

5. Het bias beest

Bias wordt gebruikt in drie contexten:

1. Er zijn dingen zorgen voor een bias in de schattingen van de parameters
2. Dingen die zorgen voor een bias van de standard error en het betrouwbaarheidsinterval
3. Dingen die zorgen voor een bias in de test statistieken en p-waarden

Assumpties

Een belangrijke bron van bias is het schenden van de assumpties van statistische toetsen. Wanneer er aannames geschonden zijn, kunnen we geen nauwkeurige conclusies trekken, omdat de aannames iets zeggen over het model.

De meeste statistische procedures die besproken worden in het boek zijn *parametrische toetsen* (toetsen die gebaseerd zijn op de normale verdeling). Voor het verkrijgen van veel gegevens die parametrisch moeten zijn, worden aannames gedaan die waar moeten zijn. Dit zijn parametrische toetsen. Het is belangrijk dat je de aannames bekijkt voordat je een statistische toets kiest. Er zijn vier grote assumpties voor parametrische toetsen:

Normaalverdeling: De gegevens moeten normaal verdeeld zijn.

Homoscedasticiteit: Deze aanname wil zeggen dat de varianties van de verschillende gegevens ongeveer even groot moeten zijn.

Lineair verband. Voor toetsen gebaseerd op een lineair model, is de assumptie dat de variabelen ook lineair met elkaar samenhangen.

Onafhankelijkheid: Deze aanname kan op verschillende manieren opgevat worden. Het kan zijn dat de deelnemers onafhankelijk van elkaar een vragenlijst in moeten vullen. Bij herhaalde metingen kan dit niet bereikt worden maar wordt verwacht dat de gedragingen van de verschillende deelnemers onafhankelijk van elkaar zijn.

Uitschieters

Uitschieters, scores die erg afwijken van de rest van de data, zijn een belangrijke bron van bias. Een uitschieter kan bijvoorbeeld het gemiddelde enorm omhoog of omlaag brengen, wat zorgt voor een vertekening. Omdat er in de SS gebruik wordt gemaakt van gekwadrateerde deviaties van het gemiddelde, heeft een uitschieter hierop een heel grote invloed. Omdat je met de SS de standaardafwijking en de standard error berekend, en die gebruikt worden voor het betrouwbaarheidsinterval, zorgt dit voor een enorme bias in de test statistiek.

Lineariteit

Zoals gezegd is het schenden van assumpties ook een belangrijke bron van bias. De eerste assumptie die wordt besproken is die van lineariteit.

De meeste statistische methodes zijn gebaseerd op lineaire modellen. De assumptie die hieraan ten grondslag ligt, is dat de variabelen ook op een lineaire wijze met elkaar samenhangen. Deze aanname is zeer belangrijk. Als de variabelen namelijk niet een lineair verband vertonen, klopt je hele model niet.

Normaliteit

In veel statistische toetsen wordt aangenomen dat de verdeling van de steekproeven normaal is. Dit is belangrijk vanwege vier belangrijke dingen die we doen in de statistiek:

- Schattingen van parameters. Veel parameters, zoals het gemiddelde, kunnen erg beïnvloed worden door niet normaal verdeelde distributies. Uitschieters kunnen een gemiddelde enorm vertekenen. Met veel uitschieters is een verdeling niet meer normaal.
- Betrouwbaarheidsinterval. Om een betrouwbaarheidsinterval te berekenen, maken we gebruik van de z-scores, dus van een standaard normale verdeling.
- Nulhypothese significantie toetsen. Bij het hypothese toetsen gaan we ervan uit dat de parameter normaal verdeeld is, omdat de distributies van de test statistieken die we gebruiken daaraan gerelateerd zijn.
- Error. Als de residuals (de error van elke score) normaal verdeeld zijn in de populatie, kan de least squares method worden toegepast om de parameters te schatten.

Hiermee samenhangend kan de ‘assumptie van normaliteit’ ook verschillende betekenissen hebben. Voor een betrouwbaarheidsinterval moet de geschatte parameter uit een normale verdeling komen. Voor het hypothesetoetsen moet de steekproefverdeling normaal zijn. Echter, in de praktijk komt de assumptie neer op het testen of je data normaal verdeeld is, omdat als de data normaal verdeeld is, je redelijkerwijs kunt aannemen dat de error en steekproefverdeling ook normaal verdeeld zijn.

Volgens de centrale limiet stelling is bij grote steekproeven ($n > 30$) de verdeling sowieso normaal, ook al is het in de populatie niet normaal verdeeld. De centrale limiet stelling betekent dat er veel situaties zijn waarbij we normaliteit kunnen aannemen, ongeacht hoe onze data eruit ziet.

Homoscedasticiteit

Homoscedasticiteit betekent homogeniteit van variantie. Deze assumptie heeft invloed op het schatten van parameters in je regressiemodel en hypothese toetsen.

Als je verschillende groepen participanten hebt, betekent deze aanname van homoscedasticiteit dat elk van deze steekproeven uit populaties met dezelfde variantie komen. Bij correlatieel onderzoek betekent het dat de variantie van je uitkomstvariabele stabiel is over verschillende waarden van de predictor.

Als deze aanname geschonden wordt, zijn je schattingen van parameters in een regressiemodel niet optimaal, maar meestal zijn de schattingen ook bij schending niet gebiased. Heteroscedasticiteit (=ongelijke varianties) zorgt echter wel voor een bias bij het schatten van de standard error in je model. Hierdoor zijn ook je betrouwbaarheidsinterval en de hypothesetoetsen niet meer accuraat, omdat die gebruik maken van de standard error.

Onafhankelijkheid

Deze assumptie betekent dat de errors in je model niet aan elkaar gerelateerd zijn. Deze assumptie wordt bijvoorbeeld geschonden wanneer participanten overleggen tijdens een vragenlijst. Het antwoord van de een beïnvloedt hierdoor het antwoord van de ander.

De standard error kun je alleen berekenen als aan deze assumptie is voldaan. Significantietoetsen en betrouwbaarheidsintervallen berekenen zijn bij schending dus geen optie. Voor het schatten van parameters in je model kan het nog wel gebruikt worden, maar krijg je geen optimale resultaten.

Spot de bias

Geïsoleerde uitschieters zijn makkelijk te herkennen als je een grafiek maakt, bijvoorbeeld een histogram, of een boxplot. Ook voor het bekijken van de vorm van de verdeling is een histogram een goede manier.

Een andere grafiek om de verdeling goed te bekijken is een *P-P plot* (probability-probability plot). Deze grafiek laat de cumulatieve kans van een variabele tegen de cumulatieve kans van een bepaalde verdeling zien. Dit betekent dat de gegevens gesorteerd worden en de bijbehorende z-score wordt berekend. Dit is dan de verwachte waarde dat de score zal hebben in een normaalverdeling. De score zelf wordt ook omgezet in een z-score. Dan wordt de werkelijke z-score uitgezet tegen de verwachte z-score. Deze grafiek is een rechte lijn met een normaalverdeling. Voor een P-P plot in SPSS ga je naar Analyze - Descriptive statistics - P-P plots.

In kleinere steekproeven kun je naar frequenties kijken om de normaliteit te controleren (via analyse - descriptive statistics - frequenties). SPSS geeft het in een tabel weer en bij charts kan je meerdere opties aanvinken voor een grafiek (zie blz. 183). Bij statistics kan je aangeven welke statistieken er in de tabel vermeld moeten worden. Scheefheid en kurtosis moeten bij een normaalverdeling 0 zijn. Als de scheefheid een positieve waarde heeft zitten de scores meer aan de linkerkant en bij een negatieve waarde zitten de scores meer aan de rechterkant. Voor kurtosis geldt dat een positieve waarde een puntige verdeling geeft en een negatieve waarde geeft een platte verdeling.

Een normaalverdeling toetsen

Door de scores in z-scores om te zetten kunnen de verschillende steekproeven met verschillende metingen beter met elkaar vergeleken worden. Bovendien kun je dan een p-waarde berekenen om te kijken of de waarde significant afwijkt van 0. Hoe verder de waardes van 0 afliggen hoe minder de verdeling op een normaalverdeling lijkt. Je zet de z-scores om door ze te delen door de standaard meetfout.

Hoewel je dus kunt testen of je data normaal verdeeld is, is dit niet altijd een goed idee. Deze toetsen zijn namelijk gebaseerd op nulhypothese significantie toetsen, en hebben dezelfde problemen. Een heel kleine afwijking van normaal is in grote steekproeven toch significant, en een heel grote afwijking kan in kleine steekproeven niet significant zijn.

Een normaalverdeling toetsen kan door naar de hele verdeling ineen te kijken. De *Kolmogorov-Smirnov test* en de *Shapiro-Wilk test* vergelijken de verdeling van de steekproef met een normaalverdeling met hetzelfde gemiddelde en een zelfde standaard deviatie. Als de test niet significant is, betekent het dat de verdeling van de steekproef niet significant verschilt van een normaalverdeling (de verdeling van de steekproef is dan normaal). En als de test wel significant is, is de verdeling van de steekproef niet normaal.

De K-S test is in SPSS te vinden met analyse- descriptive statistics - Explore (zie voor de afbeelding blz. 186). Bij plots kan je Normality plots with tests aanvinken om de K-S uit te kunnen voeren. Ook wordt er dan door SPSS een *Q-Q plot* gemaakt. Dit is hetzelfde als de P-P plot maar dan worden niet de individuele scores weergegeven maar de *kwartielen*. Kwartielen delen de gegevens in gelijke stukken. *Percentielen* delen de gegevens in 100 stukjes, *nonielen* delen de gegevens in 9 stukken enz. Kortom de Q-Q plot is makkelijker te interpreteren wanneer je veel scores hebt.

In een Q-Q plot moeten de punten op de lijn liggen voor een normaalverdeling. Er is kurtosis als de punten consequent boven of onder de lijn liggen. Wanneer de punten een S-vorm hebben is er sprake van scheefheid.

De letter waarmee de K-S test wordt aangegeven is D(). Hetgeen wat tussen de haakjes komt te staan zijn de vrijheidsgraden.

De S-W test werkt hetzelfde als de K-S test. Alleen heeft deze test meer power om verschillen in normaliteit te vinden.

De gegevens kunnen meerdere groepen bevatten. Met de split file is het in SPSS mogelijk deze groepen van elkaar te onderscheiden en ze zo één voor één goed te kunnen bekijken. Voor een voorbeeld waar alle statistieken in een frequentietabel staan zie blz. 189 en 190. De beschrijvende statistieken en histogrammen zijn een goede manier om een juist beeld van de verdeling te krijgen. Om groepen op te splitsen ga je naar data - split file - organize output by groups. Hierna kan je analyses uitvoeren met de groepen gesplitst. Als je ermee klaar bent kan je de optie weer uitzetten door bij split file de optie analyze all cases, do not create groups.

Herken de homoscedasticiteit en lineariteit

De assumpties van homoscedasticiteit en lineariteit kunnen tegelijk bekeken worden. Beide assumpties gaan namelijk over de residuals, de error in het model. Om ze te controleren, maak je een scatterplot, waarbij je de residuals uitzet tegen de uitkomst volgens het model. Hiermee kijk je of er een verband is tussen de waardes die het model voorspelt en de error. Als de assumpties van homoscedasticiteit en lineariteit niet geschonden zijn, hoort hier geen verband tussen te zijn, en is het scatter plot dus een random stipjespatroon. Op pagina 192 zie je hoe het scatter plot eruit ziet als er sprake is van heteroscedasticiteit en/of non-lineariteit.

Levene's toets

Om de homogeniteit van varianties te toetsen wordt *Levene's test* gebruikt. Deze toets test de nulhypothese waarbij de variantie in de verschillende groepen gelijk is. Het gaat samen met een één-weg ANOVA. Er wordt gekeken naar het absolute verschil tussen elke score en het gemiddelde van de groep waar het vandaan kwam. Wanneer Levene's test significant is betekent het dat de nulhypothese niet klopt en de varianties dus significant van elkaar verschillen. Niet-significantie betekent dat de varianties ongeveer gelijk zijn aan elkaar.

Levene's test werkt alleen goed als je gelijke groepsgrootte en grote steekproeven hebt. Maar bij deze omstandigheden maakt een schending van de assumptie van heteroscedasticiteit nauwelijks uit.

Met de *Hartley's F_{max}* (of *variantie ratio*), kan je de variantie ook bekijken. Hierbij wordt de ratio berekend tussen de groep met de grootste variantie en de groep met de kleinste variantie.

Deze ratio wordt dan vergeleken met kritieke waarden in een tabel. De kritieke waarden hangen af van de grootte van de steekproef.

In SPSS vind je Levene's test bij analyse - descriptive statistics - explore. Bij plots kan je Untransformed aanvinken wat Levene's test voor ruwe scores is. Levene's test wordt aangegeven met de letter F. Voor het noteren gebruik je $F(df1, df2) = \text{waarde}$.

Bias verminderen

Er zijn vier manieren waarop je problemen in je data kunt aanpakken:

1. Bepaalde extreme scores verwijderen (trimmen).
2. Uitschieters vervangen door de hoogste score die geen uitschieter is.
3. Je data analyseren met robuuste methodes.
4. De data transformeren.

Trimmen

Een extreme score verwijderen kan door het verwijderen van de data van de desbetreffende participant. Dit kan alleen als je zeker weet dat de persoon er echt buiten valt, dus uit een andere populatie komt. Je kunt ook trimmen op basis van een bepaald percentage of op basis van de standaardafwijking.

Trimmen op basis van een percentage betekent dat je de scores rangschikt op volgorde (net als bij het berekenen van de mediaan) en dan een bepaald percentage van beide einden weghaalt. Dus (bijvoorbeeld) 5% extreme scores links en 5% extreme scores rechts. Het gemiddelde heet dan een *getrimd gemiddelde*. Een M-estimator kan soms voor je bepalen hoeveel data getrimd moet worden om een robuuste meting te geven.

Trimmen op basis van de standaardafwijking betekent dat je de scores weghaalt die een bepaald aantal standaardafwijkingen boven of onder het gemiddelde liggen. Vaak wordt een waarde van 2,5 standaardafwijking aangehouden. Het grote probleem van deze methode is dat de standaardafwijking en het gemiddelde beiden erg worden beïnvloed door de uitschieters die je met trimmen juist weg probeert te krijgen.

Uitschieter vervangen

Een andere methode is de score van de uitschieter vervangen door de hoogste score die geen uitschieter is. Dit is geen fraude als je score helemaal niet representatief is voor je steekproef en een bias is voor je hele model. Als je deze methode gebruikt om de resultaten in jouw voordeel te draaien, dan is het wel fraude. Een variant hierop is de uitschieter vervangen door drie standaardafwijkingen van het gemiddelde.

Robuuste methodes

De beste methode om bias te verminderen is gebruik maken van robuuste methodes. Dit zijn testen die weinig beïnvloed worden door schendingen van assumpties. Een andere optie is kiezen voor een non-parametrische test, een test die niet gebaseerd is op de assumptie van normaliteit. Hierover gaat hoofdstuk 6.

Het getrimde gemiddelde en de M-estimator zijn voorbeelden van robuuste tests. De tweede methode is *bootstrapping*. Dit is gebaseerd op het idee dat de steekproefdata niet normaal verdeeld hoeft te zijn, zolang de steekproefverdeling maar normaal verdeeld is. Bootstrapping gebruikt de steekproef om de steekproefverdeling te schatten, door uit de steekproef steeds willekeurig een ‘ministeekproef’ samen te stellen en daar het gemiddelde van te berekenen.

Data transformeren

- De laatste optie is het transformeren van de data. Dit betekent dat je alle scores omzet in andere scores. Hierbij verander je alle scores op dezelfde manier (daarom is deze optie toegestaan), om te corrigeren voor problemen met assumpties. De relatie tussen de variabelen verandert hier niet mee, maar de verschillen tussen de verschillende variabelen veranderen wel. Een aantal transformaties:
- Log transformatie ($\log(X_i)$): Deze transformatie kan correctie geven aan positieve scheefheid en ongelijke varianties. Het verkleint het rechterdeel van de verdeling. Een nadeel van deze transformatie is dat je voor het getal 0 en alle negatieve getallen eerst een constante toe moet voegen.
- Wortel transformatie ($\sqrt{X_i}$): Deze transformatie kan correctie geven aan positieve scheefheid en ongelijke varianties. Het heeft meer effect op grote waarden dan op kleine waarden. Dus alle grote scores zullen dicht bij het gemiddelde komen. Maar ook hier heb je het probleem met de negatieve getallen.
- Wederkerige transformatie ($1 / X_i$): Deze transformatie kan correctie geven aan positieve scheefheid en ongelijke varianties. Als je 1 door elke score deelt verkleint dit de grote scores, maar de kleine scores worden na de transformatie juist groter. Dit kan je voorkomen door de transformatie met $1 / (X_{\text{hoogste}} - X_i)$ uit te voeren.

De gegevens transformeren is een kwestie van proberen welke transformatie het beste is voor de gegevens. Let wel op dat je dezelfde transformatie bij alle variabelen moet uitvoeren. Bij Levene's test (via explore) kun je de optie transformed aanklikken. SPSS geeft dan een lijst van transformaties waaruit je kan kiezen.

In SPSS ga je naar transform -> compute variable om bij de transformaties te komen (zie blz 204). Bij target variabele kan je de nieuwe naam voor de variabele geven, bij function Group vind je verschillende categorieën van functies en bij functions and special variables vind je de functies binnen de geselecteerde categorie. Op blz. 205 en 206 is een aantal functies te vinden met de iconen erbij. Op blz. 207 zijn de meest bruikbare functies te vinden.