

Hoofdstuk 12: Eenweg ANOVA

12.1 Eenweg analyse van variantie

Eenweg en tweeweg ANOVA

Wanneer we verschillende populaties of behandelingen met elkaar vergelijken, dan zal er binnen de data altijd sprake zijn van variabiliteit. Omdat dit normaal is, richten we ons voor het vergelijken van populaties of behandelingen op *gemiddelden*. We gebruiken ANOVA (*analyse van variantie*) om verschillende gemiddelden met elkaar te vergelijken. Er bestaan twee ANOVA-technieken:

- *Eenweg ANOVA*: deze techniek wordt gebruikt als er maar één manier is om de populaties te classificeren. Een voorbeeld is onderzoeken of overlevingskansen verschillen voor drie verschillende longkankerbehandelingen.
- *Tweeweg ANOVA*: in dit geval is er meer dan één manier om populaties te classificeren. Je kunt bijvoorbeeld vergelijken in hoeverre drie temperaturen (0, 20 en 30 graden) *gecombineerd* met drie hoeveelheden van licht (geen, gemiddeld en fel licht) invloed hebben op het onthouden van woordjes die een onderzoeker opleest. In het volgende hoofdstuk zal dieper ingegaan worden op tweeweg ANOVA.

Data voor eenweg ANOVA

Met eenweg ANOVA vergelijken we dus verschillende populatiegemiddelden. We trekken daarvoor een random steekproef (SRS) uit elke populatie en we gebruiken deze data om de nulhypothese te toetsen. We kunnen eenweg ANOVA ook gebruiken bij gerandomiseerde experimenten. De nulhypothese stelt dat alle populatiegemiddelden hetzelfde zijn. In ons voorbeeld stelt de nulhypothese dus dat alle drie de longkankerbehandelingen gepaard gaan met dezelfde overlevingskans.

Gemiddelden vergelijken

We gebruiken de term 'groepen' als we het hebben over de verschillende populaties waarvan we willen onderzoeken of ze hetzelfde gemiddelde hebben. Om uitspraken over het gemiddelde van de populaties te doen, maken we gebruik van steekproefgemiddelden. Als blijkt dat er een verschil in gemiddelden bestaat, dan vragen we ons af of dat door toeval of door een echt effect komt. Door middel van ANOVA kunnen we uitzoeken of geobserveerde verschillen tussen steekproefgemiddelden *statistisch significant* zijn. We kunnen hier uitspraken over doen door te kijken naar de (1) steekproefgroottes, (2) de spreiding binnen de groepen en (3) de steekproefgemiddelden.

De twee-sample t-toets

De twee-sample t-toets vergelijkt de gemiddelden van twee populaties. De aanname is hierbij dat beide populaties dezelfde (maar een onbekende) standaarddeviatie hebben en dat de steekproefgroottes gelijk zijn. Hoe vinden we deze t-toets?

- Trek eerst het gemiddelde van y van het gemiddelde van x af.
- Vermenigvuldig de gepoolde standaarddeviatie (s_p) met de wortel uit $1/n+1/n$.
- Deel tot slot de uitkomst uit de eerste stap door de uitkomst van de tweede stap.

Als we de gevonden t-toets (t^2) kwadrateren geeft dit precies hetzelfde resultaat als de ANOVA F-toets voor twee populaties. Een formule om t^2 direct uit te rekenen is:

- $t^2 = n/2(x\text{-gemiddeld} - y\text{-gemiddeld})^2/s_p^2$.

De teller in de bovenstaande formule meet de spreiding *tussen* de groepen. De teller kan groot zijn door een groot verschil tussen de steekproefgemiddelden of omdat de steekproeven groot zijn. De noemer meet de spreiding *binnen* de groepen. Als deze erg klein is, dan zorgt dit voor een grote t^2 -waarde. Dit maakt de kans op een significant resultaat groter.

Hypothesen

Eenweg ANOVA gaat uit van de nulhypothese die stelt dat *alle* populatiegemiddelden gelijk zijn. De alternatieve hypothese stelt dat niet alle populatiegemiddelden gelijk zijn. De alternatieve hypothese klopt dus al als één populatiegemiddelde afwijkt van de rest. De alternatieve hypothese wordt echter ook aangenomen als alle populatiegemiddelden van elkaar verschillen.

- Als we de nulhypothese hebben afgewezen, moeten we daarom nog uitzoeken waar de verschillen tussen populatiegemiddelden precies liggen. Dit kunnen we doen aan de hand van *contrasten*. Als we drie populaties (1,2,3) onderzoeken, kunnen we contrasten maken waarin we stellen dat ze allemaal verschillend zijn of dat populatie 1 van populatie 2 en populatie 3 verschilt. We kunnen ook een contrast maken waarin we stellen dat populatie 2 van populatie 1 en populatie 3 verschilt. Zo zijn er verschillende contrasten mogelijk. De contrasten moeten geformuleerd worden *voordat* het onderzoek uitgevoerd wordt.
- Als we geen vermoeden hebben over de specifieke relatie tussen de populatiegemiddelden, dan kunnen we gebruik maken van *multipele vergelijkingen* (*multiple-comparisons*). We onderzoeken in dat geval welke *paren* van populatiegemiddelden significant van elkaar verschillen.

Het eenweg ANOVA-model

Ook dit model gaat uit van een situatie waarbij een deel van de data bij het model past en een deel onverklaard blijft (data= fit+residuen).

- Het *eenweg ANOVA-model* is: $x_{ij} = \mu_i + \epsilon_{ij}$. In deze formule staat i voor $1, \dots, I$. Bij eenweg ANOVA staat I voor het aantal populaties. Daarnaast staat j voor $1, \dots, n_i$. In dit verband staat n_i voor de steekproef van de i -ste populatie. Tot slot staat x_{ij} voor de j -ste observatie van de i -ste populatie. De I populatiegemiddelden zijn het fit-gedeelte van het model en worden uitgedrukt in μ_i . De overige spreiding (ϵ_{ij}) staat gelijk aan het residu-gedeelte. Dit zijn de afwijkingen van de populatiegemiddelden.
- ϵ_{ij} komen uit een normaalverdeelde distributie met een onbekende standaarddeviatie en een gemiddelde van 0: $N(0, \sigma)$.
- De (onbekende) *parameters van het model* zijn de populatiegemiddelden ($\mu_1, \mu_2, \dots, \mu_I$) en de standaarddeviatie (σ) waarvan aangenomen wordt dat deze voor alle populaties gelijk is. Er wordt dus vanuit gegaan dat steekproefgroottes verschillend kunnen zijn, terwijl alle populaties dezelfde standaarddeviatie hebben.

Het schatten van populatieparameters

ANOVA is niet erg gevoelig voor ongelijke standaarddeviaties tussen de groepen. Als de grootste standaarddeviatie minder dan twee keer de kleinste standaarddeviatie is, dan zal gebruik van ANOVA leiden tot (bijna) correcte resultaten. Wanneer we aannemen dat de populatiestandaarddeviaties gelijk zijn, dan is elke steekproefstandaarddeviatie een schatter van σ . We voegen deze steekproefstandaarddeviaties samen om tot een gepoolde schatter van σ te komen:

- De *gepoolde steekproefvariantie* is: $s_p^2 = (n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 + \dots + (n_l - 1) s_l^2 / ((n_1 - 1) + (n_2 - 1) + \dots + (n_l - 1))$.
- Vervolgens moet de wortel uit de uitkomst getrokken worden om tot de gepoolde standaarddeviatie (s_p) te komen. Dit is een schatter van σ . Als de steekproefgroottes gelijk zijn, dan is s_p^2 het gemiddelde van de steekproefvarianties van alle groepen. Let op: s_p is niet het gemiddelde van de steekproefstandaarddeviaties.

Binnen-groepen variantie en tussen-groepen variantie

De verschillende populatiegemiddelden worden vergeleken en getoetst aan de hand van de F-toets. Hierbij wordt gekeken naar de spreiding binnen en tussen de groepen. We willen dat de tussen-groepenvariantie groot is en dat de binnen-groepenvariantie klein is. Dat maakt de kans op het vinden van significante resultaten groter.

- De *nulhypothese* is dat alle groepsgemiddelden gelijk zijn: $\mu_1 = \mu_2 = \dots = \mu_l$. I staat zoals eerder gezegd voor het aantal populaties. De *alternatieve hypothese* zegt dat niet alle gemiddelden (μ_i) gelijk aan elkaar zijn. We maakten tot nu toe vooral gebruik van het model $\text{data} = \text{fit} + \text{residuen}$. Nu kunnen we dat vertalen naar: $\text{totaal} = \text{tussen-groepen} + \text{binnen-groepen}$.

SS, DF en MS voor eenweg ANOVA

SS (*sums of squares*) laat zien hoeveel spreiding er in de data aanwezig is. De verschillende SS-elementen worden berekend door de gekwadrateerde afwijkingen op te tellen. Bij eenweg ANOVA zijn er *drie bronnen van spreiding*: (1) groepen, (2) error en (3) totaal. We zeggen daarom ook wel:

- $SST = SSG + SSE$. De totale spreiding wordt dus opgedeeld in tussen- en binnen-groepenvariantie. De bijbehorende *vrijheidsgraden* kunnen ook opgedeeld worden in (1) groepen, (2) error en (3) totaal. We zeggen daarom dat DFT bestaat uit DFG en DFE. Om *MS (mean square)* uit te rekenen, moet een specifieke SS gedeeld worden door de bijbehorende vrijheidsgraden. We kunnen de MS voor de error (MSE) vinden door de gepoolde standaarddeviatie (s_p) te kwadrateren (s_p^2). Dus: $s_p^2 = MSE = SSE / DFE$. Om de gepoolde standaarddeviatie te vinden moet de wortel uit MSE getrokken worden.

De F-toets

Als de nulhypothese waar is, zijn er geen verschillen tussen de populatiegemiddelden.

MSG/MSE is ongeveer 1 als de nulhypothese waar is. De waarde wordt groter dan 1 als de alternatieve hypothese klopt. We maken voor het toetsen van de nulhypothese gebruik van:

- Het aantal vrijheidsgraden uit de teller. Voor eenweg ANOVA zijn de vrijheidsgraden uit de teller $I-1$. Het aantal vrijheidsgraden uit de noemer: $N-I$.
- Samengevat geeft dit: $F(I-1, N-I)$. De F-toets voor ANOVA is behoorlijk robuust als het gaat om non-normaliteit en ongelijke varianties binnen de groepen.

De ANOVA-tabel voor eenweg ANOVA

BRON (source)	Vrijheidsgraden (DF)	SS (sum of squares)	MS (mean square)	F
Groepen	I-1 (aantal groepen-1)	$\sum_{\text{groepen}} n_i(x\text{-gemiddeld}_i - x\text{-gemiddeld})^2$	SSG/DFG	MSG/MSE
Error	N-I	$\sum_{\text{groepen}} (n_i - 1)s_i^2$	SSE/DFE	
Totaal	N-1	$\sum_{\text{obs}} (x_{ij} - x\text{-gemiddeld})^2$		

Tot slot kan bij eenweg ANOVA de *coefficient van bepaling (coefficient of determination)* uitgerekend worden: $R^2 = \text{SSG}/\text{SST}$.

12.2 Gemiddelden vergelijken

Verschillen tussen groepsgemiddelden

De ANOVA F-toets geeft ons alleen antwoord op de vraag of de gevonden verschillen tussen groepsgemiddelden significant zijn. Een kleine p-waarde zegt ons dat de groepsgemiddelden niet allemaal hetzelfde zijn. We weten dan echter nog niet welke groepsgemiddelden van elkaar verschillen. Wanneer de nulhypothese bij eenweg ANOVA afgewezen is, moeten er aanvullende analyses uitgevoerd worden om te kijken waar de verschillen precies liggen. Middels contrasten kunnen we de groepen met elkaar vergelijken. We kunnen dan bijvoorbeeld de eerste twee populaties vergelijken met de derde populatie. We kunnen er ook voor kiezen om de eerste populatie te vergelijken met de tweede en de derde populatie. We kunnen dus verschillende alternatieve hypothesen formuleren in de vorm van contrasten.

- Om een contrast over de populatie te toetsen, maken we gebruik van een *steekproefcontrast*. We kijken dan naar steekproefgemiddelden in plaats van populatiegemiddelden.

Contrasten

- Een *contrast* is een combinatie van populatiegemiddelden in de vorm van $\Psi = \sum a_i \mu_i$. De coëfficiënten van a_i tellen op tot 0.
- Het corresponderende *steekproefcontrast* is: $c = \sum a_i x\text{-gemiddeld}_i$.
- De *standaardfout van c* is: $SE_c = s_p \sqrt{\sum a_i^2 / n_i}$.
- We toetsen de nulhypothese $\Psi = 0$. We gebruiken hiervoor de *t-toets*: $t = c / SE_c$. Ook maken we gebruik van de vrijheidsgraden voor de error (DFE) die gepaard gaan met s_p . De alternatieve hypothese kan zowel eenzijdig als tweezijdig zijn.
- Het *betrouwbaarheidsinterval* voor Ψ is $c \pm t^* SE_c$.

Multipelen vergelijkingen (multiple comparisons)

Multipelen vergelijkingen worden uitgevoerd *nadat* de nulhypothese voor eenweg ANOVA verworpen is. Aan de hand van deze vergelijkingen worden steeds paren van populaties met elkaar vergeleken.

- Om multipelen vergelijkingen te toetsen, berekenen we *t-toetsen*:
- $t_{ij} = x\text{-gemiddeld}_i - x\text{-gemiddeld}_j / (s_p \sqrt{1/n_i + 1/n_j})$. Als de uitkomst van t_{ij} groter of gelijk aan t^{**} is, dan mogen we concluderen dat de populatiegemiddelden uit één paar verschillend zijn. Als dat niet het geval is, dan zijn de populatiegemiddelden gelijk aan elkaar. De waarde van t^{**} hangt af van de statistische meettechniek die we gebruiken.

Het bepalen van t^{}**

We kunnen, om t^{**} te bepalen, bijvoorbeeld kiezen voor de '*least-significant differences method*' (*LSD*) waarbij gebruik gemaakt wordt van een alfa van 5%. Dit kan gevaarlijk zijn; vooral als er veel populaties onderzocht worden. Dit komt doordat de kans op een type-I-fout dan toeneemt. In dat geval wordt de nulhypothese verworpen, terwijl deze in werkelijkheid wel klopt. Als onderzoeker neem je dan aan dat er een effect bestaat, terwijl dit niet het geval is. Om t^{**} te bepalen, kunnen we ook kiezen voor de *Bonferroni methode*. Met deze methode neemt de kans op een type-I-fout niet toe per vergelijking. De kans blijft altijd 5%.