

## 16. MANOVA

### MANOVA

*Multivariate variantieanalyse (MANOVA)* kan gebruikt worden in een situatie waarin je meerdere afhankelijke variabelen hebt. Met MANOVA kan er 1 onafhankelijke variabele gebruikt worden maar ook meer. Er kan naar interacties en contrasten gekeken worden met MANOVA. Omdat met ANOVA niet meer dan 1 afhankelijke variabele gebruikt kan worden, wordt het een *univariate* toets genoemd.

### Overeenkomsten en verschillen met ANOVA

Met meerdere afhankelijke variabelen zouden we ook meerdere ANOVA's uit kunnen voeren. Een nadeel hiervan is dat hoe meer ANOVA's je uitvoert, hoe groter de familywise meefout wordt en dus hoe groter de kans op type I fouten is. Een ander nadeel van meerdere ANOVA's is dat er niet gekeken wordt naar de relatie van de afhankelijke variabelen onderling.

Pas wel op met de afhankelijke variabelen in MANOVA. Niet elke variabele heeft een goede reden om in de analyse te zitten. Statistisch gezien kan dat soms wel een goede uitkomst kunnen geven, maar empirisch niet.

### De theorie

#### *Matrices*

Een *matrix* is een verzameling van cijfers die geordend zijn in rijen en kolommen. Een 2X3 matrix heeft bijvoorbeeld twee rijen en drie kolommen. In SPSS is de data editor bijvoorbeeld een matrix, waarbij de rijen de deelnemers zijn en de kolommen als variabelen gezien worden. De waarden in een matrix worden *componenten* of *elementen* genoemd. De rijen en kolommen heten *vectoren*.

Een *square matrix* heeft evenveel rijen als kolommen en ziet er dus als een vierkant uit. De getallen die op de diagonaal van de square matrix liggen zijn diagonale componenten en de getallen die niet op de diagonaal liggen zijn niet-diagonale componenten. Als de diagonale componenten allemaal 1 zijn en de niet-diagonale componenten allemaal 0 wordt de matrix een *identiteitsmatrix* genoemd.

#### *Belangrijke matrices*

Net als bij ANOVA zijn we geïnteresseerd in de variantie die verklaard wordt door de experimentele manipulatie, vergeleken met de variantie die daar niet door verklaard wordt. Bij MANOVA wordt de systematische variantie vergeleken met de niet-systematische variantie van meerdere afhankelijke variabelen. De systematische en niet-systematische variantie is bij MANOVA niet een enkel getal, maar wordt in een matrix met veel varianties en covarianties weergegeven.

De matrix dat de systematische variantie laat zien heet de *hypothese kwadratensom en kruisproducten matrix (hypothese SSCP)*. Deze matrix wordt aangeduid met de letter *H*. De matrix dat de niet-systematische variantie laat zien wordt de *meefout kwadratensom en kruisproducten matrix (Meefout SSCP)* genoemd. Deze matrix wordt aangeduid met de letter *E*. Er is ook nog een matrix die de totale variantie voor elke afhankelijke variabele weergeeft. Deze matrix heet de *totale kwadratensom en kruisproducten matrix (totaal SSCP)* en wordt aangeduid met de letter *T*.

De kwadratensom in de matrices wordt op dezelfde manier gebruikt als bij ANOVA. Met de kruisproducten wordt de totale waarde van de meetfouten tussen twee variabelen bedoeld. Hierdoor kan MANOVA de correlatie tussen twee variabelen weergeven en verklaren.

#### *Het berekenen van MANOVA met de hand*

Voor extra duidelijkheid worden eerst twee ANOVA's uitgevoerd. Hiervoor moeten eerst de  $SS_T$ , de  $SS_M$  en de  $SS_R$  berekend worden. Daarna moeten de gemiddelde kwadratensommen berekend worden door de kwadratensommen door het aantal vrijheidsgraden te delen. Met  $MS_M$  en  $MS_R$  kan dan de F berekend worden. Deze waarde kan dan vergeleken worden met kritieke waarden. Voor beide afhankelijke variabelen kan de ANOVA uitgevoerd worden.

De MANOVA gebruikt dezelfde kwadratensommen als ANOVA, alleen wordt er nu ook rekening gehouden met de relatie tussen de afhankelijke variabelen door middel van kruisproducten. De relatie tussen de twee afhankelijke variabelen gaat met kruisenproducten. Als eerste heb je het totale kruisproduct ( $CP_T$ ).

$$CP_T = \sum (x(\text{afhankelijke variabele 1}) - \bar{X}_{\text{groot(afhankelijke variabele 1)}})(x(\text{afhankelijke variabele 2}) - \bar{X}_{\text{groot(afhankelijke variabele 2)}})$$

De  $CP_T$  is een maat voor de totale relatie tussen twee variabelen. Het model kruisproduct ( $CP_M$ ) kijkt naar hoe de relatie tussen twee afhankelijke variabelen is beïnvloed door de experimentele manipulatie.

$$CP_M = \sum n(\bar{x}_{\text{groep(afh. var. 1)}} - \bar{X}_{\text{groot(afh. var. 1)}})(\bar{x}_{\text{groep(afh. var. 2)}} - \bar{X}_{\text{groot(afh. var. 2)}})$$

Als laatste is het residu kruisproduct ( $CP_R$ ). Dit kruisproduct kijkt naar hoe de relatie van de twee afhankelijke variabelen beïnvloed is door individuele verschillen.

$$CP_R = \sum (x(\text{afhankelijke variabele 1}) - \bar{X}_{\text{groep(afhankelijke variabele 1)}})(x(\text{afhankelijke variabele 2}) - \bar{X}_{\text{groep(afhankelijke variabele 2)}})$$

Het residu kruisproduct kan ook berekend worden met:

$$CP_R = CP_T - CP_M$$

Bij een analyse met twee afhankelijke variabelen zullen alle matrices 2X2 zijn. De totale SSCP matrix bevat de totale kwadratensom voor elke afhankelijke variabele en het totale kruisproduct tussen de twee afhankelijke variabelen. De eerste rij en kolom bevatten de eerste afhankelijke variabele en de tweede rij en kolom bevatten de tweede afhankelijke variabele.

De residu SSCP matrix bevat de residu kwadratensom voor elke afhankelijke variabele en het residu kruisproduct tussen de twee afhankelijke variabelen. Deze matrix ziet er hetzelfde uit als de totale SSCP matrix alleen gaat het hier om de meetfouten.

De model SSCP matrix bevat de model kwadratensom voor elke afhankelijke variabele en het model kruisproduct tussen de twee afhankelijke variabelen. Matrices kunnen bij elkaar opgeteld worden. Je kan dus twee matrices samenvoegen door de componenten bij elkaar op te tellen.

$$T = H + E$$

*De MANOVA teststatistiek*

Bij een ANOVA delen we de systematische variantie door de niet-systematische variantie. Het probleem bij MANOVA is dat matrices niet deelbaar zijn door andere matrices. Op een omgekeerde manier kan het wel gedeeld worden. Als we H door E willen delen, dan moet H vermenigvuldigt worden met het omgekeerde van E ( $E^{-1}$ ). Er ontstaat dan een nieuwe matrix dat  $HE^{-1}$  wordt genoemd.

Het omgekeerde van E uitrekenen is erg moeilijk en de berekening kan door SPSS worden gedaan.  $HE^{-1}$  kan gezien worden als de F-ratio in de gewone ANOVA. Een moeilijkheid hierbij is dat een F-waarde gewoon een getal is, terwijl de  $HE^{-1}$  een matrix met meerdere getallen is. De uiteindelijke matrix bevat zoveel getallen als het kwadraat van het aantal afhankelijke variabelen.

## Discriminante functie variaten

Bij een matrix blijven er altijd meerdere waarden uitkomen. Dit kan opgelost worden door ze in onderliggende dimensies (factoren) op te delen. Het is mogelijk een onderliggende lineaire dimensie te vinden bij de afhankelijke variabelen. Deze lineaire combinatie van afhankelijke variabelen wordt een *variaat (variate)* genoemd. Als ze de functie hebben om groepen te onderscheiden worden het *discriminante functie variaten* genoemd.

De eerste discriminante functie ( $V_1$ ) is de combinatie van twee afhankelijke variabelen met het grootste verschil tussen de groepen. Bij de eerste discriminante functie proberen we het hoogst mogelijke getal te berekenen. Met een vergelijking is de discriminante functie als volgt:

$$Y = b_0 + b_1X_1 + b_2X_2$$

$$V_1 = b_0 + b_1DV_1 + b_2DV_2$$

DV = dependent variable

De b-waardes zijn de gewichten van de variabelen die iets zeggen over de toegevoegde waarde van de afhankelijke variabele aan de variaat. Met discriminant functie analyse zijn de b-waardes te verkrijgen door de eigenvectors van de  $HE^{-1}$  matrix. Wanneer er twee afhankelijke variabelen en twee groepen in de onafhankelijke variabele zitten is er maar 1 variaat. Het aantal variaten is altijd minder dan het aantal afhankelijke variabelen. De  $b_0$  kan genegeerd worden, omdat die alleen de plaats van de variaat aangeeft, wat onnodig is als je het gebruikt om groepen te onderscheiden.

De variaten van de gegevens zijn orthogonaal, wat betekent dat ze ongecorrleerd zijn. Door deze variaten te gebruiken, reduceer je het aantal gegevens in de matrix. In het voorbeeld blijven er nog maar twee getallen over, in plaats van vier. Deze waardes in de diagonaal van de matrix van de variaten zijn de eigenwaardes van de originele  $HE^{-1}$  matrix. Deze eigenwaardes zijn het equivalent van de F-ratio in een ANOVA. Tenslotte moeten deze waardes worden vergeleken met wat we zouden verwachten als er geen effect is in de populatie. Dit kan op vier verschillende manieren.

Ten eerste is er de *Pillai-Bartlett trace* ( $V$ ):

$$V = \sum \frac{\lambda_i}{1 + \lambda_i}$$

Hierbij is  $\lambda_i$  de eigenwaarde van elke discriminant variaat. Pillai's trace is de som van de proportie verklaarde variantie op de discriminante functies en is dus vergelijkbaar met  $R^2$ .

Een tweede manier om naar het effect te kijken is met *Hotelling-Lawley trace* (*Hotelling's  $T^2$* ). Het is de som van de eigenwaarden van de variaten.

$$T = \sum \lambda_i$$

Een derde manier is met *Wilks's lambda* ( $\Lambda$ ). Het is het product van de onverklaarde variantie van elke variaat.

$$\Lambda = \prod \frac{1}{1 + \lambda}$$

Het teken vooraan in de formule is vergelijkbaar met het somteken, alleen moet hier alles met elkaar vermenigvuldigd worden in plaats van opgeteld. Grote eigenwaardes (die een groot experimenteel effect aangeven) zorgen voor een kleine Wilks's lambda, dus significantie wordt bereikt als Wilks's lambda klein is.

Een vierde manier is *Roy's largest root*. Het is de eigenwaarde voor het eerste variaat.

$$\Theta = \lambda_{\text{grootst}}$$

Deze statistiek laat de verklaarde variantie van het eerste variaat zien. Deze waarde is conceptueel vergelijkbaar met de F-ratio in een ANOVA. Deze statistiek heeft de meeste power.

## Praktische zaken

### *Assumpties*

- **Onafhankelijkheid:** De residuen moeten statistisch onafhankelijk van elkaar zijn.
- **Random steekproef:** De gegevens moeten gemeten worden op intervalniveau en random uit de populatie komen.
- **Multivariate normaliteit:** Bij MANOVA hebben de afhankelijke variabelen multivariate normaliteit binnen de groepen.
- **Homogeniteit van de covariantie matrices:** Voor elke afhankelijke variabele moeten de varianties gelijk zijn en de correlatie tussen elk van de afhankelijke variabelen moet hetzelfde zijn in alle groepen. Deze assumptie kan getest worden door te kijken of alle populatie *variantie-covariantie matrices* van de verschillende groepen gelijk zijn.

De assumpties zijn op dezelfde manier te testen als bij univariate toetsen. Multivariate normaliteit kan niet getest worden in SPSS en de oplossing hiervoor is voor elke afhankelijke variabele de normaliteit te testen. Hierdoor kan je zeggen dat er univariate normaliteit is, maar het geeft geen garantie voor multivariate normaliteit.

De laatste assumptie kan wel bekeken worden. De variantie-covariantie matrices kunnen vergeleken worden met de *Box's test*. Wanneer de matrices gelijk zijn, is deze test niet significant. Box's test kan ook niet significant zijn als de assumptie van multivariate normaliteit geschonden wordt. Wanneer de groeps-groottes gelijk zijn kan je de Box's test negeren omdat het onstabiel is en we in deze situatie aan kunnen nemen dat de Hotelling's en Pillai's statistieken robuust zijn. Bij ongelijke groeps-groottes kan de robuustheid niet aangenomen worden.

Er is geen niet-parametrische vervanger voor MANOVA.

### *Een teststatistiek kiezen*

Wanneer de groepsverschillen op het eerste variaat zitten kan het beste Roy's statistiek gebruikt worden omdat die de meeste power heeft. Wanneer groepen verschillen op meer dan 1 variaat is Pillai's trace het meest krachtig en Roy's het minst. Alle vier de statistieken zijn redelijk robuust tegen de assumptie van multivariate normaliteit. Roy's root is niet robuust tegen de assumptie van homogeniteit. Pillai-Bartlett trace is het meest robuust voor schending van de assumpties bij gelijke groeps-grootte. Bij ongelijke groeps-grootte is Pillai's trace accuraat als aan de assumpties van multivariate normaliteit en homogeniteit van covariantie matrices is voldaan.

Als er een significante MANOVA uitkomt kan je vervolgens op alle afhankelijke variabelen een ANOVA toepassen om te kijken waar het verschil zit. Hierbij moet wel de Bonferroni correctie worden toegepast, in verband met een inflatie van de kans op een type I fout. Deze aanpak heeft nog een nadeel, namelijk dat er geen rekening meer gehouden wordt met de relatie tussen de afhankelijke variabelen. De discriminant analyse kan als alternatief gebruikt worden omdat die de groepen van elkaar scheidt.

---

Deze analyse vindt lineaire combinaties van afhankelijke variabelen die de groepen het beste van elkaar onderscheidt.

### SPSS

Voor MANOVA in SPSS ga je naar Analyze – General Linear Model – Multivariate (zie blz. 645). De afhankelijke variabelen kunnen in Dependent Variables worden ingevoerd en de onafhankelijke variabele kan worden geplaatst bij Fixed Factor. Je kunt ook covariaten meenemen in de analyse.

Voor de onafhankelijke variabelen kunnen contrasten en post hoc testen uitgevoerd worden. Bij opties kunnen nog extra de SSCP matrices en de residual SSCP matrices aangevinkt worden zodat die ook in de output komen. Bij de residual SSCP matrices verschijnen de meetfout variantie-covariantie matrix en de matrix die gebaseerd is op *Bartlett's test of sphericity*. Deze test kijkt of de variantie-covariantie matrix proportioneel is aan een identiteitsmatrix, wat betekent dat de covarianties nul zijn en de varianties (de waardes op de diagonaal) ongeveer gelijk zijn.

## Output

Wanneer bij opties descriptives aangevinkt zijn, komt als eerst voren in de output een tabel met beschrijvende statistieken. Het bevat de groepsgemiddelden en standaarddeviaties voor de afhankelijke variabelen. Het volgende deel is de Box's test voor de assumptie van gelijke covariantie matrices. Als de matrices gelijk zijn, moet de statistiek niet significant zijn. Bartlett's test van sfericiteit test of aan de assumptie van sfericiteit is voldaan, in het geval van univariate herhaalde metingen (MANOVA heeft deze assumptie niet).

De tabel Multivariate Tests is de hoofdtabel van de resultaten van MANOVA. Deze tabel bevat de vier teststatistieken die eerder besproken zijn. Als de teststatistiek significant is, is er een verschil tussen bepaalde groepen, maar je weet nog niet tussen welke groepen het verschil is. SPSS voert univariate testen uit om dit verschil te bekijken.

Eerst wordt een tabel met de Levene's test weergegeven om te kijken of er gelijke varianties zijn bij de afhankelijke variabelen. Om aan de assumptie te voldoen moet Levene's test niet significant zijn. Daarna volgt een samenvatting van de ANOVA tabel. Deze tabel geeft dezelfde waardes als wanneer losse ANOVA's waren uitgevoerd. Er is dus geen aanpassing om inflatie van de kans op een type I fout te voorkomen.

In het voorbeeld in het boek zijn geen van de univariate testen significant, ondanks dat de MANOVA wel significant was. De reden hiervoor is dat MANOVA meer power heeft omdat het rekening houdt met de onderlinge correlaties tussen de afhankelijke variabelen. Er zal dus een discriminante functie analyse uitgevoerd moeten worden.

Als je de optie hebt geselecteerd om de SSCP matrices weer te geven, komen er twee tabellen in de output zoals op bladzijde 651.

Het rapporteren van MANOVA is hetzelfde als bij ANOVA. Vermeld de teststatistiek, de F-waarde, het aantal vrijheidsgraden en de significantie.

## Discriminant analyse

Bij een discriminant analyse kijk je hoe je het best groepen kunnen onderscheiden met lineaire combinaties van afhankelijke variabelen. Voor discriminant analyse in SPSS ga je naar Analyze – Classify – Discriminant.

De onafhankelijke variabele komt bij Grouping variable waarbij je ook de range moet geven (de hoogste en de laagste code) en de voorspellers gaan bij Independents. Bij Statistics kunnen verschillende matrices aangevinkt worden en kunnen de ongestandaardiseerde coëfficiënten worden geselecteerd.

Bij Classify kan je aangeven of de groeps grootte gelijk is of niet en het gebruik van de covariantie matrix (zie blz. 656). Het is handig om ook een grafiek te plotten, met weinig groepen kun je het beste een gecombineerde groepsplot kiezen. De laatste opties zijn bij save te vinden. De *discriminant scores* geven de scores van elke persoon op elke variat, wat handig is voor de analyse.

## Output

De covariantie matrices zijn handig omdat ze inzicht geven in hoe de relatie tussen twee afhankelijke variabelen verandert van groep tot groep. Op de diagonaal zie je de covariantie. Het is niet gestandaardiseerd, maar toch geeft het een indicatie van de relatie tussen de afhankelijke variabelen.

Na de tabel met de covariantie matrices komen de statistieken van de discriminant analyse. Hier worden de eigenwaardes waargegeven, samen met het percentage verklaarde variantie. De volgende tabel laat de significantietest zien. De eerste rij laat de significantie zien van beide variaten, de tweede rij de significantie wanneer alleen het tweede variat nog in de analyse zit. Bij twee variaten zijn er slechts twee stappen.

De tabellen Standardized Canonical Discriminant Function Coefficients en Structure matrix zijn het belangrijkste voor de interpretatie. Het laat gestandaardiseerde eigenvectors en waardes die vergelijkbaar zijn met factorladingen zien. De variaten met hoge canonical correlaties zorgen voor het grootste verschil tussen de groepen. Ze laten de relatieve contributie van elke afhankelijke variabele aan het variat zien.

Na de twee tabellen worden twee andere tabellen weergegeven die de coëfficiënten op de ongestandaardiseerde manier laten zien. Dit zijn de b-waardes in de vergelijking. Bij Functions at Group Centroids worden de gemiddelde scores van de variaten voor elke groep weergegeven. Om dit te interpreteren, moet je kijken of het een positieve of negatieve waarde heeft.

In de grafiek met gecombineerde groepen worden de variat scores voor elke persoon geplotted, kijkend naar de experimentele groep waar die persoon in zat.

De APA heeft geen duidelijk richtlijnen voor het rapporteren van de resultaten van de discriminant analyse. Aangeraden worden het percentage verklaarde variantie, de gekwadraterde canonical correlatie voor elke variat en de chi-square statistiek weer te geven.

Tenslotte moet de MANOVA in zijn geheel geïnterpreteerd worden. Het makkelijkst is om hierbij grafieken te gebruiken, zoals op pagina 661.