Voorwoord

De serie *Statistiek voor de psychologie* is, zoals de titel al zegt, een inleiding in de statistiek toegespitst op opleidingen psychologie. De serie bestaat uit vier delen en heeft de volgende uitgangspunten. Elke analyse wordt behandeld volgens dezelfde structuur, het 'elementaire rapport'. Dat maakt het leren veel gemakkelijker. Je wordt systematisch getraind op: datadesign en vraagstelling; betekenis en formulering van de conclusies; causale interpretaties en hun beperkingen. Je leert daarmee niet alleen het 'hoe' maar ook het 'wanneer', 'welke' en 'waarom' van de analyses op een praktische manier. Bij veel analyses wordt uitgebreid aandacht besteed aan visualiseren: het 'lezen' en inzichtsmatig redeneren met figuren. De stof begint extreem gemakkelijk (hoe bereken je een gemiddelde) maar eindigt op hoog niveau (dubbel multivariate repeated-measures ANOVA). In de eerste twee delen wordt uitvoerig stilgestaan bij de beperkingen van het intuïtieve statistische denken. Er is ook veel aandacht voor praktische regels. Daarnaast staan er geen onnodige formules in de boeken. De stof wordt uitgelegd met gezond verstand en veel gevarieerde voorbeelden, van het dagelijks leven tot gepubliceerd onderzoek. Met name onderzoeken die prototypisch zijn voor een bepaald gebied worden gebruikt. Tot slot is de stof vijf jaar lang getest en verbeterd naar aanleiding van vragen, problemen en suggesties van studenten van de opleiding Psychologie van de RU Nijmegen.

Op de productpagina van dit boek op www.boomlemma.nl is extra materiaal te vinden.

Bij deze wil ik Dick Willems bedanken voor het zoeken en beschrijven van diverse interessante voorbeeldonderzoeken, met name in hoofdstuk 1 en 2. Ook aan hoofdstuk 4 heeft hij diverse bijdragen geleverd. Daarnaast wil ik de studentassistenten bedanken die vanaf 1997 met hun commentaar, suggesties, tips en verbeteringen hebben bijgedragen aan dit boek. Ook veel studenten hebben hieraan bijgedragen, al was het maar door aan te geven welke passages ze niet snapten. Hun grootste bijdrage lag in het positieve commentaar. Het was altijd prettig om te horen dat ze het boek duidelijk vonden, dat ze het vaak leuk vonden om te lezen, en dat ze mijn humor waardeerden. Dat stimuleerde me om ermee door te gaan.

Jules L. Ellis
Nijmegen, augustus 2013

Inhoud

1 1-factor-ANOVA

1.1 Inleiding

Achtergrond

ANOVA (analysis of variance) oftewel 'variantieanalyse' is de verzamelnaam voor een grote hoeveelheid statistische procedures die erop gericht zijn verschillen tussen gemiddelden te onderzoeken. De naam ANOVA is wat dit betreft dus misleidend want het gaat helemaal niet om varianties. Wel is het zo dat er veel varianties in de analyse gebruikt worden. Het doel blijft echter: uitspraken doen over gemiddelden.

Omdat in ANOVA veel varianties gebruikt worden, raden we je aan de stof hierover in deel 1A te herlezen. Het komt erop neer dat de variantie van een stel scores gelijk is aan het kwadraat van de standaardafwijking van die scores. De standaardafwijking is een maat voor spreiding en de variantie is dus een gelijkwaardige maat voor spreiding: als de standaardafwijking groot is dan is ook de variantie groot, en omgekeerd.

De ANOVA-vorm die hier wordt behandeld is de eenvoudigste, de zogenaamde **1-factor-** of **1-weg-** of **simpele** ANOVA. Dit is een generalisatie van de *t*-toets voor onafhankelijke steekproeven naar designs met meer dan twee groepen.

Doel

Na bestudering van de paragrafen 1.1 tot en met 1.4 en het maken van de bijbehorende opgaven moet je in staat zijn een **elementair rapport van een 1-weg-**ANOVA te maken. Dit omvat:

- datadesign
- mate van controle
- spreidingsdiagram
- geaggregeerde data
- hypothesen
- ANOVA-tabel
- beslissing
- causale interpretatie

Alvorens deze zaken te beschrijven, geven we eerst een beknopte beschrijving van ANOVA en introduceren we het voorbeeld dat doorlopend gebruikt zal worden.

1.2 Beknopte beschrijving van een 1-weg-ANOVA

Een 1-weg-ANOVA dient om verschillen tussen **gemiddelden** te onderzoeken in een **between-subjectdesign** met twee of meer groepen. Vaak zal er sprake zijn van een experiment waarbij deze groepen onder verschillende condities zijn gemeten. Als voorbeeld nemen we een experiment waarbij iedere proefpersoon een aantal borrels krijgt (0, 3 of 5) waarna zijn prestatie op een evenwichtstaak wordt gemeten. Het aantal borrels dat iemand krijgt noemen we de **factor** Alcohol, zijn score op de evenwichtstaak noemen we de **meting** Evenwicht. In elke groep met 0, 3 of 5 borrels bepalen we het gemiddelde van de scores op Evenwicht. De nulhypothese zegt nu dat deze gemiddelden in de populatie **allemaal gelijk** zijn. De alternatieve hypothese zegt dat ten minste twee van deze gemiddelden verschillend zijn.

De kern van de ANOVA wordt gevormd door de zogenaamde **summary table** (ANOVA-tabel). Deze ziet er bijvoorbeeld uit zoals tabel 1.1.

Tabel 1.1 ANOVA-tabel voor Evenwicht

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between Aantal borrels	2	900	450	4.5	< .025	.25
Within	27	2700	100			
Total	29	3600				

Hiervan zijn vooral de *p*-waarde en de R^2-waarde van belang. De overige getallen zijn slechts tussenuitkomsten die op zichzelf weinig zeggen. De ***p*-waarde** geeft aan hoe aannemelijk of **houdbaar** de nulhypothese is. Aangezien de *p*-waarde hier nogal klein is, concluderen we dat de populatiegemiddelden verschillend zijn: Alcohol heeft een effect op Evenwicht. Hoe groot dat effect is, wordt aangegeven met de ***R*²-waarde**. Dit is een maat voor **sterkte van de samenhang** tussen de factor en de meting. De R^2-waarde van .25 wil zeggen dat de variatie in Evenwicht voor 25% wordt veroorzaakt doordat het Alcohol-niveau werd gevarieerd. De overige 75% wordt veroorzaakt door andere, onbekende factoren zoals individuele verschillen tussen de subjecten of onopgemerkte variaties binnen de condities.

Van de tussenuitkomsten zullen vooral de Sums of Squares (*SS*) onze aandacht krijgen. SS_{Total} is de zogenaamde **variatie** van de meting. Deze is eenvoudig gedefinieerd als (N - 1) maal de varia**N**tie van de meting. In dit voorbeeld deden N = 30 subjecten mee en in de tabel zien we dat SS_{Total} = 3600. We mogen dus concluderen dat de variantie van de meting gelijk was aan 124.14. De overige *SS*-en geven aan hoeveel van deze variatie is toe te schrijven aan het feit dat het Alcohol-niveau werd gevarieerd ($SS_{Between}$) dan wel aan onbedoelde variaties van overige, onbekende factoren (SS_{Within}). In plaats van $SS_{Between}$ en SS_{Within} worden ook vaak de (algemenere) termen SS_{Model} en SS_{Error} gebruikt.

Doorlopend voorbeeld

De laatste tijd krijgt het verschijnsel pesten op school steeds meer aandacht in de media. Het heeft een slechte reputatie. Vaak wordt verwacht dat kinderen die op school gepest worden daar hinder van ondervinden op sociaal gebied. Een Nijmeegse psycholoog is van plan deze relatie te onderzoeken. Hiervoor neemt hij bij 42 kinderen de *social-isolationschaal* af. Een hoge score op deze schaal betekent een hoge mate van sociaal isolement, een lage score betekent een lage mate van sociaal isolement. Vervolgens deelt hij deze kinderen op grond van sociometrische gegevens (dit zijn gegevens over de sociale status van kinderen die verkregen zijn via andere kinderen) in naar hun Treiterstatus: bully (pest veel), victim (wordt veel gepest) en non involved (pest niet veel en wordt niet veel gepest). De onderzoeksvraag is of de verschillende groepen (bully, victim, non involved) verschillende gemiddelden hebben met betrekking tot Sociale isolatie. Een volgende vraag kan dan zijn of we ook kunnen stellen dat Treiterstatus *invloed* heeft op Sociaal isolement. De data van het onderzoek staan vermeld in tabel 1.2 (bron: Sectie Ontwikkelingspsychologie, KUN).

Tabel 1.2 Sociale-isolatie-scores gegroepeerd naar Treiterstatus

	Treiterstatus	
non involved	*bully*	*victim*
1.75	0.75	1.00
0.25	0.25	2.00
0.25	0.75	0.75
1.25	0.50	3.25
0.75	0.50	0.50
0.50	0.50	2.25
0.25	0.75	3.00
1.00	0.25	1.00
0.75	0.25	1.75
1.00		1.75
1.25		0.25
0.25		0.25
1.50		1.75
0.50		
0.25		
0.25		
0.25		
0.50		
1.75		
0.25		

1.3 Elementair rapport van een 1-weg-ANOVA

1.3.1 Datadesign

Bij het datadesign beschrijf je de aard van de data die in de analyse gebruikt zullen worden. Eventuele overtollige data komen daarbij niet aan de orde. De beschrijving omvat het volgende.

Wat betreft de afhankelijke variabele:
- de naam van de variabele
- het meetniveau (in dit geval kwantitatief)
- het aantal metingen per persoon (in dit geval één)

Wat betreft de onafhankelijke variabele:
- dat het een factor is
- het 'domein' van de factor (in dit geval between-subject)
- de naam van de factor
- de niveaus van de factor

Toelichting
De 1-factor-ANOVA gebruik je als er sprake is van een **between-subjectdesign** met **twee of meer groepen**. In elke groep zitten meerdere subjecten en elk subject heeft een score die iets van zijn of haar gedrag meet. Nader beschouwd is er in het between-subjectdesign sprake van twee variabelen:
1 De *factor*. De variabele die aangeeft in welke groep iemand zit noemen we de **factor**. Dit is een kwalitatieve, onafhankelijke variabele. Een specifieke groep (bijvoorbeeld groep 2, als er drie groepen zijn) noemen we een **niveau** van de factor.
2 De *meting*. Daarnaast dient elk subject één score te hebben. De scores van de subjecten noemen we de **meting**. Dit dient een kwantitatieve variabele te zijn. Men gebruikt vaak deze notatie: X_{ki} = de score van *het* i-*de subject in groep* k.

In de ANOVA gaat het erom de **samenhang** te bestuderen tussen de factor en de meting. Daarbij fungeert de factor als onafhankelijke variabele (de vermoedelijke oorzaak) en de meting fungeert als afhankelijke variabele (het vermoedelijke gevolg). Het causale model waar we in eerste instantie van uitgaan is dus:

factor → meting

Voorbeelden
In het doorlopende voorbeeld:

afhankelijke variabele = Sociale isolatie (kwantitatief, één meting per persoon)

onafhankelijke variabele:
between-subjectfactor = Treiterstatus (bully/victim/non involved)

Andere voorbeelden zijn gegeven in tabel 1.3.

Tabel 1.3 Voorbeelden van designs met één factor

Vraagstelling	*Factor*	*Niveaus*	*Meting*
Verschillen mannen en vrouwen in gevoeligheid voor impliciete humor?	geslacht	man, vrouw	gevoeligheid voor impliciete humor (per persoon)
Verschillen leden van diverse kerken in houding t.a.v. abortus?	kerk	katholiek, protestant, moslim	houding t.a.v. abortus (per persoon)
Heeft een geconsumeerde hoeveelheid alcohol invloed op het spreekvolume?	hoeveelheid alcohol	0, 1-3, 3-5, 5 of meer borrels	decibelniveau van het spreken
Heeft de geconsumeerde hoeveelheid alcohol invloed op de evenwichts–prestatie?	hoeveelheid alcohol	0, 3 of 5 borrels	evenwichts-prestatie
Verschillen leden van verschillende jeugdculturen in houding t.a.v. drugs?	jeugdcultuur	gabber, skater, rocker, alto	houding t.a.v. drugs (per persoon)
Heeft de intensiteit van de straf na het maken van een fout invloed op de gemiddelde nauwkeurigheid?	intensiteit van een schok	10 volt, 20 volt, 30 volt	prestatie op een nauwkeurigheidstaak

1.3.2 Mate van controle

Hier beschrijf je of het onderzoek passief-observerend of experimenteel is, dan wel iets daartussenin.

Toelichting
Bij ANOVA gaat het er meestal om te onderzoeken of de factor invloed heeft op de meting. Dit zou tot uiting komen doordat het gemiddelde van de meting varieert met het niveau van de factor. Dus fungeert de factor als onafhankelijke variabele en de meting als afhankelijke variabele. De mate van controle slaat op de mate waarin de onafhankelijke variabele **adequaat is gemanipuleerd** ten behoeve van de causale interpretatie. Er zijn twee uitersten:

1 *Passief-observerend*. Dit is het geval als de onderzoeker geen invloed heeft op de waarde van de onafhankelijke variabele voor een gegeven persoon. Er is dan sprake van *natuurlijke groepen* die voor het onderzoek al bestonden.
2 *Experimenteel*. Dit is het geval als de subjecten door de onderzoeker willekeurig (at random) aan condities worden toegewezen en er tevens een aantal andere maatregelen zijn genomen om de invloed van overige factoren onder controle te houden. Er is dan sprake van meerdere *experimentele groepen*. Voor een zuiver experiment is nodig:

- *manipulatie* van de onafhankelijke variabele;
- *randomisatie* van de subjecten over de condities;
- *controle* van de storende factoren (randomiseren dan wel fixeren op één waarde);
- *meting* van de afhankelijke variabele.

Er is ook een tussenvorm: het is mogelijk om een manipulatie uit te voeren zonder dat er is gerandomiseerd.

Voorbeelden
Het onderzoek over pesten is een passief-observerend onderzoek. De kinderen zijn niet at random over de condities non involved, bully en victim verdeeld. Dit is overigens ook niet mogelijk.

Andere voorbeelden staan in tabel 1.4 (let vooral op het verschil tussen 3 en 4).

Tabel 1.4 Voorbeelden van de mate van controle

	Factor	*Niveaus*	*Meting*	*Toewijzing van niveau*	*Mate van controle*
1	geslacht	man, vrouw	gevoeligheid voor impliciete humor	natuur	passief-obs.

	Factor	*Niveaus*	*Meting*	*Toewijzing van niveau*	*Mate van controle*
2	kerk	katholiek, protestant, moslim	houding t.a.v. abortus	keuze van subjecten	passief-obs.
3	alcohol (in een café)	0, 1-3, 3-5, 5 of meer borrels	decibelniveau van het spreken	keuze van subjecten in een café	passief-obs.
4	alcohol (in een lab)	0, 3 of 5 borrels	evenwichts-prestatie	randomi-seren	experiment
5	jeugdcultuur	gabber, skater, rocker, alto	houding t.a.v. drugs	keuze van subjecten	passief-obs.
6	intensiteit van de schok	10 volt, 20 volt, 30 volt	prestatie op een nauw-keurigheidstaak	randomi-seren	experiment

1.3.3 Spreidingsdiagram

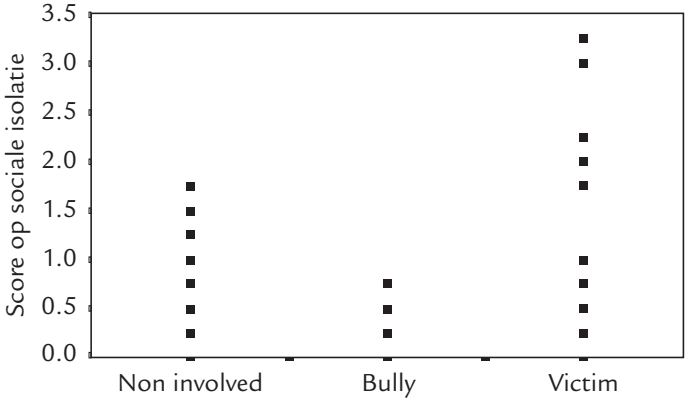
Hier geef je de data weer in een spreidingsdiagram met op de *x*-as de factor en op de *y*-as de meting (de afhankelijke variabele).

Toelichting

De data van een 1-weg-ANOVA kunnen worden weergegeven in een spreidingsdiagram. Het enige verschil met een normaal spreidingsdiagram is dat de onafhankelijke variabele nu kwalitatief is. De verschillende niveaus van de factor worden op de horizontale as uitgezet. De waarde van de meting wordt op de verticale as uitgezet. Voor ieder subject wordt een punt in de figuur gezet. De horizontale coördinaat geeft aan in welke groep het subject zit. De verticale coördinaat is de score van het subject.

Een aanname bij ANOVA is dat binnen elke groep de afhankelijke variabele normaal verdeeld is. Dat zal nader worden besproken in paragraaf 1.3.9. Aan de hand van het spreidingsdiagram kan visueel worden geïnspecteerd of er geen grote uitschieters zijn binnen een groep. Als die er wel zijn, kan het zijn dat ANOVA niet de beste analyse is.

Voorbeeld



Figuur 1.1 Spreidingsdiagram van Sociale isolatie op Treiterstatus

1.3.4 De geaggregeerde data

Om met de analyse te kunnen beginnen hebben we per groep nodig: het gemiddelde, de variantie en de steekproefgrootte. Deze noteren we als:

$\bar{x}_k$ = het **gemiddelde** in groep *k*
s_k^2 = de **variantie** in groep *k*
n_k = de **steekproefgrootte** in groep *k*

Daarnaast bepalen we:

$\bar{x}_\bullet$ = het **totaalgemiddelde**
N = de **totale steekproefgrootte**

(Met subscript $\bullet$ geeft men aan dat er wordt gemiddeld.) Verder is het raadzaam om deze uitkomsten overzichtelijk, in een tabelletje, op te schrijven.

Toelichting
De steekproefgemiddelden hebben we uiteraard nodig om iets te zeggen over de populatiegemiddelden. De varianties en steekproefgrootten hebben we nodig om iets te zeggen over de betrouwbaarheid van de steekproefgemiddelden. Dat laatste is nodig voor de statistische inferentie, de toetsing.

Voor een betrouwbare toepassing van ANOVA is deze vuistregel van belang: de verhouding tussen de grootste en de kleinste standaardafwijking mag hoogstens 2 zijn. Dit zal nader worden besproken in paragraaf 1.3.9.

Voorbeeld

Tabel 1.5 Geaggregeerde data van Sociale isolatie

Maat	*Non involved*	*Bully*	*Victim*	*Totaal*
Gemiddelde	$\overline{x}_1 = 0.725$	$\overline{x}_2 = 0.500$	$\overline{x}_3 = 1.500$	$\overline{x}_\bullet = 0.917$
Variantie	$s_1^2 = 0.282$	$s_2^2 = 0.047$	$s_3^2 = 0.958$	
Grootte	$n_1 = 20$	$n_2 = 9$	$n_3 = 13$	N = 42

Bij het totaalgemiddelde is in tabel 1.5 het zogenaamde 'gewogen' gemiddelde vermeld. Daarbij is het eerste groepsgemiddelde 20 keer meegeteld omdat $n_1 = 20$, het tweede groepsgemiddelde is 9 keer meegeteld omdat $n_2 = 9$, enzovoort.

Er moet worden gecontroleerd of de varianties ongeveer gelijk zijn: (grootste sd) / (kleinste sd) = $\sqrt{0.958} / \sqrt{0.047} = 4.51 > 2$, dus ANOVA is misschien niet de beste analyse. Maar omdat de berekeningen van een ANOVA hier nu eenmaal moeten worden geïllustreerd, gaan we er toch mee door.

1.3.5 De hypothesen

Hier specificeer je de hypothesen die onderzocht worden. H_0 houdt altijd in dat al deze groepen hetzelfde gemiddelde hebben in de populatie. H_a houdt in dat ten minste twee groepen een verschillend gemiddelde hebben in de populatie. De populatiegemiddelden geven we aan met:

μ_k = het populatiegemiddelde van groep *k*

De hypothesen zijn dus:

H_0 : $\mu_1 = \mu_2 = \ldots = \mu_K$
H_a : ten minste twee van deze gemiddelden verschillen

Toelichting
Het doel van 1-factor-ANOVA is om na te gaan of in de populatie de onafhankelijke variabele (de factor) invloed heeft op het gemiddelde van de afhankelijke variabele (de meting). Dit houdt in dat de verschillende groepen verschillende gemiddelden van de scores zouden hebben. De nulhypothese zegt van niet, de alternatieve hypothese zegt van wel. De alternatieve hypothese is altijd **ongericht** bij ANOVA.

Voorbeeld
Bij het onderzoek over pesten zien de hypothesen er als volgt uit:

H_0 : $\mu_{\text{non involved}} = \mu_{\text{bully}} = \mu_{\text{victim}}$
H_a : ten minste twee van deze gemiddelden verschillen

Andere voorbeelden

Een aantal voorbeelden van mogelijke populatiegemiddelden staan in tabel 1.6. Bij sommige voorbeelden is H_0 waar, bij andere niet. Merk op dat de gemiddelden niet allemaal verschillend hoeven te zijn als H_0 onwaar is.

Tabel 1.6 Relatie tussen populatiegemiddelden en nulhypothese

Populatiegemiddelde					
μ_1	μ_2	μ_3	μ_4	H_0	*Opmerking*
2	2	2	2	waar	
2	3	2	2	onwaar	er hoeft slechts 1 gemiddelde af te wijken
2	2	2	2.01	onwaar	ook heel kleine verschillen tellen mee
30	2	40	1	onwaar	het verband hoeft niet lineair te zijn
1	2	3	4	onwaar	het verband mag lineair zijn

1.3.6 De ANOVA-tabel

In de ANOVA-tabel wordt stapsgewijs de toetsingsgrootheid uitgerekend. In de titel van de tabel wordt de naam van de afhankelijke variabele genoemd. De tabel bestaat uit zes kolommen, met de kopjes Bron, *df*, *SS*, *MS*, *F*, *p*, en R^2. De tabel wordt van links naar rechts doorgerekend. De kolommen zullen in die volgorde worden besproken.

Voorbeeld

Tabel 1.7 ANOVA-tabel voor Sociale isolatie (1)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between Treiterstatus	2	6.72	3.36	7.60	< .01	.28
Within	39	17.24	0.44			
Total	41	23.96				

Bron

Als bronnen schrijf je op: Between, Within, en Total. Bij de Between-bron schrijf je ook de naam van de betrokken factor. De Within-bron wordt ook vaak Error genoemd.

Toelichting

De scores op de meting zijn in principe allemaal verschillend. Deze 'variatie' wordt bij een 1-factor-ANOVA naar twee bronnen uitgesplitst. Dit zijn: variatie **tussen** groepen en variatie **binnen** groepen. Deze twee samen veroorzaken de **totale** variatie. Deze drie namen moet je dus in de kolom Bron zetten. Vaak worden de Engelse ter-

men 'Between' en 'Within' gebruikt. Voor de duidelijkheid is het goed bij 'Between' ook de naam van de factor te schrijven, zodat de lezer gelijk ziet waar het over gaat.

Variatie *tussen* groepen heeft betrekking op verschillen tussen groepsgemiddelden. Variatie *binnen* groepen heeft betrekking op verschillen tussen individuele scores uit dezelfde groep.

Voorbeeld
Invullen van de eerste kolom van de ANOVA-tabel levert tabel 1.8 op.

Tabel 1.8 ANOVA-tabel voor Sociale isolatie (2)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between						
Treiterstatus						
Within						
Total						

Degrees of freedom *(df)*
De degrees of freedom *(df)* zeggen waar we straks in de tabel de *p*-waarde moeten zoeken. Voor elke bron, tussen en binnen, is er een eigen *df*. Je kan ze al direct berekenen op grond van informatie van het **design**, nog voor de data bekend zijn.

De formules zijn:

$df_{Between}$ = aantal groepen - 1
df_{Within} = aantal personen - aantal groepen
df_{Total} = aantal personen - 1

Voorbeeld
In dit geval is het aantal groepen gelijk aan 3 en het aantal personen gelijk aan 42. Dus $df_{Between}$ = 3 - 1 = 2 en df_{Within} = 42 - 3 = 39. Invullen levert tabel 1.9 op.

Tabel 1.9 ANOVA-tabel voor Sociale isolatie (3)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between	2					
Treiterstatus						
Within	39					
Total	41					

Sum of squares (*SS*)
Voor elke bron (tussen, binnen, en totaal) is er een zogenaamde sum of squares (*SS*). Deze *SS* geeft aan hoeveel **variatie** door die bron wordt veroorzaakt. Daarbij is het begrip variatie in alle gevallen gedefinieerd als:

SS = variatie = varia**N**tie $*$ (N - 1)

Voor verschillende *SS*-en worden echter **verschillende soorten scores** als invoer gebruikt. Uit de formule blijkt dat een *SS*, anders dan een variantie, automatisch toeneemt met het aantal subjecten. Tevens is een *SS* afhankelijk van de schaal waarop de afhankelijke variabele wordt uitgedrukt: als de ruwe scores in centimeters worden uitgedrukt zal een *SS* 10 000 keer zo groot zijn dan wanneer de ruwe scores in meters worden uitgedrukt. Om deze twee redenen is het bijvoorbeeld heel goed mogelijk dat een *SS* heel groot is, bijvoorbeeld groter dan 1 000 000. Daarom zegt de grootte van een *SS* op zichzelf weinig. Alleen de *verhoudingen* tussen verschillende *SS*-en zijn van belang.

Om $SS_{Between}$ te berekenen, voer je de **groepsgemiddelden** in als scores. Voor elk groepsgemiddelde neem je daarbij als frequentie het aantal personen in die groep. Vervolgens bereken je van deze would-be-scores de variantie (druk op de σ_{n-1}-toets of s_{n-1}-toets en kwadrateer). Deze variantie vermenigvuldig je met (N - 1). De uitkomst is $SS_{Between}$.

Om SS_{Within} te berekenen, vermenigvuldig je de **variantie** van elke groep met (n_k - 1), dat is het aantal personen in die groep min 1. Je krijgt dan voor elke groep de *SS* binnen die groep. Deze uitkomsten tel je bij elkaar op. Dat is SS_{Within}.

Ten slotte bereken je SS_{Total} door de twee voorgaande *SS*-en **op te tellen**:

$$SS_{Total} = SS_{Between} + SS_{Within}$$

De aldus berekende sums of squares geven aan in welke mate de scores worden beïnvloed door de volgende bronnen:

- de factor ($SS_{Between}$)
- overige, onbekende factoren, zoals individuele variaties (SS_{Within})
- deze bronnen samen (SS_{Total})

Toelichting

$SS_{Between}$

$SS_{Between}$ is *gedefinieerd* als de hoeveelheid variatie die er zou zijn als elk subject een score had die gelijk is aan het gemiddelde in zijn groep. Subjecten uit dezelfde groep zouden dan dus allen dezelfde score hebben.

De voorgeschreven berekening volgt precies de definitie. Dit leidt ertoe dat $SS_{Between}$ toeneemt naarmate de groepsgemiddelden meer van elkaar verschillen. De mate waarin de scores binnen een groep van elkaar verschillen, heeft echter geen invloed op $SS_{Between}$.

$SS_{Between}$ geeft daarom aan in welke mate de scores worden beïnvloed door de factor. Deze maat zal groot zijn als de groepsgemiddelden sterk verschillen in de steekproef. Dat duidt er dan op dat de alternatieve hypothese waar is. De factor heeft blijkbaar een groot *effect* op de scores. Als bijvoorbeeld protestanten gemiddeld veel meer lezen dan katholieken dan kun je zeggen dat de factor 'kerk' veel invloed heeft op de meting 'lezen'.

$SS_{Between}$ is ook een maat voor de **afstand** tussen de **groepsgemiddelden** enerzijds en het **totaalgemiddelde** anderzijds. Als je de definitie ontleedt, zie je namelijk dat:

$$SS_{Between} = \sum(\bar{x}_k - \bar{x}_\bullet)^2$$
$$= \text{som (groepsgemiddelde - totaalgemiddelde)}^2$$

Hierbij wordt een groepsgemiddelde *voor elk subject in de groep* een keer meegeteld. Grote afstanden tussen groepsgemiddelden en totaalgemiddelde zullen daardoor leiden tot een grote $SS_{Between}$. De groepsgemiddelden kunnen alleen van het totaalgemiddelde verschillen als ze ook van elkaar verschillen.

SS_{Within}

SS_{Within} is *gedefinieerd* als de hoeveelheid variatie die er zou zijn als van elke score het groepsgemiddelde werd afgetrokken. Dan zouden alle groepen dus hetzelfde gemiddelde hebben, namelijk 0. Wel zouden subjecten in een groep nog verschillende scores hebben.

De voorgeschreven berekening volgt niet letterlijk de definitie, maar levert wel dezelfde uitkomst. Zowel aan de definitie als aan de berekeningswijze is te zien dat SS_{Within} toeneemt naarmate de scores in de groepen meer van elkaar verschillen (de groepsvarianties groter zijn). De mate waarin de groepsgemiddelden van elkaar verschillen, heeft echter geen invloed op SS_{Within}.

SS_{Within} geeft daarom aan in welke mate de scores onderhevig zijn aan variaties die niet het gevolg zijn van de factor. Deze mate zal groot zijn als de varianties per groep groot zijn. Dat duidt erop dat de scores in belangrijke mate door andere, onbekende factoren worden beïnvloed. Daarbij moet je denken aan individuele verschillen (de ene persoon is nu eenmaal niet de andere persoon, daardoor kunnen ze verschillende scores hebben) en aan onopgemerkte verschillen in de condities (bijvoorbeeld doordat de scores op verschillende tijdstippen werden gemeten). De variatie binnen groepen kun je zien als een soort **ruis**. Deze vertroebelt het beeld van de populatie. Immers, als er grote individuele verschillen zijn in de scores, dan ligt het voor de hand dat – door puur toeval – ook de steekproefgemiddelden enigszins van elkaar verschillen.

SS_{Within} is ook een maat voor de **afstand** tussen de **individuele scores** enerzijds en de bijbehorende **groepsgemiddelden** anderzijds. Als je de definitie ontleedt, zie je namelijk dat:

$$SS_{Within} = \sum(x_{ki} - \bar{x}_k)^2$$
$$= \text{som (score - groepsgemiddelde)}^2$$

Grote afstanden tussen scores en groepsgemiddelden in een groep zullen daardoor leiden tot een grote SS_{Within}.

SS_{Total}

SS_{Total} is *gedefinieerd* als de hoeveelheid variatie in de scores, van alle groepen bij elkaar genomen. Om de SS_{Total} volgens deze definitie te berekenen, zou je als volgt te werk moeten gaan. Voer de ruwe scores in, van alle groepen bij elkaar, alsof je te maken hebt met één groep. Vervolgens bereken je de variantie van deze scores. Deze variantie vermenigvuldig je met (N - 1). De uitkomst is SS_{Total}.

De voorgeschreven berekening ($SS_{Between}$ en SS_{Within} optellen) is natuurlijk veel eenvoudiger en leidt tot dezelfde uitkomst.

SS_{Total} geeft aan in welke mate de scores variëren door alle oorzaken samen.

Voorbeeld (berekening)

$SS_{Between}$

De volgende groepsgemiddelden waren gevonden:

gemiddelde (non involved)	= 0.725
gemiddelde (bully)	= 0.500
gemiddelde (victim)	= 1.500

Voer nu 20 keer 0.725 in op je rekenmachine, 9 keer 0.500 en 13 keer 1.500. De variantie is dan $(0.4049)^2 = 0.1639$. $SS_{Between}$ wordt dan $(0.1639) * 41 =$ **6.72**.

SS_{Within}

De volgende varianties waren gevonden:

variantie (non involved)	= 0.282
variantie (bully)	= 0.047
variantie (victim)	= 0.958

SS_{Within} is nu gelijk aan $19 * 0.282 + 8 * 0.047 + 12 * 0.958 =$ **17.24**
(Berekend met de niet-afgeronde varianties, anders krijg je 17.23.)

SS_{Total}

Dit is gelijk aan de som van de vorige twee *SS*-en:

$$SS_{Total} = SS_{Between} + SS_{Within}$$
$$= 6.72 + 17.24 = \mathbf{23.96}$$

Ingevuld ziet de summary table er nu uit zoals tabel 1.10.

Tabel 1.10 ANOVA-tabel voor Sociale isolatie (4)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between	2	6.72				
Treiterstatus						
Within	39	17.24				
Total	41	23.96				

Alleen de verhoudingen tussen de sums of squares zijn van belang. De relaties tussen de diverse *SS*-en zijn:

$$SS_{Total} = SS_{Between} + SS_{Within}$$

Voorbeeld (definities)
Hoewel het voor de berekening niet nodig is, zullen we nu de in de 'Toelichting' besproken definities illustreren. De daar besproken principes komen in veel analyses terug. Allereerst bereken je voor elk subject:

voorspelde score = het groepsgemiddelde
residu = geobserveerde score - voorspelde score

Neem het volgende voorbeeld. Subject 1 heeft Treiterstatus 'non involved'. De gemiddelde Sociale isolatie van deze groep is 0.725. Zou Sociale isolatie alleen van iemands Treiterstatus afhangen – en niet van andere persoonlijke factoren – dan zouden we ook voor subject 1 een Sociale isolatie van 0.725 verwachten. Dit is daarom zijn voorspelde score. De werkelijke Sociale-isolatie-score van dit subject is 1.75. Dat is de geobserveerde score. Het residu is voor dit subject dus 1.75 - 0.725 = 1.025. Zo doen we dat voor alle subjecten. Als je dat in de datamatrix schrijft, krijg je de 'uitgebreide datamatrix' van tabel 1.11.

Tabel 1.11 Uitgebreide datamatrix van Sociale isolatie en Treiterstatus

Kind	*Treierstatus*	*Isolatie*	*Voorspelde isolatie*	*Residu isolatie*
1	non-involved	1.75	.725	1.025
2	non-involved	0.25	.725	-.475
3	non-involved	0.25	.725	-.475
4	non-involved	1.25	.725	.525
5	non-involved	0.75	.725	.025
6	non-involved	0.50	.725	-.225
7	non-involved	0.25	.725	-.475
8	non-involved	1.00	.725	.275
9	non-involved	0.75	.725	.025
10	non-involved	1.00	.725	.275

Kind	*Treierstatus*	*Isolatie*	*Voorspelde isolatie*	*Residu isolatie*
11	non-involved	1.25	.725	.525
12	non-involved	0.25	.725	-.475
13	non-involved	1.50	.725	.775
14	non-involved	0.50	.725	-.225
15	non-involved	0.25	.725	-.475
16	non-involved	0.25	.725	-.475
17	non-involved	0.25	.725	-.475
18	non-involved	0.50	.725	-.225
19	non-involved	1.75	.725	1.025
20	non-involved	0.25	.725	-.475
21	bully	0.75	.500	.250
22	bully	0.25	.500	-.250
23	bully	0.75	.500	.250
24	bully	0.50	.500	.000
25	bully	0.50	.500	.000
26	bully	0.50	.500	.000
27	bully	0.75	.500	.250
28	bully	0.25	.500	-.250
29	bully	0.25	.500	-.250
30	victim	1.00	1.500	-.500
31	victim	2.00	1.500	.500
32	victim	0.75	1.500	-.750
33	victim	3.25	1.500	1.750
34	victim	0.50	1.500	-1.000
35	victim	2.25	1.500	.750
36	victim	3.00	1.500	1.500
37	victim	1.00	1.500	-.500
38	victim	1.75	1.500	.250
39	victim	1.75	1.500	.250
40	victim	0.25	1.500	-1.250
41	victim	0.25	1.500	-1.250
42	victim	1.75	1.500	.250

	Isolatie	*Voorspelde isolatie*	*Residu isolatie*
gemiddelde	0.917	0.917	0
variantie	0.5843	0.1639	0.4204
SS	23.96	6.72	17.24
Bron	Total	Between	Within

Vervolgens bereken je voor elk van de drie soorten scores het gemiddelde en de variantie. Deze zijn onder in de tabel opgeschreven. Onder de varianties zijn de *SS*-en geschreven. Deze zijn berekend door de variantie met (N - 1) te vermenigvuldigen. Je ziet, het zijn precies de al eerder berekende *SS*-en.

Mean squares (*MS*)

Voor elk van de twee bronnen 'tussen' en 'binnen' is er een mean square (*MS*). De *MS* van een bron wordt berekend door de *SS* van die bron te delen door de *df* van die bron. De formules voor de mean squares zijn dus:

$$MS_{Between} = SS_{Between} / df_{Between}$$
$$MS_{Within} = SS_{Within} / df_{Within}$$

$MS_{Between}$ meet in hoeverre de groepsgemiddelden van elkaar verschillen. MS_{Within} meet in hoeverre de scores binnen de groepen van elkaar verschillen.

Toelichting

De MS_{Within} is te zien als de **gemiddelde binnengroepsvariantie**. Als binnen elke groep de variantie 2.17 is, dan zal MS_{Within} ook gelijk aan 2.17 zijn. In werkelijkheid hebben de groepen natuurlijk altijd varianties die enigszins van elkaar verschillen. Dan is MS_{Within} het gemiddelde van die varianties. Bijvoorbeeld als er twee even grote groepen zijn met varianties 7 en 8 dan is MS_{Within} 7.5. Als de groepen niet even groot zijn dan is MS_{Within} een gewogen gemiddelde van de varianties. Hoe groter de groep, hoe zwaarder hij meetelt. Aan de formules voor SS_{Within} en df_{Within} is te zien dat elke groep dan wordt gewogen met een factor (n_k - 1).

De $MS_{Between}$ kun je het beste zien als een *herschaling* van de $SS_{Between}$. Hij geeft aan welke waarde je zou moeten beredeneren voor binnengroepsvariantie, als je alleen de groepsgemiddelden zou kennen en als de nulhypothese waar is. Hier is dus iets vreemds aan de hand: $MS_{Between}$ wordt berekend uit de variantie *tussen* groepen, maar hij geeft aan wat de variantie *binnen* groepen zou zijn als de nulhypothese waar is. Deze overgang is gebaseerd op de wortel-N-wet van deel 2B.

Het begrijpen van de betreffende redenering behoort niet tot de leerdoelen van dit boek, maar hij is gebaseerd op een belangrijk statistisch principe en hij wordt in vrijwel elk boek over ANOVA uitgelegd. Daarom zullen we hem ook hier kort weergeven. Ga uit van *k* groepen. Voor het gemak veronderstellen we dat de groepen allemaal even groot zijn, met *n* waarnemingen per groep. Zoals opgemerkt, is het een assumptie van ANOVA dat de scores in elke groep normaal verdeeld zijn met gelijke varianties in elke groep. Noem deze variantie σ^2. Dat is dus de populatiewaarde van de binnengroepsvariantie. Veronderstel dat de nulhypothese waar is. Dan zijn de *k* groepen in feite *k* steekproeven van grootte *n* uit eenzelfde populatie die variantie σ^2 heeft. De groepsgemiddelden zullen nu alleen van elkaar verschillen als gevolg van steekproeftoeval. Volgens de wortel-N-wet zou de standaardafwijking van de groepsgemiddelden dan ongeveer $\sigma / \sqrt{n}$ zijn, en de variantie zou

dus σ^2 / n zijn. Als we daarentegen de variantie van de geobserveerde groepsgemiddelden berekenen, en daarbij elk gemiddelde één keer meetellen, dan komt daar uit: $SS_{Between} / n(k - 1)$. Dit zou dus ongeveer gelijk moeten zijn aan σ^2 / n. Daaruit volgt dat $\sigma^2 \approx SS_{Between} / (k - 1)$. Deze laatste uitdrukking levert dus een schatting voor de binnengroepsvariantie als de nulhypothese waar is.

Kortom, *als* de nulhypothese waar is dan zijn $MS_{Between}$ en MS_{Within} allebei schatters voor dezelfde populatiewaarde van de binnengroepsvariantie. Alleen zijn zij gebaseerd op verschillende, onafhankelijke informatiebronnen. MS_{Within} gebruikt de *varianties* als informatiebron. $MS_{Between}$ gebruikt de *gemiddelden* als informatiebron. Dat laatste is nogal indirect en het klopt alleen als de nulhypothese waar is.

Hoewel dat niet nodig is, kun je op analoge wijze ook MS_{Total} berekenen: $MS_{Total} = SS_{Total} / df_{Total} = 23.96 / 41 = 0.5839$. Dat is de gewone variantie van de ruwe scores.

Alle mean squares zijn te zien als varianties. Dat betekent dat ze kwadratisch afhangen van de schaal waarin de scores zijn uitgedrukt. Stel bijvoorbeeld dat de afhankelijke variabele eerst was gemeten in meters, en dat de MS_{Within} dan 2.17 is. Dan is die 2.17 uitgedrukt in vierkante meters. Als je de afhankelijke variabele uitdrukt in centimeters, dan worden alle scores 100 keer zo groot. De MS_{Within} wordt dan $100^2 = 10\,000$ keer zo groot, dus 21 700.

Voorbeeld
Voor de *MS*-en vinden we:

$$\begin{aligned} MS_{Between} &= 6.72 / 2 &&= 3.36 \\ MS_{Within} &= 17.24 / 39 &&= 0.442 \end{aligned}$$

Als H_0 waar is, zouden deze twee *MS*-en ongeveer even groot moeten zijn. Dat $MS_{Between}$ veel groter is, betekent dat de verschillen tussen de groepsgemiddelden veel groter zijn dan te verwachten valt op grond van MS_{Within} plus toeval. Ingevuld ziet de tabel er nu uit als tabel 1.12.

Tabel 1.12 ANOVA-tabel voor Sociale isolatie (5)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between Treiterstatus	2	6.72	3.36			
Within	39	17.24	0.44			
Total	41	23.96				

_F_-waarde
De *F*-waarde is de toetsingsgrootheid. Het is een maat voor de **hoeveelheid bewijs**

tegen H_0 in de data. Hoe groter *F*, hoe meer bewijs tegen H_0. De *F*-waarde wordt berekend als:

$$F_{Between} = MS_{Between} / MS_{Within}$$

Als de nulhypothese waar is, zal de *F*-waarde naar verwachting **ongeveer 1** zijn, ongeacht de waarde van N. Als de alternatieve hypothese waar is, zal de *F*-waarde naar verwachting **groter dan 1** zijn, en rechtevenredig toenemen met N.

Toelichting

Een *F*-waarde van 1 wil zeggen dat de verschillen tussen de groepsgemiddelden (gemeten met $MS_{Between}$) precies even groot zijn als je zou verwachten op grond van steekproeftoeval wanneer de nulhypothese waar is. Deze verwachting is gebaseerd op de mate van individuele variaties binnen groepen (gemeten met MS_{Within}). Een *F*-waarde van 1 of kleiner zal er dan ook toe leiden dat de nulhypothese wordt behouden.

Een *F*-waarde groter dan 1 wil zeggen dat de groepsgemiddelden meer van elkaar verschillen dan je zou verwachten als de nulhypothese waar is. Bij *F*-waarden groter dan 4 zal de H_0 in de regel worden verworpen.

De *F*-waarde neemt toe naarmate de groepsgemiddelden meer van elkaar verschillen. Er is dan meer bewijs tegen de nulhypothese dat de gemiddelden eigenlijk gelijk zijn. De *F*-waarde neemt af naarmate de groepsvarianties groter worden. Er zit dan meer ruis in de data, waardoor de scores van verschillende groepen elkaar meer overlappen en de verschillen tussen de groepen minder duidelijk zijn. De *F*-waarde hangt niet af van de meeteenheid: als alle scores met 100 worden vermenigvuldigd, zal de *F*-waarde hetzelfde blijven. Bij gelijkblijvende groepsgemiddelden en groepsvarianties zal de *F*-waarde rechtevenredig toenemen met N. Hoe groter de steekproef, hoe meer een gegeven verschil tussen de gemiddelden kan worden gezien als bewijs tegen de nulhypothese.

Voorbeeld

Voor de *F*-waarde vinden we: $F = 3.36 / 0.442 = 7.60$. Je kan dus zeggen dat $MS_{Between}$ 7.60 keer te groot is. Ingevuld ziet de tabel er nu uit als tabel 1.13.

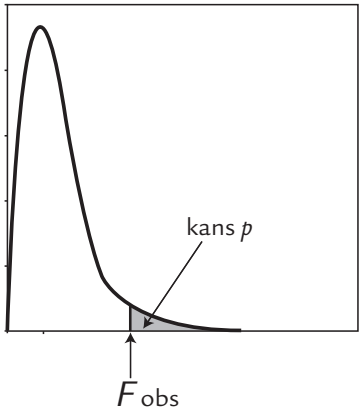
Tabel 1.13 ANOVA-tabel voor Sociale isolatie (6)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between	2	6.72	3.36	7.60		
Treiterstatus						
Within	39	17.24	0.44			
Total	41	23.96				

_p_-waarde

De *p*-waarde is een maat voor de aannemelijkheid of **houdbaarheid** van H_0. De *p*-waarde is gedefinieerd als de kans op een steekproef met een *F*-waarde die even groot of groter is dan de gevonden *F*-waarde indien H_0 waar is (zie figuur 1.2). Bij gelijkblijvende *df*'s geldt: hoe groter *F*, hoe *kleiner p*. De *p*-waarde kun je opzoeken in de *F*-tabel (tabel A.5 in de appendix). Bij het opzoeken moet je voor de 'teller' $df_{Between}$ nemen en voor de 'noemer' df_{Within}, zoals bij de berekening van *F*. Het is gebruikelijk om alleen zo scherp mogelijk te vermelden in welke van de volgende niveaus de *p*-waarde ligt:

$p \geq .1$
$p < .1$
$p < .05$
$p < .01$
$p < .001$



Figuur 1.2 Relatie tussen de geobserveerde F*-waarde en de* p*-waarde*

Toelichting

De kansen in deze tabel zijn berekend op grond van de steekproeventheorie. Zij vormen samen de zogenaamde **referentieverdeling** van *F*. In het voorbeeld met *df* = (2, 39) zou je de *p*-waarde van *F* in principe zonder tabel kunnen bepalen door de volgende stappen uit te voeren met een computer.

1 Creëer drie zeer grote populaties met scores waarbij geldt:
   - De drie populaties hebben hetzelfde gemiddelde (**H_0 is waar**, essentiële aanname).
   - De drie populaties hebben dezelfde variantie (technische aanname).
   - In elk van de drie populaties zijn de scores normaal verdeeld (technische aanname).
2 Trek 1 000 000 steekproeven uit deze populaties. Laat elke steekproef bestaan

uit totaal 42 subjecten, net zoals in het echte onderzoek: 20 subjecten uit populatie 1, 9 subjecten uit populatie 2, en 13 subjecten uit populatie 3.

3 Bereken bij elke steekproef van N = 42 de *F*-waarde. Het histogram van de resulterende 1 000 000 *F*-waarden staat in figuur 1.2. Dit histogram noemt men de *F*(2, 39) verdeling. Het is de steekproevenverdeling van *F* als H_0 waar is.

4 Bepaal bij hoeveel procent van de steekproeven de *F*-waarde groter dan of gelijk is aan de *F*-waarde in het echte onderzoek, 7.60. Dit percentage is de *p*-waarde.

Als je deze procedure volgt, zul je vinden:

$$p = .001632$$

Dit betekent dat bij 1632 van de 1 000 000 steekproeven de *F*-waarde groter dan of gelijk aan 7.60 is. *Dat is dus bijna nooit*. Conclusie: de *F*-waarde die in het echte onderzoek is gevonden (7.60) is veel groter dan je onder H_0 zou verwachten. Aangezien de *F*-waarde bovendien een maat is voor de hoeveelheid bewijs tegen H_0, is dit een reden om H_0 te verwerpen.

De zojuist gevonden exacte *p*-waarde staat overigens niet in de tabel vermeld omdat de tabel dan veel te groot zou worden. Met de tabel kun je slechts de zwakkere uitspraak doen dat:

$$p < .01$$

Dit is óók waar en is nog steeds voldoende reden om H_0 te verwerpen.

Voorbeeld

We vonden een *F*-waarde van 7.60. We weten dat $df_{Between} = 2$ (aantal vrijheidsgraden in de teller) en dat $df_{Within} = 39$ (aantal vrijheidsgraden in de noemer). Nu kunnen we de *p*-waarde vinden in de *F*-tabel (tabel A.5 in de appendix). Deze staat in de cel die door de bovenstaande vrijheidsgraden wordt gespecificeerd. We zien dat bij 'aantal vrijheidsgraden in de noemer' de waarde 39 niet voorkomt. Daarom benaderen we deze door in plaats van 39 de waarde 40 te nemen. We zien in die cel: bij $F = 5.18$ hoort $p = 0.010$ en bij $F = 8.25$ hoort $p = 0.001$. Onze *F* van 7.60 is groter dan 5.18, dus $p < 0.01$. Ingevuld levert dit tabel 1.14 op.

Tabel 1.14 ANOVA-tabel voor Sociale isolatie (7)

Bron	*df*	*SS*	*MS*	*F*	*p*	R^2
Between Treiterstatus	2	6.72	3.36	7.60	< .01	
Within	39	17.24	0.44			
Total	41	23.96				

In dit voorbeeld is de dichtstbijzijnde *df* van de tabel genomen. Je kan wel raden dat er betere manieren zijn. Maar ik vind het niet nuttig daar aandacht aan te besteden, want in de praktijk wordt de *p*-waarde tegenwoordig altijd per computer berekend.

R^2-waarde

De R^2 geeft aan hoe groot de variatie tussen groepen is in verhouding tot de totale variatie. De formule is

$$R^2 = SS_{Between} / SS_{Total}$$

Dit is een maat voor de **sterkte van het effect** van de factor op de meting. Je kan R^2 ook zien als een maat voor de **sterkte van de samenhang** tussen de factor en meting. Denk daarbij aan het spreidingsdiagram. R^2 wordt ook vaak de **proportie verklaarde variantie** genoemd.

Een R^2 van 1 wil zeggen dat de scores geheel worden bepaald door de factor: de groepsgemiddelden verschillen en subjecten uit eenzelfde groep hebben altijd dezelfde score. Een R^2 van 0 wil zeggen dat de factor helemaal geen invloed heeft: de groepsgemiddelden zijn gelijk maar binnen een groep bestaan wel individuele variaties in de scores. Een R^2 van .80 wil zeggen dat de scores voor 80% kunnen worden verklaard door de factor en voor 20% door individuele verschillen tussen subjecten. De verschillen tussen groepen zijn dan veel groter dan de verschillen binnen groepen.

Toelichting

Een oude mythe is dat een 'significant' verschil tussen gemiddelden betekent dat het verschil behoorlijk groot of belangrijk is. Als je dat ook denkt, is het je waarschijnlijk aangepraat door autoriteiten uit je jeugd, die probeerden te verbloemen dat ze niets van statistiek wisten en daarom het woord letterlijk vertaalden. De naïeve reflex bij deze mythe is:

'$p < .001 \rightarrow$ zeer significant $\rightarrow$ Hoera!'

Zet deze gedachte onmiddellijk uit je hoofd, het is onzin! De *p*-waarde zegt alleen iets over de *zekerheid* van de conclusie. Dat hangt in belangrijke mate af van de *grootte van de steekproef*. Zelfs als de populatiegemiddelden maar een heel klein beetje van elkaar verschillen, zal bij een grote steekproef automatisch een kleine *p*-waarde verschijnen. Als bijvoorbeeld de gemiddelde lengte van vrouwen 175 cm is en die van mannen 175.0000001 cm, dan is dat verschil naar alle redelijke maatstaven klein en onbelangrijk. Toch zal men vinden dat $p < .001$ als de steekproef maar groot genoeg is. Een kleine *p*-waarde betekent *niet* dat de verschillen tussen de gemiddelden groot zijn; **een kleine *p*-waarde betekent slechts dat de steekproef groot genoeg is** om er erg zeker van te zijn dat de populatiegemiddelden niet *exact* gelijk aan elkaar zijn.