

Hoofdstuk I De statistiek in het onderzoek

Doelstellingen:

- De student kent de diverse fasen van het onderzoek en begrijpt de rol van de statistiek in dat kader.
- De student begrijpt de besproken begrippen.
- De student kent de diverse vormen van steekproeftrekking en ook de mogelijkheden en beperkingen van elk van deze vormen
- De student kan een datafile aanmaken in SPSS, een file opslaan en openen.

1.1 Wat is beschrijvende statistiek?

Statistiek: de wetenschap van het verzamelen van gegevens, he classificeren, samenvatten, organiseren, analyseren en interpreteren van deze informatie.

Beschrijvende statistiek: gegevens van een onderzoek voor te stellen via kengetallen en/of een grafische presentatie

Inductieve statistiek: op grond van steekproefgegevens iets trachten te besluiten over de populatie

1.2 Het onderzoek

Onderzoek: op basis van waarnemingen ware en algemene uitspraken doen over de werkelijkheid

Een uitspraak: een bewering waarin één of meerdere objecten een eigenschap wordt toegeschreven

Wetenschappelijk onderzoek behoort aan 4 voorwaarden te voldoen:

1. Objectiviteit: afzien van vooroordelen en zich openstellen voor de werkelijkheid zoals ze is
2. Controleerbaarheid: de onderzoeker geeft kansen om bevindingen te controleren via het aangeven van de wijze waarop deze bevindingen tot stand zijn gekomen
3. Herhaalbaarheid: andere onderzoekers moeten het onderzoek op exact dezelfde wijze kunnen herhalen of repliceren
4. Systematiek: de onderzoeker volgt een bepaalde lijn, er zit een systeem achter

Natuurwetenschappen → deterministische uitspraken (= niets ontsnapt aan een wetmatigheid)

Menswetenschappen → probabilistische uitspraken (= uitspraken die nooit zeker kunnen zijn)

Variabele: een eigenschap/kenmerk van de onderzoekseenheid (bv. Geslacht) $\leftrightarrow$ **constante**

1. Afhankelijke variabele: de eigenschap ter studie → veranderbaar
2. Onafhankelijke variabele: oorzak van de verschillen in de afhankelijke variabele → constant

Soorten onderzoek:

1. Experiment: één of enkele variabelen worden gemanipuleerd en de impact op de afhankelijke variabele wordt onderzocht
→ mogelijk om causale verbanden vast te stellen!!
2. Veldexperiment: survey- of enquêteonderzoek: er worden geen variabelen gemanipuleerd, maar er wordt een steekproef getrokken uit de populatie.
→ geen mogelijkheid tot causale verbanden !!

1.3 Fasen in het onderzoeksproces

Zes fasen in een onderzoek:

1. De vraagstelling
2. De operationalisering
3. De steekproefopzet
4. Het verzamelen van de gegevens
5. De beschrijving van de gegevens
6. De formulering van de conclusies

1.3.1 De vraagstelling

De vraag waarop het onderzoek een antwoord moet geven

1.3.2 De operationalisering:

Hoe kunnen we een begrip concreet maken?

1.3.3 De steekproefopzet:

Populatie: het geheel van individuen waar de onderzoeker iets over wil vaststellen en die bijgevolg in aanmerking komen voor het onderzoek

- Census: indien alle individuen in de populatie uitgenodigd worden om aan het onderzoek mee te doen

Steekproef: een selectie van individuen uit de populatie

Aselecte steekproeftrekking/at random sampling: elk individu van de populatie heeft evenveel kans om in de steekproef terecht te komen

Niet-aselecte steekproef: elk individu heeft NIET evenveel kans om getrokken te worden

1.3.3.1 De aselechte proefopzet

Steekproefkader: een lijst van al de individuen van de populatie

Representativiteit: de samenstelling van de steekproef loopt gelijk met die van de populatie

Vijf vormen van een aselechte steekproef:

1. Volledig aselechte steekproeftrekking: steekproefkader opstellen, elk individu nummeren, computer geeft willekeurig een aantal nummers. Deze individuen worden gekozen.
2. Systematische steekproef: er wordt een eerste willekeurige pp uitgekozen, vervolgens wordt een nummer gekozen dat x waarden hoger ligt.
3. Gestratificeerde steekproef: populatie in strata (subgroepen) indelen, daarna aselekt een steekproef trekken uit elke subgroep
 - a. Proportioneel: de verhoudingen tussen de subgroepen in de steekproef lopen parallel met die van de populatie
 - b. Disproportioneel: de verhoudingen tussen de subgroepen in de steekproef lopen NIET parallel met die van de populatie
4. Clustersteekproef: de populatie wordt ingedeeld in subgroepen die zo heterogeen mogelijk zijn, vervolgens wordt een aantal subgroepen geselecteerd.
5. Getrapte steekproef: een combinatie van bovenstaande trekkingen

1.3.3.2 De niet-aselechte steekproef

= snel & goedkoop

→ NIET veralgemeenbaar naar de populatie

Vijf vormen van een niet-aselechte steekproef:

1. Convenience sampling of gemakkelijkssteekproef: individuen worden gekozen omdat ze binnen handbereik zijn en/of het gemakkelijkst benaderd kunnen worden
2. Judgement sample of beoordelingssteekproef: bevraging van de bevoorrechte getuigen (bv. De zware gebruikers van een product ondervragen)
3. Snowball sampling of sneeuwbalsteekproef: voor populaties die zich moeilijk laten benaderen; één persoon zoeken die voldoet aan de voorwaarden en dan via die persoon andere zoeken
4. Quota sampling of quotasteekproef: populatie wordt in gedeeld in subgroepen en vervolgens worden er in elke deelpopulatie deelnemers gezocht volgens het gemakkelijksheidsprincipe
5. Random walk of routemethode: de proefleider zal pp kiezen volgens een vooraf vastgelegd wandeltraject

1.3.4 Het verzamelen van de gegevens

ALTIJD eerst de variabele view ingeven!

Beperkingen bij ingeven van een variabele:

- Een naam mag maximaal uit 64 tekens bestaan (conventie: 8 tekens)
- Een naam moet beginnen met een letter
- Een naam mag niet eindigen op een punt of een underscore
- Spaties en gereserveerde tekens (&, !, ?, *, /, enz.) mogen niet gebruikt worden
- De woorden all, and, by, eq, ge, gt, le, lt, ne, not, or, to, with zijn niet toegestaan
- De variabelennaam is niet hoofdlettergevoelig
- Een naam mag slechts 1 keer voorkomen

Name: naam van de variabele met bovenstaande beperkingen

Type: type van variabele (numeriek, alfanumeriek, enz.)

Width: maximaal aantal karakters

Decimals: hoeveel decimale cijfers

Label: verdere specificatie van variabele

Values: enkel bij kwalitatieve variabele (bv. Geslacht: 1 = vrouw, 2 = man)

Missing: indien respondenten niets hebben ingevuld, waarde gebruiken die voorkomt bij de

variabele

Columns: hoeveel ruimte heb je nodig per cel

Align: mogelijk om waarden per cel links of rechts uit te lijnen of te centreren

Measure: meetniveau van de variabele

Extensies:

- Databestand: .sav
- Outputbestand: .spv

1.3.5 De beschrijving van gegevens

Hoe kunnen we de gegevens op een overzichtelijke manier samenvatten?

1.3.6 Het formuleren van conclusies

Welke betekenis hebben de gegevens van de steekproef ten opzichte van de populatie?

Deductieve redenering: uitgaande van bestaande waarnemingen nieuwe kennis opbouwen

Inductieve redenering: op basis van empirische observaties een wet formuleren

- Alternatieve hypothese: de stelling die de onderzoeker tracht te bewijzen
- Nulhypothese: het tegengestelde van de alternatieve hypothese

Rapportage: in een tijdschriftartikel, presentatie, rapport voor opdrachtgever, enz.

Hoofdstuk II Operationalisering en meten van variabelen

Op welke wijze kunnen begrippen in het kader van een onderzoek geoperationaliseerd worden en wat zijn de diverse niveaus van meting?

Doelstellingen:

- De student begrijpt de begrippen variabele, operationalisering, kwantitatieve versus kwalitatieve variabelen, discrete versus continue variabele, dichotome variabele, betrouwbaarheid en validiteit.
- De student kent en begrijpt de betekenis van de vier soorten meetniveaus in de statistiek.
- De student begrijpt de consequenties van deze vier types van meetniveaus
- De student kan via SPSS variabelen bewerken via Transform – compute, recode en count.
- De student kan de Cronbach's Alpha berekenen van een schaal.

2.1 Variabelen

Operationaliseren: precies aangeven hoe een begrip zal gemeten worden

Kwantitatieve variabele: waarden bestaan uit getallen/scores (scale)

Kwalitatieve variabele: waarden bestaan uit indeling in categorieën (nominaal en ordinaal)

Discrete variabele: aantal waarden voor de variabele is beperkt, kan enkel gehele waarden aannemen

Continue variabele: het aantal tussenliggende waarden is onbegrensd

2.2 Meetniveaus

We onderscheiden vier meetniveaus: Nominaal, Ordinaal, Interval en Ratio (NOIR)

→ in SPSS maken we enkel gebruik van nominaal, ordinaal en scale (interval + ratio)

2.2.1 Nominaal meetniveau

Men kan geen rangordering maken van deze waarden, er is geen meeteenheid en een nulpunt ontbreekt.

Alle nominale variabelen zijn kwalitatieve variabelen. Ze kunnen wel gekoppeld worden aan getallen. Ze kunnen de variabelen ook in groepen verdelen, er wordt ook wel eens gesproken van categorische variabelen.

De waarden moeten exhaustief (elke deelnemer moet geplaatst kunnen worden) en mutueel exclusief (elke deelnemer kan slechts in één categorie geplaatst worden) zijn

Dichotome variabele: een nominale variabele die slechts 2 waarden kan aannemen (bv. Een Ja-Nee vraag).

Bv. Geslacht, beroep, ...

2.2.2 Ordinaal meetniveau

We kunnen de variabelen ook enken indelen in categorieën, maar we kunnen ze ook een rangordering geven. Er is echter geen meeteenheid.

Bv. Hoogst genoten opleiding (lager onderwijs, lager secundair, hoger secundair, bachelor, master)

2.2.3 Intervalmeetniveau

We kunnen een rangordering maken en we hebben informatie over de onderlinge afstanden tussen de opeenvolgende waarden. We hebben dus een meeteenheid, maar geen absoluut nulpunt.

Bv. Intelligentie, temperatuur, jaartelling, testscore, ...

2.2.4 Rationmeetniveau

We kunnen een rangordering maken, de onderlinge afstanden tussen opeenvolgende waarden zijn gelijk, er is een absoluut nulpunt.

Bv. Leeftijd, tijd om een test op te lossen, lichaamslengte, ...

Komt weinig voor in de gedragswetenschappen.

De meetniveaus vormen een hiërarchie. Alle eigenschappen van de lagere meetniveaus komen eveneens voor bij de hogere meetniveaus.

2.3 Betekenis van deze meetniveaus

De aard van het meetniveau zal bepalend zijn voor de wijze van grafische voorstelling van de gegevens en voor de berekening van tal van kengetallen.

Eenzelfde variabele kan van niveau wijzigen al naargelang de vraagstelling. Bekwame onderzoekers proberen zoveel mogelijk gebruik te maken van een zo hoog mogelijk meetniveau.

Likertschaal: schaal die vraagt naar houdingen, overtuigingen, gevoelens, ...

1. Helemaal akkoord
2. Akkoord
3. Neutraal
4. Niet akkoord
5. Helemaal niet akkoord

Technisch gezien behoort deze schaal tot de ordinale schaal, maar door de associatie tussen deze antwoorden en getallen, beschouwen we deze als intervalschaal.

2.4 Betrouwbaarheid en validiteit

Betrouwbaarheid: heeft te maken met de stabiliteit van de meting, indien deze onstabiel is, zeggen we dat deze onbetrouwbaar is. We kunnen de betrouwbaarheid meten aan de hand van vier procedures:

1. Hertesting
2. Halvering
3. Parallelvorm
4. Interne homogeniteit

Interne homogeniteit: worden de items van eenzelfde test door de respondenten op eenzelfde manier beantwoord? De samenhang wordt bestudeerd. → **Cronbachs alfa**

Validiteit: meet de test wat hij moet meten? Hiervoor bestudeert men de samenhang tussen de uitslag van de test en een andere meting (= criterium) extern aan deze test.

De betrouwbaarheid vormt een noodzakelijke voorwaarde voor de validiteit!

2.5 Voorbereidende bewerkingen in SPSS

Transform-compute: mogelijkheid om nieuwe variabelen aan te maken op basis van bestaande variabelen.

Transform-recode into different variables: nieuwe variabelen aanmaken door een bestaande te hercoderen.

Transform-count: mogelijkheid om per waarnemingsobject bepaalde scores te tellen.

2.5.1 Transform-recode

= hercoderen van variabelen

Bv. Het grootste deel van items zijn positief opgesteld, enkelen echter negatief. We hercoderen dan de negatieve naar positieve items zodat men verdere bewerkingen kan uitvoeren.

- Into different variables: de oorspronkelijke gegevens blijven behouden (ALTIJD gebruiken)
- Into same variables: de oorspronkelijke gegevens worden vervangen

2.5.2 Transform-compute

Op basis van gegevens een totaalscore maken op 2 manieren:

1. Sommatie maken van scores per item, MAAR problemen indien er missing values zijn.
2. MEAN: SPSS geeft voor elke respondent een uitslag, ook indien er missing values zijn.

2.5.3 Transform-count

Hoe vaak wordt een bepaalde score gehaald?

2.5.4 Cronbach's alpha

In hoeverre zijn de items van een schaal op eenzelfde wijze ingevuld door de respondenten?

In de economie is .60 al voldoende, in de psychologie is er echter een betrouwbaarheid van .70 nodig.

Heel goede betrouwbaarheid indien de Cronbach's alpha hoger is dan .80.

Hoofdstuk III De frequentieverdeling

Na het verzamelen van gegevens tracht de onderzoeker de resultaten op een overzichtelijke wijze te presenteren zodat bepaalde tendensen zichtbaar worden.

Doelstellingen:

- De student begrijpt de begrippen die hierbij aan bod komen, zoals datamatrix, frequentietabel, enz.
- De student kan een univariate en meerdimensionale tabel van de resultaten opstellen.
- De student kent de diverse vormen van grafische voorstellingen van de gegevens.
- De student weet hoe het schaalniveau een impact heeft op de wijze van grafische voorstelling.
- De student kan percentiel/kwartielen bepalen uit een frequentietabel.
- De student kan via SPSS een frequentietabel en een eenvoudige grafische voorstelling van de gegevens maken.

3.1 De frequentietabel

Eendimensionale tabel: slechts één ingang, dus ook één variabele.

Meerdimensionale tabel of kruistabel of contingentietabel: relatie tussen twee nominale variabelen

Absolute frequentie: aantal keren dat een waarde voorkomt (bij elk niveau)

Relatieve frequentie: absolute frequentie gedeeld door het aantal respondenten = **proportie** (bij elk niveau)

Cumulatieve frequentie: hoeveel observaties zijn lager of gelijk aan een bepaalde waarde (minimaal ordinaal)

Cumulatieve proportie: hoeveel procent van de pp hebben eenzelfde of een lager score (minimaal ordinaal)

3.2 Grafische voorstelling van de gegevens

Een grafiek oogt veel beter dan een tabel.

Een taartdiagram (pie chart): voor kwalitatieve variabelen, vooral gebruikt bij nominale variabelen.

→ analyse – descriptive statistics – frequencies – charts

Staafdiagram (bar chart): verticale balken grenzen niet aan mekaar, vooral bij ordinale variabelen.

→ analyse – descriptive statistics – frequencies – charts

Histogram: lijkt op een staafdiagram, maar de balken grenzen wel aan mekaar. Voor interval- of rationniveau. Indien we een veelheid hebben aan waarden, zal SPSS deze automatisch gaan groeperen.

→ analyse – descriptive statistics – frequencies – charts

Stamdiagram (stem-and-leafplot): voordeel is dat de individuele observaties gepresenteerd worden. Anderzijds oogt het niet echt mooi.

→ analyse – descriptive statistics – explore – plots

Boxplot:

- De doos: 50 % middelste observaties
- Boven en onder de doos: de overige 50 %
- Horizontale lijn in de doos: boven deze lijn 50 %, onder deze lijn 50 %

→ graphs – legacy dialogs – boxplot – simple

Via graphs – legacy dialogs – bar – simple/clustered/stacked

Ook mogelijk om slechts een deel van de pp te gebruiken: data – select cases

3.3 Percentielen

Hoeveel procent van de deelnemers eenzelfde of een zwakkere score behaalde.

→ analyze – descriptive statistics – frequencies – statistics – percentiles

Kwartiel: Q_1 , Q_2 , Q_3

Deciel: P_{10} , P_{20} , ... P_{100}

Bij nominale waarden kan er geen sprake zijn van Percentielen of kwartielen.

Hoofdstuk IV Centrummaten

De centrummaten geven aan wat de centrale tendens is van de verdeling. De variabiliteitsmaten geven aan hoe groot de onderlinge verschillen zijn tussen de diverse uitslagen.

Doelstellingen:

- De student kent de diverse begrippen over de centrummaten.
- De student kent de impact van de aard van de schaal op de bepaling van de centrale tendens.
- De student kan – handmatig – de centrale tendens berekenen voor een (kleine) verdeling van uitslagen.
- Via SPSS kan de student de centrale tendens van een reeks gegevens berekenen en hij kan hierbij rekening houden met het schaalniveau.
- De student kan de output van SPSS interpreteren.

4.1 De modus

= de waarde met de hoogste frequentie.

Kan in principe worden toegepast op elk meetniveau, maar wordt vooral gebruikt op nominaal of ordinaal niveau.

Bimodaal: indien er 2 modi voorkomen bij een verdeling

Multimodaal: indien de verdeling meer dan 2 modi heeft

4.2 De mediaan

= de middelste waarde (wanneer de observaties van laag naar hoog zijn gerangschikt)

Kan NIET op nominaal niveau.

4.3 Het rekenkundig gemiddelde

Enkel mogelijk op interval- en rationiveau.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Gewogen gemiddelde: gemiddelde van de samengestelde proefgroep waarin beide groepen zijn opgenomen:

Alle scores hebben een impact op het gemiddelde.

Getrimde gemiddelde: het gemiddelde berekenen na weglating van een bepaald percentage hoogste en laagste uitslagen.

Eigenschappen:

- De som van de afwijkingen van de scores tot het gemiddelde is gelijk aan nul.
- Bij een lineaire transformatie zal het rekenkundig gemiddelde op dezelfde manier getransformeerd worden:
- Het gemiddelde varieert minder van steekproef tot steekproef in vergelijking met de mediaan.

Een rechtsscheve verdeling of positieve scheefheid: $\text{modus} < \text{mediaan} < \text{gemiddelde}$

Een linksscheve verdeling of negatieve scheefheid: $\text{modus} > \text{mediaan} < \text{gemiddelde}$

4.4 SPSS en de centrale tendens

Analyse – descriptive statistics – frequencies – statistics – mean/median/mode/sum

Vergelijking van groepen: analyze – compare means – means

Hoofdstuk V Variabiliteitsmaten

Liggen alle observaties dicht bij elkaar of eerder verspreid over het gehele continuüm?

Minimaal niveau dat nodig is: ordinaal

Doelstellingen:

- De student kent de diverse maten van variabiliteit en de voor- en nadelen van elk van deze maten.
- De student kan uit een eenvoudig bestand deze maten handmatig en via SPSS berekenen. Bovendien kan hij deze resultaten interpreteren.
- De student begrijpt het effect van een lineaire transformatie op het rekenkundig gemiddelde en de spreidingsmaten.
- De student kan omgaan met Z-waarden en op deze wijze uiteenlopende scores met elkaar vergelijken.
- De student kan een variabele omzetten in Z-waarden en omgekeerd een Z-waarde omzetten naar een verdeling met een willekeurig gemiddelde en standaarddeviatie.
- De student kent de betekenis van T-scores, C-scores en IQ's.

5.1 De range of het bereik

Minimaal van ordinaal niveau. Het verschil tussen de grootste en laagste waarde.

Nadeel: slechts twee observaties hebben invloed op de range.

5.2 De interkwartielafstand

Het 'gebied' waartussen 50 % van de observaties valt oftewel $75\% - 25\% = Q_3 - Q_1$

Semi-interkwartielbereik: de helft van de interkwartielafstand

5.3 De gemiddelde afwijkingsscore

Of deviatiescore: van alle uitslagen het gemiddelde berekenen en daarna voor elke score het verschil bepalen ten opzichte van die gemiddelde. Is natuurlijk altijd nul.

5.4 De gemiddelde absolute afwijkingsscore

Het gemiddelde berekenen van de absolute waarde van de afwijkingsscores. Maar deze maat wordt nooit gebruikt in de statistiek.

5.5 De variantie

De som bepalen van de gekwadrateerde afwijkingsscores en daarna het gemiddelde nemen van deze. Goede en veelgebruikte maat.

→ analyse – descriptive statistics – frequencies – statistics – variance

Indien we handmatig de variantie berekenen, zien we dat er een kleine afwijking is. Dit komt doordat SPSS een iets andere formule gebruikt, namelijk met $(n-1)$.

5.6 De standaarddeviatie

Het is moeilijk om de variantie te vergelijken met het rekenkundige gemiddelde, om dit op te lossen nemen we de vierkantswortel van de variantie; de standaarddeviatie:

→ analyse – descriptive statistics – frequencies – std. variance

Ook in dit geval is er een verschil tussen het handmatig bepalen en de berekening met SPSS. Ook bij de standaarddeviatie maakt SPSS gebruik van de formule met (n-1).

5.7 De variatiecoëfficiënt

= de standaarddeviatie gedeeld door het gemiddelde

Op deze manier wordt de meeteenheid weggewerkt en krijgen we een relatieve maat van variabiliteit:

Minimaal niveau van ratio nodig.

5.8 Lineaire transformaties

= delen, vermenigvuldigen, optellen en aftrekken

Niet-lineaire transformatie: kwadrateren, enz.

- Vermeerderen of verminderen: standaarddeviatie en variantie wijzigen niet

- Vermenigvuldigen of delen: standaarddeviatie wordt gedeeld of vermenigvuldigd met hetzelfde getal, de variantie wordt gedeeld of vermenigvuldigd met het kwadraat van dat getal.

Standardscore: hoeveel standaarddeviaties is een score verwijderd ten opzichte van het gemiddelde?

Eigenschappen:

- Gemiddelde = 0
- Standaarddeviatie = 1

→ analyse – descriptive statistics – descriptives

Van Z-score naar oorspronkelijke score:

T-scores: gemiddelde = 50,

standaarddeviatie = 10

C-scores: gemiddelde = 5,

standaarddeviatie = 2

→ eerst omzetten naar Z-scores, dan naar T of C

Hoofdstuk VI De normale verdeling

= een vorm van verdeling, ontstaat bij een meting van zeer groot aantal ppn, gekenmerkt door een klokvorm. Dit model kan ook fungeren als kansmodel.

Doelstellingen:

- De student kent de eigenschappen van een normale verdeling.
- De student kent de eigenschappen van de standaardnormale verdeling en kan bepalen wat de linker – c.q. rechteroverschrijdingskans is van bepaalde Z-waarden.
- De student kan in een standaardnormale verdeling de proportie van observaties berekenen binnen een bepaald interval.
- Uitgaande van een bepaalde proportie linker- of rechteroverschrijdingskans kan de student de corresponderende Z-waarden aanduiden.
- De student kan deze principes toepassen in om het even welke normale verdeling.

6.1 De normale verdeling

komt vaak voor, zowel voor psychologische als voor biologische eigenschappen als bedrijfskundige processen.

Eigenschappen:

- De uitslagen liggen geconcentreerd rond het rekenkundig gemiddelde.
- Verdeling is volledig symmetrisch.
- Bij lage en hoge uitslagen: convexe curve
- Bij het midden: concave curve
- Gemiddelde = modus = mediaan
- Minder dan één std deviatie van gemiddelde = 68 %
- Minder dan twee std deviaties van het gemiddelde = 95 %
- Minder dan drie std deviaties van het gemiddelde = 99 à 100 %

6.2 De standaardnormale verdeling

= een verdeling van Z-waarden die normaal verdeeld is.

Linkeroverschrijdingskans: de kans om een uitslag te vinden die kleiner is of gelijk aan (terug te vinden in de tabel)

Rechteroverschrijdingskans: de kans om een uitslag te vinden groter of kleiner aan. (in de tabel: 100% - de Linkeroverschrijdingskans)