welke verwachtingen verschillen. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

   g. Geef het 95%-betrouwbaarheidsinterval voor de verwachte hoogte in bos 2.

h. Geef het 95%-betrouwbaarheidsinterval voor de overall verwachte hoogte.

i. Geef het 95%-betrouwbaarheidsinterval voor het verschil in verwachte hoogte van bos 1 en bos 2.

j. Geef 95% simultane betrouwbaarheidsintervallen voor de verschillen in verwachte hoogte.

k. Geef de modelveronderstellingen, waarop alle voorgaande antwoorden gebaseerd zijn. Zijn deze modelveronderstellingen met bovenstaande gegevens te controleren?

2. Men wenst het effect van twee factoren op de slijtvastheid van rubber na te gaan. Factor A betreft het voorbehandelen van de rubber. Dit wordt volgens drie methoden uitgevoerd. Factor B is de grondstof. Deze factor wordt op vier niveaus gevarieerd. De getallen in de tabel zijn een maat voor de slijtvastheid van rubber. Hoe groter het getal, des te beter is de slijtvastheid.

	grondstof			
voorbehandelen	1	2	3	4
1	626	241	55	730
2	766	441	340	780
3	979	698	344	1083

a. Geef de modelveronderstellingen voor de two-way analysis of variance, en geef interpretaties van de parameters in de context van deze opgave.

b. Bereken de geschatte rij- en kolom-effecten, de geschatte overall waarde $\hat{\mu}$ en de residuen. Presenteer dit geheel in tabelvorm zoals in voorbeeld 3.3.6.

c. Geef commentaar op de resultaten.

d. Bereken de ANOVA-tabel.

e. Toets of het voorbehandelen effect heeft. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

f. Toets of er sprake is van een grondstof-effect. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.01$.

g. Bereken het 95%-betrouwbaarheidsinterval voor het verwachte effect van grondstof 3.

h. Bereken het 95%-betrouwbaarheidsinterval voor μ. Geef een interpretatie van het resultaat.

i. Bereken het 95%-betrouwbaarheidsinterval voor het verwachte verschil tussen voorbehandelingsmethode 1 en 3.

j. Geef 95% simultane betrouwbaarheidsintervallen voor de verwachte verschillen in het effect van voorbehandelen. Geef commentaar op het resultaat.

k. Leg uit waarom het interval voor het verwachte verschil tussen voorbehandelingsmethode 1 en 3 bij onderdeel i anders is dan bij j.

l. Geef 95% simultane betrouwbaarheidsintervallen voor de verwachte verschillen in grondstof-effect. Geef commentaar op het resultaat.

3. De volgende tabel geeft de consumptie in aantal kilogrammen per hoofd van de bevolking voor de jaren 1965, 1970, 1977, 1979 en voor de producten kaas, vlees, vis en verse zuidvruchten.

	1965	1970	1977	1979
kaas	8.7	9.3	12.4	12.9
vlees	46.9	48.9	55.8	61.7
vis	10.4	10.5	12.0	12.2
zuidvr.	25.0	27.1	31.3	32.5

a. Pas de logaritmische transformatie toe op de data. Bereken voor de getransformeerde data de geschatte rij- en kolomeffecten, de geschatte overall-waarde $\hat{\mu}$ en de residuen. Presenteer dit geheel in tabelvorm zoals in voorbeeld 3.3.6.

b. Geef een interpretatie van de geschatte rij- en kolomeffecten.

4. De volgende ANOVA-tabel hoort bij een additief twee-factor model met 19 niveaus voor factor A en 3 voor factor B.

Source	df	SS	MS	F
Factor A		151.75		
Factor B		103.37		
Error		35.24		
Totaal		290.36		

a. Vul de openstaande plaatsen in de tabel aan.

b. Toets of er sprake is van een significant effect van factor A. Neem als onbetrouwbaarheidsdrempel 0.05.

5. Men onderwerpt 5 groepen van ieder 4 cavia's aan 4 verschillende vormen van dieet. De 4 cavia's binnen een groep worden als gelijkwaardig beschouwd. Iedere cavia binnen een gegeven groep krijgt een ander dieet. De gegevens betreffende de gewichtstoename in grammen leveren de volgende ANOVA-tabel op.

Source	df	SS	MS	F
Dieet		8.15		
Groep		0.39		
Error		0.78		
Totaal		9.32		

a. Beschrijf het model voor variantie-analyse, toegespitst op deze situatie, inclusief interpretatie van de parameters.

b. Vul de open plaatsen in de ANOVA-tabel aan.

c. Toets of er verschil is tussen de vormen van dieet. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

d. Is het nodig om onderscheid tussen de groepen te maken? Zo nee, welk eenvoudiger model kan gehanteerd worden en hoe luidt het antwoord op vraag c in dit model?

6. Een fabrikant wenst een machine aan te schaffen en kan kiezen uit drie verschillende typen. Van ieder type ontvangt hij een exemplaar om uit te proberen. Hij laat 4 van zijn personeelsleden ieder één dag aan elk van de 3 typen machines werken. De opbrengsten per machine en per man staan hieronder

	werknemer			
machine	1	2	3	4
1	37	12	27	28
2	49	41	45	37
3	16	22	12	34

a. Beschrijf het model voor variantie-analyse, toegespitst op deze situatie en geef in woorden weer wat de parameters betekenen.

b. Geef schattingen van de parameters μ, α_i en β_j.

7. Beschouw het additieve model

$$X_{ij} = \mu + \alpha_i + \beta_j + U_{ij}$$

met

$$\sum_{i=1}^{I} \alpha_i = 0, \ \sum_{j=1}^{J} \beta_j = 0 \text{ en } U_{11}, \ldots, U_{IJ} \text{ o.o. } N(0, \sigma^2)\text{-verdeeld.}$$

a. Leid een uitdrukking af voor $E\left\{\sum_i \sum_j (X_{\bullet j} - X_{\bullet\bullet})^2\right\}$.

b. Wat is het verschil tussen U_{ij} en het residu R_{ij}?

8. Een onderzoek naar het voorkeursgedrag van consumenten betrof drie verschillende manieren van het groeperen van artikelen bij 4 verschillende supermarkten. De resultaten waren als volgt:

supermarkt	manier van groeperen	aantal verkochte artikelen
1	1	171
2	3	214
3	1	16
4	2	228
1	3	239
2	1	156
3	2	237
4	1	65
1	2	340
2	2	262
3	3	80
4	3	168

a. Geef de data weer in een tweevoudige tabel.

b. De ANOVA-tabel luidt:

Source	df	SS	MS	F
supermarkten		33863.3		
groeperingen		54507.2		
error		5044.1		
totaal		93414.6		

Vul de tabel verder in en onderzoek m.b.v. een geschikte statistische toets met gebruikmaking van de ANOVA-tabel of er verschil is tussen de verschillende manieren van groepering van de artikelen. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

9. Men wil de effecten onderzoeken van 3 verschillende soorten toevoegsels op 3 verschillende soorten benzine. Voor ieder van de 9 mogelijke combinaties bekijken we het aantal afgelegde kilometers (bij dezelfde hoeveelheid brandstof).

	toevoegsel		
benzine	1	2	3
1	124.1	131.5	127.0
2	126.4	130.6	128.4
3	127.2	132.7	125.6

a. Geef een kansmodel dat dit experiment beschrijft. Geef in woorden de betekenis van de voorkomende parameters.

b. Geef het 95%-betrouwbaarheidsinterval voor het verwachte effect van toevoegsel 2.

c. Geef 99%-simultane betrouwbaarheidsintervallen voor de verschillen in verwacht effect van de soorten benzine.

d. Geef het 99%-betrouwbaarheidsinterval voor de verwachte overall waarde.

e. Stel de ANOVA-tabel op.

f. Toets de hypothese dat de toevoegsels gelijkwaardig zijn. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

10. Twee verschillende medewerkers moeten ieder 12 taken uitvoeren. De tijden (in minuten) die zij nodig hebben om de taken uit te voeren staan hieronder.

	taak											
	1	2	3	4	5	6	7	8	9	10	11	12
medewerker 1	72	56	83	42	35	61	50	77	28	55	65	42
medewerker 2	77	54	85	47	30	52	61	84	35	48	60	48

Men wenst te onderzoeken of er een systematisch verschil in tijd is tussen medewerker 1 en 2. De ANOVA-tabel is als volgt

Effecten	df	SS	MS	F
medewerker		9.375		
taak		6392.125		
residuen		237.125		
totaal		6638.625		

a. Formuleer het kansmodel, waarmee dit experiment wordt beschreven en vermeld in woorden de betekenis van de parameters.

b. Vul de ANOVA-tabel verder in.

c. Toets of er systematisch verschil is tussen medewerker 1 en 2. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

d. Zij X_{1j} de tijd die medewerker 1 nodig heeft voor taak j en zij X_{2j} de tijd die medewerker 2 nodig heeft voor taak j ($j = 1, \ldots, 12$). Een eenvoudiger model wordt verkregen door te veronderstellen dat we twee onafhankelijke aselekte steekproeven uit normale verdelingen met gelijke varianties hebben. De tijden die medewerker 1 nodig heeft zijn nu realisaties van een aselekte steekproef van X_1 die $N(\mu_1, \sigma^2)$-verdeeld is. De tijden die medewerker 2 nodig heeft zijn realisaties van een aselekte steekproef van X_2 die $N(\mu_2, \sigma^2)$-verdeeld is. Het eenvoudiger model luidt dus

$X_{ij} = \mu_i + U_{ij}$ met U_{ij} o.o. $N(0, \sigma^2)$-verdeelde stochastische variabelen ($i = 1, 2, j = 1, \ldots, 12$). Waarom is het bezwaarlijk dit eenvoudiger model te nemen?

11. Een nieuw type camera wordt in 4 regio's geïntroduceerd. Binnen elke regio zijn 3 deelgebieden te localiseren met verschillend concurrentie-niveau. De wekelijkse (gecodeerde) verkoop van de camera's is hieronder weergegeven voor een periode van 4 weken

	concurrentie-niveau											
regio	1				2				3			
1	4	3	2	3	8	8	6	7	12	10	9	11
2	8	7	6	6	3	2	1	4	6	7	8	7
3	10	14	13	15	3	3	2	4	3	2	3	4
4	3	2	1	2	15	16	12	14	8	7	7	6

De ANOVA-tabel is als volgt

Source	SS	df	MS	F
regio	34.729			
conc.niveau			2.146	
interactie				87.19
error	50.250			
total				

a. Vul de ANOVA-tabel verder in.

b. Toets of er sprake is van interactie. Neem als onbetrouwbaarheidsdrempel $\alpha = 0.05$.

c. Bereken het 95%-betrouwbaarheidsinterval voor het interactie-effect bij regio 2 en concurrentieniveau 1.

d. Bereken het 95%-betrouwbaarheidsinterval voor het verschil in hoofdeffect van regio 4 en regio 2.

e. Op grond van onderdeel d. kan geconcludeerd worden dat de verkoop in regio 4 hoger is dan in regio 2. Waarom?

f. Betekent de conclusie in onderdeel e. ook dat op elk concurrentie-niveau regio 4 beter scoort dan regio 2? Argumenteer uw antwoord.