**Statistiek A – beschrijvende statistiek**

**Michael Savelkoul| Vakcode: SOW-MTB1002**

**Deze samenvatting is afkomstig uit het collegejaar 2014-2015. Het kan zijn dat sommige onderdelen iets verschillen van de huidige tentamenstof. Op pagina 9 vind je het stappenplan dat je moet doorlopen om de juiste toets te kiezen. Let hier op!**

Cursusomschrijving:

De student:  
• kan gegevens invoeren in spreadsheets (SPSS en Excel) en kan basale data cleaning uitvoeren;  
• kan data modificeren en eenvoudige schaalconstructies maken;   
• is in staat om de gangbare toepassingen van data management uit te voeren;  
• kan de verzamelde gegevens adequaat grafisch weergeven;  
• kan met behulp van de gangbare maten voor centrum en spreiding kwalitatieve en kwantitatieve gegevens beschrijven;  
• heeft kennis van en kan werken met kruistabellen en de daarmee verbonden associatiematen;  
• heeft kennis van en kan werken met correlaties en lineaire regressie-analyse;  
• heeft basale kennis van de toetsende statistiek;  
• kent de meest gangbare procedures binnen SPSS voor de beschrijvende statistiek en kan deze toepassen met behulp van menu gestuurde opdrachten en SPSS-syntax bestanden.

|  |  |  |  |
| --- | --- | --- | --- |
| ***1 Variabele*** | | | |
| **Meetniveau** *(blz.16)* | **Maar voor centrale tendentie** | **Spreidingsmaat** | **Statistische Toets** |
| *Nominaal*   * Kwalitatief * Laatste niveau * categorieën kunnen niet worden geordend | 1. Modus   (Beschrijven met staaf- en cirkeldiagrammen, *blz. 30*) | N.v.t. | * Chi-kwadraattoets (X2) = representativiteit *(blz. 348-360)* * Toets voor 1 proportie/fractie (H8)   (Betrouwbaarheidsintervallen rondom 1 proportie) |
| *Ordinaal*   * Kwalitatief * middelste niveau * categorieën kunnen worden geordend | 1. Mediaan: Q2 *(blz. 42)* 2. Modus *(blz. 44)* 3. Kwartielen: Q1,Q2,Q3 *(blz. 56)*   (Beschrijven met staaf- of cirkeldiagrammen, *blz. 28-31*) | 1. Interkwartielrange: Q3 – Q1 *(blz. 56)* | * Chi-kwadraattoets (X2) = representativiteit *(blz. 348-360)* |
| *Interval / Ratio*   * Kwantitatief * Hoogste niveau * Hebben van nature een numerieke schaal | 1. Gemiddelde: μ, *(blz. 41, 122, 132, 146)* 2. Mediaan *(blz. 42)* 3. Modus *(blz. 44)* 4. Kwartielen: Q1,Q2,Q3 *(blz. 56)*   (Beschrijven met punten-, stam & blad- of histogrammen, *blz. 31-39*) | 1. Standaarddeviatie: σ, s *(blz. 48, 124, 132, 146)* 2. Interkwartielrange: Q3 – Q1 *(blz. 56)* 3. Variantie: σ2, s2 *(blz. 47, 124, 132)* 4. Bereik *(blz. 45)* | Bij n > 30:   * Z-toets voor 1 populatiegemiddelde *(blz. 238)*   Bij n < 30:   * T-toets voor 1 populatiegemiddelde *(blz. 245)*   (SPSS geeft altijd de t-toets)   * Betrouwbaarheidsinterval *(blz. 201, 206, 209, 211)* |

|  |  |  |  |
| --- | --- | --- | --- |
| **2 Variabelen** | | | |
|  | **Nominaal** | **Ordinaal** | **Interval / Ratio** |
| **Nominaal** | 1. Kruistabel:  * Cramers V (samenhangsmaat) * Chi-kwadraattoets: X2 *(blz. 348-360)* | - | - |
| **Ordinaal** | 1. Kruistabel:  * Cramers V (samenhangsmaat) * Chi-kwadraattoets: X2 *(blz. 348-360)*   2. Wilcoxon (2 onafh.)  3. Kruskal - Wallis (≥ 2 onafh. steekpr.) | 1. Rspearman (-1;1) 2. Wilcoxon 2 related samples 3. Friedman bij >3 variabelen 4. evt Kruistabel  * Cramers V (samenhangsmaat) * Chi-kwadraattoets: X2 *(blz. 348-360)*   5. evt. Kruskal-Wallis | - |
| **Interval / Ratio** | 1. Nominaal kenmerk mag maar uit 2 categorieën bestaan:  * Z-toets voor verschil van populatiegemiddelde (bij n > 30) * T-toets voor verschil van populatiegemiddelde (bij n < 30)   (SPSS geeft altijd de t-toets)  • variantie-analyse (>2 onafh. steekpr.)  • ( evt. Wilcoxon of Kruskal-Wallis bij n<30) | 1. Rspearman (-1;1) *(bij correlatie = samenhang)* 2. evt. Kruskal – Wallis 3. Variantieanalyse | 1. Rpearson (-1;1) *(blz. 389-393) (bij correlatie = samenhang)* 2. Enkelvoudige lineaire regressie *(H12)* 3. Multipele regressie *(H13)* 4. Het vergelijken van 2 populatieverwachtingen (gepaarde waarnemingen)  - Z-toets voor gepaarde waarnemingen : test voor gemiddelde op 2 kenmerken (N>30) - T-toets voor gepaarde waarnemingen: test voor gemiddelde op 2 kenmerken (N<30)   (weergeven in spreidingsdiagram of puntenwolk, *blz. 60-61*) |
| **Overige bij**  **Interval/Ratio – Interval Ratio** | **Vergelijken van twee populatieverwachtingen: gepaarde waarneming** (Verschillen in het steekproefgemiddelde(), bij dezelfde x die meerdere keren wordt gemeten):  Bij N > 30:   * Z-toets voor gepaarde waarnemingen *(blz. 280)*   Bij N < 30:   * T-toets voor gepaarde waarnemingen *(blz. 280)*   (SPSS geeft altijd de t-toets)  **Vergelijken van twee populatieverwachtingen: onafhankelijke steekproeven** (gevolgtrekkingen over 2 verschillende parameters)  Bij N > 30:   * Z-toets voor gepaarde waarnemingen *(blz. 270)*   Bij N < 30:   * T-toets voor gepaarde waarnemingen *(blz. 272) \*SPSS geeft altijd de t-toets* | | |

**Samenvatting Statistiek A en B**

* *Stappenplan toetsing*
* *SPSS handelingen bij alle toetsen*
* *Het maken van dummy’s*
* *Tabellen (wanneer je welke toets moet gebruiken)*
* *Welke toets je moet gebruiken wanneer meerdere toetsen mogelijk zijn*
* *Parametrische en non parametrische toetsen*
* *Nuttige (extra) weetjes*
* *Samenvatting Syntax boekje*

**Toetsen met SPSS**

***0. Algemeen***

* Welk meetniveau een variabele heeft, kan worden afgelezen in SPSS bij “Values” onder “Variable View” 🡪 deze klopt echter niet altijd! Door de FREQUENCIES op te vragen van een variabele (via frequencies dus) kun je goed zien welke variabele het is.
* Sig. (2-tailed) bij eenzijdige toets delen door 2
* Met de hand berekenen: Test waarde buiten kritieke waarde (dus in kritiek gebied): H0 verwerpen

***0. Betrouwbaarheidsinterval*** *(p.201)*

* Analyse > Descriptive statistics > Explore
* Voeg onder “Dependent List” de variabele toe waarvoor je het betrouwbaarheidsinterval wilt weten
* Klik op “Statistics”, zorg dat “Discriptives” is aangevinkt, en kies het betrouwbaarheidsinterval (bv. 95%)
* Klik op “Plots” om verschillende beschrijvende figuren aan te vinken.
* Klik op “OK”
* Je kan met (bv 95%) zekerheid zeggen dat het gemiddelde binnen het interval ligt

***1a. Chi-kwadraat (X2)*** *(p. 349)*

***1b. Kruistabel met Chi-kwadraat & Cramers V*** *(p. 357)*

* Analyse > Descriptive Statistics > Crosstabs
* Zet in “Row’s” de Afhankelijke Variabele
* Zet in “Columns” de Onafhankelijke Variabele
* Klik op “Statistics” > Vink “Chi-square” en “Phi and Cramers V” aan en klik op “Continue”
* Klik op “Cells” > Vink onder percentages “Columns” aan en klik op “Continue”
* Klik op “OK”
* X2 Test mag alleen gebruikt worden als niet meer dan 20% van de cellen een “expected count” hebben van “less than 5”
* Sig. van de X2 test = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* X2 Waarde = testwaarde, dus: X2 < X2 α: H0 niet verwerpen X2 > X2 α: H0 wel verwerpen (zie 17+18)

***2a. Z-toets (bij > 30) of T-toets (bij n <30) voor 1 populatiegemiddelde μ*** *(Z: p. 238, T: 253)*

* Analyse > Compare Means > One Sample T Test
* Vul bij “Test Value” de waarde in die onder H0 staat (bv: H0 = 50 > Test Value: 50)
* Klik op “Options” en geef het betrouwbaarheidsinterval aan (90%, 95% of 99%), klik op “Continue”
* Klik op “OK”
* Sig. van One-Sample test = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* Z-waarde = testwaarde

***2b. Z-toets (bij > 30) of T-toets (bij n <30) voor verschil in 2 populatiegemiddelden bij N > 30*** *(Z: p. 270, T: 272)*

* Analyse > Compare Means > Independent-Samples T Test
* Zit in “Test variable(s)”, de kwantitatieve (interval of ratio) variabele
* Zet in “Grouping variable”, de ordinale variabele
* Klik op “Define Groups” en kies bv group 1:0 en group 2:1, klik op “Continue”
* Klik op “Options” en geef het betrouwbaarheidsinterval aan (90%, 95% of 99%), klik op “Continue”
* Klik op “OK”
* F-waarde: geeft aan of Std. Dev’s van de populatiegemiddelden ver uit elkaar liggen
* Sig. van F < α: rij met “Eqial variances NOT assumed” bekijken
* Sig. van F > α: rij met “Eqial variances assumed” bekijken
* Sig. van t-test = p, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* Z-waarde = testwaarde

***3. Variantieanalyse*** *(p. 313)*

* Analyze > Compare Means > Means
* Zet in “Dependent List” de afhankelijke (kwantitatieve: interval of ratio) variabele.
* Zet in “Independent List” de onafhankelijke (kwalitatieve: nominaal of ordinaal) variabele.

Wanneer er sprake is van meerdere onafhankelijke (kwalitatieve) variabele klik dan op “Next” zodat je bij “Layer 2 of 2” de tweede onafhankelijke variabele in de “Independent List” kan zetten.

* Klik op “Options” > Vink “Anova table and eta” aan en klik op “Continue”
* Klik op “OK”
* Sig. van “Between Groups” = p, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen

***4. R van Spearman*** *(p. 643)*

* Analyse > Correlate > Bivariate
* Voeg de 2 variabele toe
* Vink “Spearman” aan
* Kies eenzijdig of tweezijdig toetsen
* Klik op “OK”
* Sig. = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* Correlation coëfficiënt = sterkte van de samenhang (-1:1)

***5. R van Pearson*** *(p. 389)*

* Analyse > Correlate > Bivariate
* Voeg de 2 variabele toe
* Vink “Pearson” aan
* Kies eenzijdig of tweezijdig toetsen
* Klik op “OK”
* Sig. = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* Pearson correlation = sterkte van de samenhang (-1:1)

***6. Enkelvoudige lineare regressie*** *(p.387)*

* Analyse > Regression > Linear
* Voeg onder “Dependent” de Afhankelijke (Y) variabele toe
* Voeg onder “Independent” de Onafhankelijke (X) variabele toe
* Klik op “OK”
* Model Summary: R Square = mate waarin Y wordt verklaard door de onafhankelijke (X)variabele (=R2\*100%)
* Anova: F = bruikbaarheid van het gehele model (y = β0 + β1x1) Als F significant is (Sig. < α) dan is het model bruikbaar om de afhankelijke (Y) variabele te voorspellen
* Coefficients: Sig. van de onafhankelijke (X) variabele = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* (Constant )B = β0, dus de waarde van Y waarop de regressielijn de Y-as snijd
* (x1) B = β1, dus, de waarde die Y toeneemt, als x1, met 1 toeneemt

***6. Meervoudige lineare regressie*** *(p.423, 430)*

* Analyse > Regression > Linear
* Voeg onder “Dependent” de Afhankelijke (Y) variabele toe
* Voeg onder “Independent” de meerdere onafhankelijke (X) variabelen toe
* Klik op “OK”
* Model Summary: R Square = mate waarin Y wordt verklaard door de onafhankelijke (X)variabele (=R2\*100%)
* Anova: F = bruikbaarheid van het gehele model (y = β0 + β1x1 + β2x2 ……..+ βixi) Als F significant is (Sig. < α) dan is het model bruikbaar om de afhankelijke (Y) variabele te voorspellen
* Coefficients: Sig. van de onafhankelijke (X) variabelen = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* (Constant )B = β0, dus de waarde van Y waarop de regressielijn de Y-as snijd
* (xi) B = βi, dus, de waarde die Y toeneemt, als xi, met 1 toeneemt

***7. Z-toets (bij > 30) of T-toets (bij n <30) voor gepaarde waarnemingen*** *(p. 280)*

* Analyse > Compare Means > Paired Samples T Test
* Voeg de 2 variabelen toe waarvoor je wilt kijken of ze voor 1 populatie verschillen
* Klik op “Options” en geef het betrouwbaarheidsinterval aan (90%, 95% of 99%), klik op “Continue”
* Klik op “OK”
* Sig. van de paired sample test = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen
* Z-waarde = testwaarde

***8. Rangsommentoets Wilcoxon*** *(Bij N > 10: p. 626, bij N < 10: p. 624)*

* Analyze > Nonparametic Tests > 2 Independent Samples
* Zet in “Test Variable List” de afhankelijke variabele.
* Zet in “Grouping Variable” de onafhankelijke variabele.
* Klik op “Define Groups” en vul de groepen in. Vb. Group 1: 1, Group 2: 2. (Bijv. bij variabele geslacht man=1 vrouw=2 dan voer je dat in)
* Klik op “Continue”
* Klik op “OK”
* Asymp. Sig. = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen

***9. Kruskal-Wallis H-toets*** *(p.636)*

*\** Analyze > Nonparametric Tests > K independent Samples.

\* Vul in de Test Variable List de afhankelijke variable in.

\* Vul in het hokje bij Grouping Variable de onafhankelijke variable.

\* Klik op Define Range en vul daar het minimun en maximum in (bijv. 1 en 7 bij aantal dagen tv kijken).

\* Klik op Continue

\* Klik op OK.

\* Sig > α: H0 niet verwerpen / Sig < α: H0 verwerpen

***10. Rangtekentoets Wilcoxon*** *(Bij N > 25: p. 626, bij N < 25: p. 624)*

* Analyze > Nonparametic Tests > 2 Related Samples
* Zet in “Test Pairs” beide (onafhankelijke) variabelen.
* Bij “Test Type” moet je “Wilcoxon” aanvinken.
* Klik op “OK”
* Asymp. Sig. = P, dus: Sig > α: H0 niet verwerpen / Sig < α: H0 wel verwerpen

***11. Friedman******toets*** *(p. 854) (te gebruiken bij 3 of meer variabelen)*

\* Analyze > Nonparametric Tests > K Related Samples.

\* Klik één voor één een variabele aan in het linkerhok.

\* Breng steeds één variabele van dit hok naar het rechterhok (Test Variables).

\* Klik op OK

\* Sig > α: H0 niet verwerpen / Sig < α: H0 verwerpen

**Hoe maak je een Dummy? (zowel via syntax als via het menu)**

**Het maken van een dummy (VB: variabele OPL3CAT heeft de categorieën laag=1, midden=2, en hoog=3)**

* Transform > Compute Variable
* Klik op “If” > Vink “Include if case satifies condition” aan en haal variabele OPL3CAT naar rechts.
* Typ achter OPL3CAT “=1”, je krijgt dus “OPL3CAT=1”
* Klik op “Continue”
* Vul bij “Numeric Expression” het getal 1 in.
* Vul bij “Target Variable” de naam van de nieuwe variabele in. Bijv: “dumopl1” (dit is dan de dummy alleen voor de categorie laag van de variabele OPL3CAT)
* Klik op “OK”
* Klik weer op Transform > Compute Variable en klik even op “Reset” om alle gegevens weg te halen.
* Klik op “If” > Vink “Include if case satifies condition” aan en haal nu de variabele “dumopl1” naar rechts. (als je die zo genoemd hebt)
* Zorg ervoor dat dit in het vak staat: “MISSING(dumopl1)”
* Klik op “Continue”
* Vul bij “Numeric Expression” het getal 0 in.
* Vul bij “Target Variable” de naam van de dummy in. In dit geval: “dumopl1”
* Klik op “OK”
* SPSS vraagt: “Change existing variable?”, klik hierbij op “OK”
* LET OP! Nu is alleen nog de dummy aangemaakt voor de categorie “laag” van de variabele OPL3CAT. Doe dit ook voor de variabelen “midden” en “hoog”

**Het maken van een dummy via de syntax (dit gaat sneller, maar misschien moeilijker)**

**(VB: variabele OPL3CAT heeft de categorieën laag=1, midden=2, en hoog=3)**

* Typ LETTERLIJK in de syntax het volgende: (let ook op het gebruik van spaties en zo)

COMPUTE dumopl1= 0 .

If (OPL3CAT=1) dumopl1=1.

* Doe dit voor alle drie de categorieën, dus je krijgt uiteindelijk in totaal dit:

COMPUTE dumopl1= 0 .

If (OPL3CAT=1) dumopl1=1.

COMPUTE dumopl2= 0 .

If (OPL3CAT=2) dumopl2=1.

COMPUTE dumopl3= 0 .

If (OPL3CAT=3) dumopl3=1.

* Selecteer dit wat je net getypt hebt in de syntax en klik op het play tekentje.
* Nu heb je wel voor alle drie de categorieën een dummy aangemaakt.

**Stappenplan:**   
  
Stap 1: bepaal het aantal variabelen. Wanneer er meer dan 2 variabelen zijn gaan je in principe meteen naar meervoudige lineaire regressie. Bij 1 variabele kijk je in de tabel ‘1 variabele’ en diens juiste meetniveau. Bij 2 variabelen kijk je bij de tabel ‘2 variabel’ en diens juist meetniveaus.   
  
Stap 2: bepaal juiste meetniveau(s):

*Kwalitatieve gegevens:*   
**Nominaal**: zonder rangorde : geslacht ; burgerlijke staat; regio waarin je woont ; respondentnummer ; lid kerk etc.   
**ordinaal**: met rangorde (alles in categorieën!): opleidingsniveau ; sterren restaurant ; ook bijv. een leeftijd die gecategoriseerd is.   
*Kwantitatieve gegevens:*   
**Interval** : willekeurig 0-punt : bijv. 0 graden, tijdstip 0, (je kunt dit niet vermenigvuldigen of delen).   
**ratio**: natuurlijk 0-punt: aantal uren per week, leeftijd zonder categorieën, 0 minuten, kinderaantal, 0 gram, 0 personen, (je kunt hierbij wel vermenigvuldigen en delen) 🡪 als je 0 kunt zeggen

Stap 3: kijk naar het schema en zoek de geschikte statistische toets op  
  
Stap 4: formuleer de nul- en alternatieve hypothese. Let op: is het eenzijdig of tweezijdig (ofwel wordt er richting aangegeven in de toets of niet? richting aangeven wordt vaak gedaan d.m.v. woorden als hoger, lager, kleiner, minder etc. Eenzijdige toets is bijv. : verwacht wordt dat mannen minder vaak naar de kapper gaan dan vrouwen)

Stap 5: bepaal de alpha. Hiermee bepaal je de zekerheid van je conclusie. Let op: je mag alpha niet door 2 delen als je van een 2-zijdige toets naar een eenzijdige toets gaat. Je deelt SIG ofwel de overschrijdingskans door 2 maar niet alpha.   
  
Stap 6: voer je toets uit in spss. Alle mogelijke toetsen en hoe je dit uitvoert in spss staan hierboven.

Stap 7 : trek je conclusie :   
- sig/p/overschrijdingskans is kleiner dan alpha dan verwerp je H0 *wel*. SIg < Alpha 🡪 H0 verwerpen  
- sig/p/overschrijdingskans groter dan alpha dan verwerp je H0 *niet*. SIG > alpha 🡪 H0 NIET verwerpen  
  
(SIG < Alpha 🡪 model is significant en H0 verwerpen  
SIG > Alpha 🡪 model is niet significant en H0 niet verwerpen.)

**Keuze maken uit meerdere mogelijke toetsen:**  
Je hebt de meetniveaus vastgesteld, maar er staan vaak meerdere mogelijke toetsen in de tabel. Om vast te stellen welke toets je gebruikt kijk je naar woorden als *relatie, samenhang, gemiddelde, afhankelijk, onafhankelijk, sterkte verband etc.*  Hier wordt duidelijk welke toets je gebruikt :

1. Onderzoeken of er een relatie bestaat tussen variabele A en B  
   🡪 **KRUISTABEL (**Chi2 en Cramer’s V)   
   Dit is tevens de enige toets waarbij je ne nulhypothese niet wil verwerpen!  
   *H0: de twee variabelen zijn onafhankelijk  
   Ha: de twee variabelen zijn afhankelijk*
2. Onderzoeken of de waarde van een bepaald steekproefgemiddelde afwijkt van een van tevoren verwacht theoretisch gemiddelde (µ) onder de nulhypothese  
   🡪 **T-TOETS OP ÉÉN GEMIDDELDE (ONE-SAMPLE T-TEST)**    
   *H0: µ = …Ha: µ ≠ …*
3. Onderzoeken of de gemiddelden van twee groepen van elkaar afwijken  
   1) Vergelijking van twee onafhankelijke groepen  
    🡪 **INDEPENDENT SAMPLES T-TEST**    
   *H0: µ1- µ2 = 0Ha: µ1- µ2 ≠ 0*  
   2) Vergelijking van twee afhankelijke groepen  
    *🡪* ***PAIRED SAMPLES T-TEST*** *H0: µD = 0Ha: µD ≠ 0*
4. Onderzoeken of twee kwantitatieve variabelen verband met elkaar houden en kijken hoe sterk dit verband is  
   1) één afhankelijke en één onafhankelijke variabele  
    🡪 **ENKELVOUDIGE LINEAIRE REGRESSIE ANALYSE**    
   *H0: B1 = 0Ha: B1 ≠ 0*  
   2) één afhankelijke en méér dan één onafhankelijke variabelen  
    🡪**MULTIPELE LINEAIRE REGRESSIE ANALYSE**  (zelfde principe als enkelvoudige regressie analyse)  
   *H0: B1 = 0 H0: B2 = 0Ha: B1 ≠ 0 Ha: B2 ≠ 0*
5. Onderzoeken of de gemiddelden van méér dan twee onafhankelijke groepen van elkaar afwijken  
    🡪 **VARIANTIE-ANALYSE (ANalysis Of VAriance)**   
   *H0: B1= B2 = 0Ha: Tenminste één B-waarde ≠ 0*

**Wanneer gebruik je (non-) parametrische toets**  
Wanneer gebruik je een non-parametrische (niet normaal verdeeld) en de parametrische (normaal verdeelde) toets?   
🡪 hier uitleg over de verschillende soorten toetsen en welke je wanneer moet gebruiken.   
  
*1. Numerieke gegevens : twee ongepaarde groepen (vergelijken gemiddelden of medianen)*   
**Soorten toetsen**:   
**2 steekproeven T- toets (parametrisch):**   
Wanneer te gebruiken:   
- wanneer de groepen groot zijn (meer dan 30observaties per groep, dus wanneer N>30);  
- standaardeviatie van de variabelen die tussen de groepen vergeleken wordt is in beide groepen even groot (vuistregel: STD in groep 1 is niet 4 maal zo groot als in groep 2)   
is dit het geval dan kun je het best deze toets gebruiken.  
**Mann-whitney test (wilcoxon rank sum test.(non parametrisch))**In principe wanneer N<30.

1. *Numerieke gegevens: meer dan twee ongepaarde groepen*   
   **parametrische variantie analyse (ANOVA)**   
   🡪 wanneer de uitkomstvariabele bij benadering normaal verdeeld is en dat de spreiding in de verschillende groepen gelijk is.   
   **non-parametrische Kruskall-wallis**
2. Numerieke gegevens : twee gepaarde groepen  
   **Paired t-toets (parametrisch)**  
   🡪 hiervoor gelden dezelfde voorwaarden als bij punt 1 besproken zijn bij de t-toets.  
   **Wilcoxon signed rank toets (non-parametrisch)**

🡪 ***Rangsommentoets Wilcoxon*** *wordt gebruikt wanneer N>10  
🡪* ***Rangtekentoets Wilcoxon*** *wordt gebruikt wanneer N>30*

* *Deze is volgens mij niet compleet, aanvullingen zijn welkom ;)*

**Nog wat leuke weetjes (zowel van de opdrachten van statistiek B als A)**

* Paired houdt in dat je verschillende gegevens gaat vergelijken van dezelfde respondenten;
* Std. Error of the mean = standaardafwijking van de steekproefverdeling;
* Mean difference = mean1-mean2 🡪 als men er bijv. van uitgaat dat mannen meer werken dan vrouwen, dan moet dit getal positief zijn wanneer mannen mean1 zijn en vrouwen mean2’
* Sig < alpha 🡪 significant;
* Sig > alpha niet significant;
* Meer dan 2 variabelen is altijd multiple linear regression, diens algemene formule is:   
  Y=B0 + B1 \* X1 + B2 \* X2 + B3 \* X3 + E (error);
* Verklaringskracht van een model is R2;
* Wanneer R2 groter is dan 0.2 is er sprake van een well fit, is het lager dan is er een low fit;
* F-toets 🡪 bruikbaarheid van het model;
* Wanneer SIG < Alpha is het model bruikbaar (bij f toets);
* Nulhypothese bij f-toets: geen enkele onafhankelijke variabele heeft invloed op de afhankelijke.;
* Je berekent de f waarde door:

(sum of squares regression / df1) / (sum of squares residual / df2) ;

* De standaarderror bij de B waarde geeft de spreiding aan die de B waarde heeft van de afhankelijke tov de gemiddelde b waarde;
* Friedman toets gebruik je wanneer meer dan 2 variabelen van ordinaal niveau zij;
* Afhankelijke variabele moet altijd in de rij, onafhankelijke variabele altijd in de kolom;
* Standaard error bereken: Standaarddeviatie / wortel N
* Wanneer bij variantieanalyse de H0 verworpen wordt, kunnen we kijken tussen welke categorieën van de onafhankelijke variabele er verschillen zijn. Dit doe je met de zogenaamde post hoc test.
* Wanneer Sig < alpha dan is het heterogeen en pak je je Games-Howell
* Wanneer dit niet zo is, dan is het heterogeen. Tukey pak je dan wanneer N even groot is, Bonferroni wanneer dit niet het geval is;
* In SPSS : One way anova 🡪 options 🡪 homog. Of variances test.

**Spss met syntax** (samenvatting)  
*Manfred te Grotenhuis en Chris Visscher*  
  
1. **Werken met SPSS-Syntax**  
 1.6 voorbeeld van syntaxbestanden   
typ in syntax bestand:   
GET FILE ‘’c:/data/hoofdstukk1.sav’’.   
DISPLAY DICTIONARY   
🡪 hierna krijg een outputvenster met een beschrijving van alle variabelen.   
  
*Wat hier wordt besproken*:  
-ophalen van bestand dat gebruikt wordt:  
GET FILE ‘’c:/data/hoofdstuk1.sav’’.   
- opvragen van informatie over variabelen   
DISPLAY DICTIONARY.   
-aanmaken van leeftijd (steekproef in dit geval uit 1995)   
DESRCRIPTIVES LEEFTIJD GEBJAAR  
 /STATISTICS MEAN STDDEV MIN MAX.   
-opvragen van frequentieverdeling van geslacht   
FREQUENCIES GESLACHT.   
  
2. **ALGEMENE VORM VAN COMMANDOS**   
2.1 een syntax moet je ALTIJD met een punt afsluiten, anders werkt deze niet.   
  
GET FILE ''c:/data/hoofdstuk2.sav''.  
FREQUENCIES  
 /VARIABLES OPLEIDIN  
 /FORMAT DVALUE  
 /STATISTICS MEAN MEDIAN  
🡪 het bestand wordt hier geopend. Vervolgens voor de variabelen opleiding een frequentieverdeling opgevraagd. De format dvalue regel geeft aan dat het in oplopende volgorde moet worden weergegeven. Laatste regel geeft aan dat ook het gemiddelde en de mediaan van de variabele moeten worden weergegeven.   
  
🡪 bij meerdere subcommando’s dient met de volgorde aan te houden waarin ze in de uitleg behandeld worden. Er zit geen verschil in wel of geen hoofdletters gebruiken. Er mogen geen witregels (enters) tussen binnen een commando plaatsvinden. Dit leidt tot gedeeltelijke uitvoering.   
  
2.3 strings bevatten vaal niet meer dan een stukje tekst. Zo wordt het ook behandeld in spss. Strings worden tussen aanhalingstekens geplaatst   
VARIABLE LABELS OPLEIDIN  
 ‘’Opleidingsnivea van de respondent’’

2.4 Aritmische expressies bevatten berekeningen. Je kunt in deze paragraaf opzoeken hoe je die berekeningen kunt plaatsen (aanhalingstekens etc.). vrij simpel verder. Bijv. ‘LEEFTID - 18 -10 geeft aan dat de leeftijd met 1.8 verminderd wordt.   
  
2.5 logische expressies lijkt op een aritmische alleen is hier het eindresultaat niet een uitkomst van een berekening maar van één van de toestanden waar dan wel onwaar. Bijv. ‘LEEFTIJD >18 is waar voor iedereen van 19 en ouder maar onwaar voor de rest.   
Neem bijv. SELECT IF.   
SELECT IF LEEFTIJD < 25.   
  
woorden als NOT en AND kunnen gebruikt worden om een bepaalde populatie wel dan wel niet aan te geven.   
  
2.6 het is mogelijk om tekst toe te voegen die niks doet binnen het syntax bestand. Dit doe je door deze tekst tussen twee sterren \* te plaatsen.   
  
**3. Bewerken van gegevens**  
hierbinnen worden verschillende commando’s besproken in op zich zelfstaande paragrafen.   
3.2 get wordt gebruikt voor het opvragen van bestanden. Met het subcommando KEEP kan worden aangegeven welke variabelen uit het bestand geopend moeten worden. Achter KEEP komt een lijst met deze variabelen.   
DROP wordt gebruikt wanneer een variabele genegeerd moet worden.   
RENAME wordt gebruikt om variabelen een andere naam te geven binnen het bestand   
het is handig om hier EXECUTE. Achter te zetten zodat het ook op de juiste manier wordt uitgevoerd.  
  
 3.3. Het commando SAVE wordt gebruikt om gegevens op te slaan. Vaak nuttig naar hercodering of andere bewerking van data. Handig is niet opslaan over het originele bestand om eventuele fouten te herzien.   
SAVE OUTFILE komt overeen met GET FILE.   
  
3.5 met VARIABLE LABELS kunnen variabelen uitgebreide omschrijvingen ofwel labels krijgen. Hierna volgen de namen van een of meer variabelen.   
  
3.6 bij VALUE LABELS wordt het label niet aan een variabele maar aan de categorie van de variabele toegekend. Vooral handig bij ordinaal/nominaal.   
  
3.7 MISSING VALUES worden ongeldige categorieën aangegeven. (bijv. weet niet/wil niet zeggen uitsluiten bij analyse). Voorbeeld : MISSING VAUES STEMMEN (14 15 16)   
  
3.8 DISPLAY : mogelijk om informatie over variabelen weer te geven.   
  
3.9 RECODE wordt gebruikt om variabelen te hercoderen.   
voorbeeld: RECODE STEMMEN (1 5 12 = 1) (2 4 9 10 = 2) (3 6 THRU 8 11 = 3) (ELSE = 999) INTO POLITIEK  
🡪 hierin worden de partijen in die groepen (links centrum rechts) verdeeld. De andere partijen, wil niet zeggen etc. worden gehercodeerd in 999. De nieuwe variabele is POLITIEK.   
  
3.10 COMPUTE = construeren variabele op basis van een formule.   
bijvoorbeeld: COMPUTE UITHUIS = (1995 – GEBJAAR) – ZELFST.   
ZELFST is wanneer respondent ouderlijk huis verliet. Gedeelte tussen haakjes berekent leeftijd respondent. Uitkomst is aantal jaren dat de respondent niet meer thuis woont bij de ouders.   
  
3.11 COUNT = telt voor iedere respondent het aantal keren dat een bepaalde waarde voorkomst in een lijst van variabelen.   
bijvoorbeeld: COUNT MISSINGS = BURG ZELFST (MISSING) URENWERK (999)   
🡪 aantal ongeldige scores die de respondent heeft op de variabelen.   
  
ontbrekende values zijn SYSMIS en ongeldige scores MISSING VALUES  
  
3.12 IF= om een (nieuwe) variabele waarden toe te kennen als aan bepaalde voorwaarden voldaan wordt.   
Bijvoorbeeld : COMPUTE AFNAME = 0   
IF (KERKNU < KER15) AFNAME =1  
🡪 variabele wordt aangemaakt die het afnemen van kerkbezoek aangeeft. Variabele respondenten wordt eerst op 0 gezet, betekent dat ze niet minder kerk zijn gaan bezoeken. Door de IF krijgen alle respondenten die ten tijde van het interview minder naar de kerk gingen dan op hun vijftiende de waarde 1.   
  
3.13 WRITE = wordt ook gebruikt om gegevens op te slaan. Write slaat echter op in een meer algemeen leesbaar tekstbestand.

3.14 DATA LIST = een data bestand dat bijv. gemaakt is met WRITE kan niet met GET geopend worden en moet dus op zo’n dergelijke manier geopend worden.   
  
3.15 DO IF/ELSE IF/ELSE/END IF= do if markeert het begin van commando die allen uitgevoerd worden als aan de voorwaarde achter do if wordt voldaan. Er kunnen hierbij allerlei databewerkingen uitgevoerd worden.   
🡪 meer info in boek.   
  
3.16 SELECT IF = gebruikt om respondenten permanent uit het gegevensbestand te verwijderen.   
Voorbeeld: SELECT IF (GESLACHT = 1).   
EXECUTE.   
🡪 dit zorgt ervoor dat het bijv. alleen betrekking heeft op de mannen in het bestand.   
  
3.17 FILTER = tijdelijk een selectie van respondenten te maken. Bij select if worden ze permanent uitgesloten bij filter dus niet! er is hierbij geen execute nodig.   
voorbeeld :   
COMPUTE MAN = (GESLACHT = 1).   
FILTER BY MAN.  
FREQUENCIES  
 /VARIABLES OPLEIDIN   
FILTER OFF.   
🡪 compute creëert een variabele (MAN) die voor mannen 1 is en voor vrouwen 0. Vervolgens wordt op die variabele gefilterd waardoor alleen nog mannelijke respondenten actief zijn (dus niet volledig uitgeschakeld!). hierna volgen een of meerdere commando’s die allen voor mannen dan ook uitgevoerd worden. Filter wordt aan het eind uitgeschakeld door FILTER OFF. Uitkomst is wel gelijk aan die van 3,16 alleen dan niet permanent zoals als 10 maal gezegd is.   
  
3.18 SORT CASES = hiermee kunnen respondenten op een of meerdere variabelen gesorteerd worden. Er is geen execute nodig.   
Voorbeeld: SORT CASES BY REGIO (D) PROVINCI.   
🡪 hiermee worden respondenten op regio en provincie gesorteerd waarbij de regiocodes in afnemende volgorde en de provinciecodes in toenemende volgorde komen te staan. De D staat dus voor afname en dit moet achter de variabele geplaatst worden.   
  
3.19 SPLIT FILE = splitst de respondenten in een aantal groepen die vervolgens apart geanalyseerd worden.   
voorbeeld:   
SORT VASES BY GESLACHT.  
SPLIT FILE BY GESLACHT.  
FREQUENCIES KERKLID.   
SPLIT FILE OFF.   
🡪 eerst wordt gesorteerd op geslacht (mannen 1 en vrouwen 2). Hierna wordt bestand opgesplitst en wordt voor zowel mannen als vrouwen afzonderlijk de frequentieverdeling kerklid gemaakt. Split off file staat voor het ongedaan maken.   
  
3.20 TEMPORARY = wordt gebruikt om het daarop volgende commando slechts tijdelijk uit te voeren. Het wordt ongedaan gemaakt nadat een tweede commando uitgevoerd wordt. Het ka gebruikt worden met missing value, recode, compute, count, select if, filter, split file.   
Voorbeeld:   
TEMPORARY.  
SELECT IF (REGIO = 2).   
FREQUENCIES GESLACHT.   
FREQUENCIES GESLACHT.   
🡪 frequencies wordt alleen uitgevoerd voor respondenten uit regio 2. Commando’s hierna gelden weer voor alle respondenten.   
  
3.21 MATCH FILES = wordt gebruikt om variabelen uit verschillende databestanden te koppelen. Kan op 2 manieren:   
1. Bestanden kunnen dezelfde analyse eenheden (respondenten bijv.) bevatten, maar dan met andere variabelen 🡪 bestanden worden naast elkaar gezet.  
2. Een of meer van de te koppelende bestanden een afwijkende analyse eenheid.   
voorbeeld :   
SORT CASES BY PROVINCE  
MATCH FILES FILE\*  
 /TABLE ‘’C:/data/provincies.sav’’  
 /BY PROVINCI.  
EXECUTE  
🡪individuele gegevens uit het geopende bestand de variabele onkerk toegevoegd. Achter table is aangegeven waar dit bestand te vinden is. het bestand is al gesorteerd op variabele provinci, het openstaande bestand wordt hier eerst met het commando sort cases gesorteerd. Nieuwe gekoppelde gegevens zichtbaar na execute.  
🡪 GROOT VOORBEELD HIERVAN TE VINDEN IN HET BOEK, DOORLEZEN!

3.22 ADD FILES = wordt gebruikt om respondenten uit verschillende bestanden bij elkaar te voegen. Resultaat is nieuw databestand waarin de oorspronkelijke bestanden onder elkaar komen te staan. De waarden uit variabelen met zelfde naam komen onder elkaar te staan waarbij labels en aanduiding van ongeldige waarden uit het eerste bestand gehaald worden.   
voorbeeld:   
ADD FILES FILE ‘’c:/data/hoofdstuk3.sav’’  
/FILE ‘’c:/data/hoofdstuk3add.sav’’  
/FILE ‘’c:/data/hoofdstuk3add2.sav’’  
EXECUTE.   
🡪 20 respondententen uit de onderste 2 bestanden worden aan het bovenste bestand toegevoegd. Het totaal aantal respondenten stijgt nu dus tot 1994.   
  
3.23 DO REPEAT/END REPEAT = wordt gebruikt om een stel berekeningen meerdere keren uit te voeren op verschillende variabelen of met verschillende getallen. Het bespaart veel tijd bij eventuele hercoderingen waarbij voor veel variabelen steeds dezelfde bewerkingen uitgevoerd moet worden. Maken van dummies is hierbij handig.   
voorbeeld.:   
DO REPEAT DUMMY = lo lbo mavo mbo havo vwo hbo uni  
/NUMMER =1 TO 8.  
IF (OPLEIDIN = NUMMER) DUMMY = 1.  
IF (OPLEIDING <> NUMMER) DUMMY = 0.   
END REPEAT.   
🡪 er worden 8 dummies gemaakt voor de 8 opleidingscategorieën. Lees in boek voor perfecte uitleg.   
  
3.24 LOOP = lijkt op do repeat in die zin dat ook hier per regel (per respondent) een aantal commando’s steeds opnieuw uitgevoerd kan worden in een run. Het commando loop is echter flexibeler waardoor het mogelijk is om meer complexe datastructuren te maken.   
voorbeeld:   
LOOP #NUMMER = 1 TO 100.  
COMPUTE LEEFT = 3NUMMER.  
XSAVE OUTFILE = ‘’c:/data/leeftijd.sav’’  
/keep = RESPNR LEEFT.  
END LOOP.  
EXECUTE  
🡪 hier wordt 100 maal een loop doorlpen, daarbij wordt steeds leeftij (beginnend bij 1) en het respondentnummer weggeschreven met xsave. Om te zien of het bestand in orde is wordt gebruik gemaakt van get file. Elke respondent tussen 1 en 100 krijgt bij deze execute een eigen dataregel.   
🡪 uitgebreidere uitleg boek.  
  
3.25 WEIGHT = kan op 2 manieren worden gebruikt.   
1. Om onder en oververtegenwoordiging in een steekproef op te heffen   
2. Om analyses te doen met databestanden waar een regel betrekking heeft op meerdere respondenten tegelijk.   
achter weight volgt BY en de naam van de zogenaamde weegvariabele. Om de weging ongedaan te maken zet je weight off neer.   
🡪 boek voor goed voorbeeld   
  
3.26 SAMPLE =kan een aselecte trekking worden gedaan uit een databestand. De reden kan zijn dat anders het bestand te groot is voor analyse. Ook wordt dit gebruikt als een eenvoudige vorm van gevoeligheidsanalyse om de stabiliteit van verkregen uitkomsten na te gaan.   
voorbeeld:   
SAMPLE .10.  
FREQUENCIES   
/VARIABLES GESLACHT.   
🡪 hier een proportie. Kan ook 2 getallen zij waarbij de eerste het exacte aantal cases en de tweede de steekproefgrootte weergeeft.   
  
🡪 overzicht alle gebruikte syntaxen in het boek!!  
  
**4. Analyses**   
🡪 er staan veel subcommando’s in het boek aangegeven. Heb je deze nodig dan kan je ze in het boek opzoeken maar je weet in ieder geval wat de commando’s inhouden en wanneer deze toe te passen zijn.   
  
4.2 DESCRIPTIVES: berekent bepaalde statistische kenmerken van een variabele, neem gemiddelde, mediaan en standaarddeviatie. Commando kan ook gebruikt worden om variabelen te standaardiseren via de z-transformatie: gem= 0 standaardafwijking = 1  
voorbeeld:  
DESCRIPTIVES GEBJAAR KERKVERL ZELFST  
 /STATISTICS MEAN VARIANCE RANGE.  
🡪 uitgebreid voorbeeld in het boek.   
  
4.3 FREQUENCIES : gebruikt om frequentieverdelingen weer te geven. Statistische maten als gem., standaardafwijking en percentielen kunnen opgevraagd worden maar ook grafieken.   
voorbeeld:   
FREQUENCIES GESLACHT OPLEIDIN.  
🡪 zelfde als bij 3.16 en 3.20 qua uitkomst. Uitgebreid voorbeeld in het boek.   
  
4.4 CROSSTABS : wordt gebruikt om kruistabellen weer te geven. Naar keuze kunne in de cellen aantallen en/of percentages weergegeven worden. Verder kan dmv samenhangsmaten gekeken worden of variabelen gerelateerd zijn aan elkaar.   
voorbeeld:   
CROSSTABS CHRISTP BY KERKLID.  
🡪 groot voorbeeld in het boek  
  
4.5 EXAMINE : wordt gebruikt om zowel numeriek als grafisch inzicht te krijgen op de verdeling van een variabele met minstens ordinaal meetniveau. Achter examine volgt een lijst met variabelen waardvoor de plots opgevraagd worden. Hierna komt subcommando plot met daarachter boxplot om een doosdiagram om te vragen.   
voorbeeld ;  
EXAMINE GEBJAAR  
 /PLOT BOXPLOT.   
  
4.6 GRAPH : wordt gebruikt om allerlei soorten grafieken en plots te produceren. Handiger om deze via menu’s te specificeren (syntax wordt anders namelijk super groot). In het boek staat uitgelegd hoe je precies zo’n grafiek moet produceren.

4.7 MEANS : wordt gebruikt om gemiddelden en andere statistische grootheden van bepaalde interval/ratio verdelingen uitgesplitst naar groepen weer te geven. Is ook mogelijk om verschillen globaal op significantie te toetsen. T-test blz. 90 voor uitgebreidere methode om groepen te vergelijken.   
achter means volgt variabele waar gem. van weergegeven moet worden. Hierachter komt woord by en daarachter de variabele die de groepen van elkaar onderscheidt. Bij nog een variabele komt er een nieuwe by bij.   
voorbeeld:   
MEANS GEBJAAR BY GSLACHT BY REGIO.   
🡪 hier worden gem., aantal cases en standaardeviaties opgevraagd van de variabele gebjaar uitgesplitst naar geslacht en regio.   
  
4.8 T-TEST: hiermee kan getoetst worden of 2 gemiddelden van elkaar verschillen. Kan gebruikt worden om te toetsen of het gem. afwijkt van een populatiegemiddelde of om te toetsen of de populatiegemiddelden van 2 groepen van elkaar afwijken. Deze groepen kunnen onafhankelijk en afhankelijk zijn.   
voorbeeld:   
(meer voorbeelden in het boek)!  
T-TEST GROUPS BURG (2 3)  
 /VARIABLES LEFTIJD.   
🡪 wordt onderzocht of de gem. leeftijd van getrouwde respondenten afwijk van respondenten die zijn gescheiden. De getallen 2 en 3 zijn de waarden die zijn toegekend aan de categorieën getrouw en gescheiden.   
  
4.9 CORRELATIONS : hiermee kan de Pearsons correlatiecoëfficiënt worden berekend inclusief het significantieniveau. Achter correlations volgt een lijst me variabelen waartussen correlatie berekend moet worden.   
voorbeeld :   
CORRELATIONS KERKVERL ZELFST.   
🡪 HIER WORDT DE CORRELATIE BEREKEND TUSSEN DE LEEFTIJD WAAROP DE RESPONDENT ZICH NIET LANGER BESCHOUWDE ALS KERKLID EN DE LEEFTIJD WAAROP DE RESPONDENT HET OUDERLIJKE HUIS VERLIET.   
🡪subcommando’s en dus meer uitleg in het boek.  
  
4.10 REGRESSION : commando waarmee (multipele) lineaire regressieanalyse uitgevoerd kan worden. Daarnaast bevat het een grote hoeveelheid extra mogelijkheden zoals het weergeven van residuen-plots en collineariteitstesten. Belangrijkste zijn in deze paragraaf te vinden.   
voor het simpelweg schatten van een regressiemodel wordt na het regression commando het subcommando dependent gegeven met erachter de afhankelijke (y). vervolgens worden dmv het subcommando method en daarachter het woord enter de onafhankelijke (x) opgegeven.   
voorbeeld:   
REGRESSION   
 /DEPENDENT RELIGIE  
 /METHOD ENTER LEEFTIJD OPL2 GESLACHT.   
🡪 hier wordt een regressiemodel geschat dat de (factor)scores (blz 106) op religie verklaart met behulp van leeftijd, opl2 en geslacht.   
🡪 regressie wordt erg uitgebreid uitgelegd in het syntax boek. Staan veel voorbeelden, leest evt. door voor meer duidelijkheid en hoe en wanneer het te gebruiken is.   
  
4.11 LOGISTISC REGRESSION : voor het uitvoeren van een logistische regressie-analyse (van toepassing indien de afhankelijke dichotoom is) wordt het commando logistic regression gebruikt.   
na logistic regression volgt de naam van de afhankelijke, deze is dichotoom wat wil zeggen er zijn slechts twee verschillende geldige waarden. In principe zijn dat de waarden 0 en 1, zo niet dan codeert spss ze zo. Hierna wordt het subcommando method met erachter enter aangegeven welke onafhankelijke variabelen opgenomen moeten worden.   
voorbeeld:   
LOGISTIC REGRESSION KERKLID  
 /METHOD ENTER LEEFTIJD GESLACHT.   
🡪in dit vb. is kerklidmaatschap (0 is geen lid, 1 is wel lid) de afhankelijke variabele en zijn leeftijd en geslacht de onafhankelijke of predicatoren.   
🡪 ook hier weer veel subcommando’s en uitgebreide uitleg, waard om door te lezen voor volledig beeld.   
🡪 belangrijk: met CONTRAST kan worden aangegeven welke variabele gedummificeerd moeten worden opgenomen.   
  
4.12 FACTOR:wordt gebruikt om factor- en componentanalyses uit te voeren (groot aantal gegevens beschrijven met een kleiner aantal relevante grootheden). Het commando wordt gevolgd door het subcommando variables met erachter de variabelen waarop de analyse moet worden uitgevoerd.  
Voorbeeld:   
FACTOR   
 /VARIABLES RELIG01TO RELIG05.  
🡪 in dit eenvoudige voorbeeld wordt een componentanalyse uitgevoerd op de onderste 2 variabelen.   
  
4.13 RELIABILITY:wordt gebruikt om bij het opstellen van een schaal te bepalen hoe goed de verschillende items in de schaal passen om de betrouwbaarheid (cronbachs a) van de schaal te berekenen.   
voorbeeld:  
RELIABILITY:  
 /VARIABLES RELIG01 TO RELIG05  
 /STATISTICS CORRELATIONS  
 /SUMMARY TOTAL.   
COMPUTE SCHAALR = RELIG01 + RELIG02 + RELIG03 + RELIG04 + RELIG05.   
DESCRIPTIVES SCHAALR.   
🡪 in dit voorbeeld wordt de cronbachs a berekend voor een schaal die bestaat uit 5 variabelen die allen verwijzen naar religiositeit. Verder wordt opgevraagd in hoeverre de cronbachs a verandert indien een item weggelaten zou worden. Met compute (bz. 41) wordt vervolgens de schaal gemaakt en opgevraagd met descriptives. Merk op dat respondenten slechts een schaalscore krijgen als de scores op alle 5 items bekend zijn. Mocht dat tot te veel uitval leiden, dan is mean een goede optie.   
  
🡪 overzicht gebruikte syntaxes in 4.14 blz 112.