De bijbel der SPSS procedures

Onze SPSS die op de computer zijt
Geheiligd zijn uw functies
Uw regressies kome,
Uw output geschiede op papier als op scherm.
Geef ons heden onze dagelijkse anova
en vergeef ons dat wij niet naar de lessen gingen
gelijk ook wij vergeven aan onze proffen.
En leid ons niet naar een onvoldoende,
maar verlos ons van een tweede zit.
Amen

Samenvatting van de SPSS procedures die deel uitmaken van de cursus Data-Analyse I.

Joris Pieters – 2006 joris.pieters@ugent.be

Opmerkingen

● Ter verduidelijking van output: Alle zaken die handig zouden kunnen zijn bij de rapportage staan vermeld. Per puntje (a.), b.),…) staat eerst vermeldt wat het is, achter de dubbele punt staat de naam van de SPSS-tabel. Na het pijltje staat in de naam van de kolom. Daarachter staat tussen vierkante haakjes de rij van de tabel.

Voorbeeld:

"e.) Significantie van het model: ANOVA → Sig. [Regression]"

Betekent:

Om te weten of in dit model de afhankelijke variabele significant wordt voorspeld door de onafhankelijke variabelen ga je kijken in de output in de tabel "ANOVA". Kijk dan op het kruispunt van de kolom "Sig." en de rij "Regression".

● We zien enkel modellen met één afhankelijke variabele. Voor modellen met meerdere afhankelijke variabelen (multivariate modellen) zie data-analyse II en toegepaste data-analyse.

1. Univariaat lineair model

1.1 Algemeen lineair model (Afhankelijke variabele op intervalniveau)

1.1.1 Enkelvoudige regressie (1 predictor op intervalniveau)

Procedure

Analyze → Regression → Linear…

Dependent:	Afhankelijke variabele
Independent(s):	Onafhankelijke variabele
[Ok]	

Output

a.) Verklaarde variantie in steekproef: Model Summary → R Square
b.) Verklaarde variantie in populatie: Model Summary → Adjusted R Square
c.) β0: Coefficients → B [Constant]
d.) β1: Coefficients → B [naam van de predictor]
e.) Significantie van het model: ANOVA → Sig. [Regression]
f.) Significantie van predictor: Coefficients → Sig. [naam van de predictor]

OPMERKING: (e) en (f) zijn hier (uiteraard) gelijk aan elkaar.

1.1.2 Meervoudige regressie (Meer dan 1 predictor op intervalniveau)

A.) Gewoon: Volledig model vergelijken met een model met geen enkele parameter

Procedure

Analyze → Regression → Linear…

Dependent:	Afhankelijke variabele
Independent(s):	Onafhankelijke variabelen
[Ok]	

Output

a.) Verklaarde variantie in steekproef: Model Summary → R Square
b.) Verklaarde variantie in populatie: Model Summary → Adjusted R Square
c.) β0: Coefficients → B [Constant]
d.) β1: Coefficients → B [naam van de eerste predictor]
e.) β…: Coefficients → B [naam van de … predictor]
f.) Significantie van het model: ANOVA → Sig. [Regression]
g.) Significantie van predictor: Coefficients → Sig. [naam van de predictor]

B.) Modelvergelijking - hiërarchische regressie: Model vergelijken met beperkter model (MODEL 2 bevat meer predictoren dan MODEL 1. MODEL 0 is het model zonder predictoren)

Procedure

Analyze → Regression → Linear…

Dependent: Afhankelijke variabele

Independent(s): Onafhankelijke variabelen MODEL 1

[Next]

Independent(s): Onafhankelijke variabelen die wel in MODEL 2 maar niet in MODEL 1 zitten.

[Statistics…]

☑ R squared change

[Continue]

[Ok]

Output

a.) Verklaarde variantie MODEL 1 (of 2) t.o.v. MODEL 0 in steekproef:
Model Summary → R Square [1 (of 2)]

b.) Verklaarde variantie MODEL 1 (of 2) t.o.v. MODEL 0 in populatie:
Model Summary → Adjusted R Square [1 (of 2)]

c.) Verklaarde variantie MODEL 2 bovenop verklaarde variantie MODEL 1:
Model Summary → R Square Change [2]

d.) Significantie van MODEL 1 t.o.v. MODEL 0:
Model Summary → Sig. F Change [1]

e.) Significantie van MODEL 2 t.o.v. MODEL 1:
Model Summary → Sig. F Change [2]

f.) $\beta 0$, $\beta 1$, β… & Significantie van de predictoren: Zie eerder.

C.) Confidentie-interval voor de impact van een predictor (β)

Procedure

Analyze → Regression → Linear…

Dependent: Afhankelijke variabele

Independent(s): Onafhankelijke variabele

[Statistics…]

☑ Confidence intervals

[Continue]

[Ok]

Output

a.) 95% Confidentie-interval voor de impact van een variabele:
Coefficients → Lower/Upper Bound [naam van de predictor]

D.) Modelvergelijking waarin ook (of enkel) categorische predictoren zitten.
OPMERKING: Indien je met categorische predictoren zit maar geen modelvergelijking moet doen gebruik dan Anova of Ancova, dan moet je niet hercoderen.

Procedure voor hercoderen

(Hier enkel effectcodering. Voorbeeld voor een variabele met 3 niveaus → 2 hulpveranderlijken)

Transform → Recode → Into Different Variables…
Input Variable -> Output Variable: [naam van de variabele]
Name: EffectEduclev1 (of iets anders, noem het zoals je wilt)
[Change]
[Old and New Values…]
Old Value: 1
New Value: -1
[Add]
Old Value: 2
New Value: 1
[Add]
Old Value: 3
New Value: 0
[Add]
[Continue]
[Ok]

Transform → Recode → Into Different Variables…
Input Variable -> Output Variable: [naam van de variabele]
Name: EffectEduclev2
[Change]
[Old and New Values…]
Old Value: 1
New Value: -1
[Add]
Old Value: 2
New Value: 0
[Add]
Old Value: 3
New Value: 1
[Add]
[Continue]
[Ok]

De modelvergelijking loopt net hetzelfde als eerder besproken, maar in plaats van de categorische variabele in te voegen bij 'Independent' voeg je daar de hulpveranderlijken in. Als je meer dan één hulpveranderlijke nodig had is het niet mogelijk om te zien of het effect van deze variabele significant is aangezien de significantie van de hulpveranderlijken getoond wordt. Oplossing: De twee of meer gebruikte hulpveranderlijken invoegen als een apart model.

E.) Interactietermen in regressie. Stel: B modereert de relatie tussen A en Z

Procedure voor niet-gestandaardiseerde methode

Analyze → Descriptive Statistics → Descriptives…
  Variable(s): A, B
  [Ok]

Transform → Compute…
  Target Variable: Dev[naam van eerste variabele] (vb: "DevA")
  Numeric Expression: [naam van eerste variabele] – [gemiddelde]
  (zie output onder 'Mean' → vb: "A – 5")
  [Ok]

Transform → Compute…
  Target Variable: Dev[naam van tweede variabele] (vb: "DevB")
  Numeric Expression: [naam van tweede variabele] – [gemiddelde]
  (zie output onder 'Mean' → vb: "B – 9")
  [Ok]

Transform → Compute…
  Target Variable: IntAB
  Numeric Expression: DevA * DevB
  [Ok]

Analyze → Regression → Linear…
  Dependent: Z
  Independent(s): DevA, DevB, IntAB
  [Ok]

Output

a.) β0, β1, β… & Significantie van de predictoren: Zie eerder.
  OPMERKING: Als het effect van de interactie significant is, dan mag je de hoofdeffecten van de variabelen NIET interpreteren!

Procedure voor gestandaardiseerde methode

Analyze → Descriptive Statistics → Descriptives…
  Variable(s): A, B, Z
  ☑ Save standardized values as variables (SPSS maakt nu 3 nieuwe variabelen aan: de standaardscores van A, B, Z → ZA, ZB, ZZ.)
  [Ok]

Transform → Compute…

Target Variable: ZIntAB

Numeric Expression: ZA * ZB

[Ok]

Analyze → Regression → Linear…

Dependent: ZZ

Independent(s): ZA, ZB, ZIntAB

[Ok]

Output

a.) β0, β1, β… & Significantie van de predictoren: Zie eerder.

OPMERKING: Als het effect van de interactie significant is, dan mag je de hoofdeffecten van de variabelen NIET interpreteren!

→ Ongeacht de niet-gestandaardiseerde of de gestandaardiseerde procedure zijn de significantiewaarden van de predictoren identiek. De β's verschillen.

→ Op basis van voorgaande een predictie doen:

Procedure voor niet-gestandaardiseerde methode

$\acute{Z}_i = b_0 + b_1{*}A_i + b_2{*}B_i + b_3(A{*}B)_i$

a.) b's: zie Coefficients → B

b.) A = Geobserveerde waarde voor eerste variabele – Gemiddelde voor die variabele (Want het gaat om deviatiescores!)

c.) B = Geobserveerde waarde voor tweede variabele – Gemiddelde voor die variabele (Want het gaat om deviatiescores!)

Procedure voor gestandaardiseerde methode

$\acute{Z}_i = \{[b_0 + b_1{*}A_i + b_2{*}B_i + b_3(A{*}B)_i] * \text{Standaarddeviatie } Z\} + \text{Gemiddelde van } Z$

a.) b's: zie Coefficients → B

b.) A = (Geobserveerde waarde – Gemiddelde) / Standaarddeviatie

c.) B = (Geobserveerde waarde – Gemiddelde) / Standaarddeviatie

F.) Meer gecompliceerde modelvergelijkingen

Procedure

Stel: Een model met 4 predictoren → 5 β's: 1 voor het intercept en 1 per predictor. $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4}$ (De predictoren noemen A, B, C, D. De afhankelijke variabele noemt Z)

Voorbeeld: 3 hypotheses. Schrijf ze zodanig dat het steeds een vergelijking ten opzichte van 0 is.

1.) $\beta_1 = 0$ (Merk op dat wanneer je enkel dit wil toetsen je ook een modelvergelijking kan doen.) [1.1.2B; p.5]

2.) $2\beta_2 = \beta_1 + \beta_3$ → $-\beta_1 + 2\beta_2 - \beta_3 = 0$

3.) $\beta_2 = \beta_3$ → $\beta_2 - \beta_3 = 0$

(Merk op dat bij hypothese 2 en 3 appelen met peren vergeleken worden: om te weten of één predictor even veel effect heeft op de afhankelijke variabele als twee andere samen (zoals in hypothese 2) of om te weten of twee predictoren een even groot effect uitoefenen op de afhankelijke variabele (zoals in hypothese 3) moet je de predictoren eerst standaardiseren. Eigenaardig genoeg gebeurt dit niet op de slides van professor De Corte.)

Geeft volgende tabel:

	β_0	β_1	β_2	β_3	β_4
Hypothese 1	0	1	0	0	0
Hypothese 2	0	-1	2	-1	0
Hypothese 3	0	0	1	-1	0

Analyze → General Linear Model → Univariate…

Dependent Variable: Afhankelijke variabele

Covariate(s): Categorische onafhankelijke variabelen

(Onafhankelijke dus NIET bij Fixed Factor(s)!!! + Zet variabelen in de volgorde die overeenkomt met de β's!!!)

[Paste]

```
UNIANOVA
 Z  WITH A B C D
 /lmatrix "Toets parameter"
 all 0 1 0 0 0 ;
 all 0 -1 2 -1 0 ;
 all 0 0 1 -1 0
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /print = parameter test(lmatrix)
 /CRITERIA = ALPHA(.05)
 /DESIGN = A B C D .
```

(Dit is volledig equivalent aan de tabel. Vergeet de punt-komma's niet!!!)

[▶]

(Merk op dat bovenstaande qua procedure een Ancova is en geen meervoudige regressie! Meer hierover in het laatste deel.)

Output

a.) Significantie contrast 1: Contrast Results → [L1: Sig.]

b.) Significantie contrast …: Contrast Results → [L…: Sig.]

c.) Significantie van de toetsen samen: Test Results → Sig.

1.1.3 Anova (Predictoren zijn categorisch)

A.) Volledige toets

Procedure

Analyze → General Linear Model → Univariate…
Dependent Variable: Afhankelijke variabele
Fixed Factor(s): Categorische onafhankelijke variabelen
[Options…]
Display Means for: Alle variabelen inclusief de interactie(s)
☑ Parameter estimates
[Continue]
[Plots…]
Horizontal Axis: De variabele met de meeste waarden
Separate Lines: De variabele met de minste waarden
[Add]
[Continue]
[Ok]

Output

a.) Verklaarde variantie: Test of Between-Subjects Effects [Onder de tabel]
b.) Significantie van de predictoren: Test of Between-Subjects Effects → Sig.
OPMERKING: Als het effect van de interactie significant is, dan kan je de hoofdeffecten van de variabelen NIET eenduidig interpreteren, wel interpretatie van de eenvoudige hoofdeffecten.
c.) Geschatte (verwachte) waarde van een celgemiddelde:
- Parameter estimates → Tel het intercept en alle waarden die voor die cel ter zake doen bij elkaar op.
Voorbeeld: Wat is de voorspelde waarde van de cel (2,1)?
Oplossing: Intercept + $\alpha_2 + \beta_1 + \gamma_{21}$
- Andere manier: kijk in de tabel van de desbetreffende interactie → Mean

Modelvergelijking: Procedure

Dit is nodig als je de significantie van 1 van de predictoren of van de interactie wil kennen:
a.) Hercodeer de onafhankelijke variabelen [1.1.2D; p.6]
b.) Maak de interactievariabelen aan. Opgelet, per interactieterm heb je (I-1)(J-1) nieuwe variabelen nodig. Stel predictor A heeft 3 niveaus, en predictor B heeft er 2 → 2 interactieniveaus (Interact1 & Interact2)

Transform → Compute…
Target Variable: Interact1
Numeric Expression: EffectA1 * EffectB
[Ok]

Transform → Compute…

Target Variable: Interact2

Numeric Expression: EffectA2 * EffectB

[Ok]

c.) Doe een modelvergelijking [1.1.2B; p.5]

- Bij Block 1 zet je alle variabelen behalve diegene waarvan je het effect wil kennen binnen het volledige model.
- De overige variabele zet je in Block 2.
- Voor iedere categorisch onafhankelijke variabelen voeg je NIET die variabele in, maar wel de gehercodeerde variabele(n)!
- Vergeet niet om de variabele(n) voor de interactie in te voegen in Block 1, of in Block 2 als je het effect van interactie afzonderlijk wil bekijken.

B.) Toetsen van contrasten

Procedure

(met A en B zijn onafhankelijke variabelen met respectievelijk 3 en 2 waarden, Z is de afhankelijke)

Voorbeeld 1: Eén contrast

$\mu_{21} = \mu_{.2}$

→ $\mu_{21} - \mu_{.2} = 0$

Probleem: $\mu_{.2}$ bestaat niet op zich, maar bestaat uit μ_{12}, μ_{22}, μ_{32}

→ $\mu_{21} - 1/3\ \mu_{12} - 1/3\ \mu_{22} - 1/3\ \mu_{32} = 0$

→ Deze 4 termen bestaan eveneens niet op zichzelf, maar bestaan uit:

$$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21}$$
$$\mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12}$$
$$\mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22}$$
$$\mu_{32} = \mu + \alpha_3 + \beta_2 + \gamma_{32}$$

Met andere woorden: α krijgt het eerste subscript, β krijgt het tweede subscript, γ krijgt beide. Als we nu de vergelijking uitschrijven krijgen we dit:

$$\mu + \alpha_2 + \beta_1 + \gamma_{21} - 1/3(\mu + \alpha_1 + \beta_2 + \gamma_{12}) - 1/3(\mu + \alpha_2 + \beta_2 + \gamma_{22}) - 1/3(\mu + \alpha_3 + \beta_2 + \gamma_{32}) = 0$$

Vereenvoudigd en in de meest logische volgorde geschreven is dit:

$$- 1/3\ \alpha_1 + 2/3\ \alpha_2 - 1/3\ \alpha_3 + \beta_1 - \beta_2 - 1/3\ \gamma_{12} + \gamma_{21} - 1/3\ \gamma_{22} - 1/3\ \gamma_{32} = 0$$

Maken we nu een tabel van deze vergelijking:

μ	α_1	α_2	α_3	β_1	β_2	γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}
0	-1/3	2/3	-1/3	1	-1	0	-1/3	1	-1/3	0	-1/3

Betreffende de γ's: De achterste index loopt het snelste op! Dit is belangrijk, want SPSS gaat er van uit dat je het zo doet.
Ter controle: tel na of de som van alle getallen in de tabel gelijk is aan 0.

Analyze → General Linear Model → Univariate…
Dependent Variable: Afhankelijke variabele
Fixed Factor(s): Categorische onafhankelijke variabelen
[Paste]

```
UNIANOVA
 Z  BY A B
 /lmatrix "Toets contrast"
 A -1/3 2/3 -1/3              (Dit zijn de waarden voor de α's in de tabel)
 B 1 -1                       (Dit zijn de waarden voor de β's in de tabel)
 A*B 0 -1/3 1 -1/3 0 -1/3     (Dit zijn de waarden voor de γ's in de tabel)
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /CRITERIA = ALPHA(.05)
 /print = test(lmatrix)
 /DESIGN = A B A*B.
```

[▶]

(Opgelet: Er wordt hier maar één contrast getoetst, er mogen hier daarom geen punt-komma's staan!)

<u>Voorbeeld 2: Twee contrasten</u>

$\mu_{21} = \mu_{.2}$ & $\mu_{.1} = \mu_{32}$

→ Contrast 1: (Zie voorbeeld 1)

$\mu_{21} = \mu_{.2}$

$\mu_{21} -1/3\ \mu_{12} -1/3\ \mu_{22} - 1/3\ \mu_{32} = 0$

$- 1/3\ \alpha_1 + 2/3\ \alpha_2 - 1/3\ \alpha_3 + \beta_1 - \beta_2 - 1/3\ \gamma_{12} + \gamma_{21} - 1/3\ \gamma_{22} - 1/3\ \gamma_{32} = 0$

→ Contrast 2:

$\mu_{.1} = \mu_{32}$

$\mu_{.1} - \mu_{32} = 0$ ($\mu_{.1}$ bestaat uit μ_{11}, μ_{21}, μ_{31})

$1/3\ \mu_{11} + 1/3\ \mu_{21} + 1/3\ \mu_{31} - \mu_{32} = 0$

$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11}$

$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21}$

$\mu_{31} = \mu + \alpha_3 + \beta_1 + \gamma_{31}$

$\mu_{32} = \mu + \alpha_3 + \beta_2 + \gamma_{32}$

$$1/3(\mu + \alpha_1 + \beta_1 + \gamma_{11}) + 1/3(\mu + \alpha_2 + \beta_1 + \gamma_{21})$$
$$+ 1/3\,(\mu + \alpha_3 + \beta_1 + \gamma_{31}) - (\mu + \alpha_3 + \beta_2 + \gamma_{32}) = 0$$

Uitgewerkt:

$$1/3\ \alpha_1 + 1/3\ \alpha_2 - 2/3\ \alpha_3 + \beta_1 - \beta_2 + 1/3\ \gamma_{11} + 1/3\ \gamma_{21} + 1/3\ \gamma_{31} - \gamma_{32} = 0$$

	μ	α_1	α_2	α_3	β_1	β_2	γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{31}	γ_{32}
Contrast1	0	-1/3	2/3	-1/3	1	-1	0	-1/3	1	-1/3	0	-1/3
Contrast2	0	1/3	1/3	-2/3	1	-1	1/3	0	1/3	0	1/3	-1

Analyze → General Linear Model → Univariate…

Dependent Variable: Afhankelijke variabele

Fixed Factor(s): Categorische onafhankelijke variabelen

[Paste]

```
UNIANOVA
  Z  BY A B
  /lmatrix "Toets contrast"
  A -1/3 2/3 -1/3
  B 1 -1                          (Contrast 1)
  A*B 0 -1/3 1 -1/3 0 -1/3 ;          (Vergeet de punt-komma niet
  A 1/3 1/3 -2/3                       tussen de contrasten!!!)
  B 1 -1                          (Contrast 2)
  A*B 1/3 0 1/3 0 1/3 -1
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /print = test(lmatrix)
  /DESIGN = A B A*B .
```

[▶]

Output

a.) Significantie contrast 1: Contrast Results → [L1: Sig.]

b.) Significantie contrast …: Contrast Results → [L…: Sig.]

c.) Significantie van de toetsen samen: Test Results → Sig.

C.) Hiërarchische opzetten (genest(eld)e factoren)

Voorbeeld: A = School, B = Provincie, C = Inrichtende macht
→ A is genesteld binnen B * C → Notatie: A(B*C)

Procedure

Analyze → General Linear Model → Univariate…
Dependent Variable: Afhankelijke variabele
Fixed Factor(s): (Niets ingeven hier!)
[Paste]

```
UNIANOVA
 Z by A B C
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /CRITERIA = ALPHA(.05)
 /design = A within B by C   B   C   B*C .
```

(Vergeet niet het puntje achter ALPHA(.05) te wissen en achter B*C te zetten!!!)

Output

a.) Het interactie-effect bevat hier A niet en A zelf staat vermeld als A(A*B) beide omwille van het feit dat A genesteld is binnen B * C.

1.1.4 Ancova (Zowel interval als categorische predictoren)

Hieronder worden twee speciale toepassingen van Ancova beschreven, gebaseerd op de cursus. Algemeen echter is Ancova gewoon die procedure waarbij je zowel categorische als continue predictoren ingeeft. Voorbeelden hiervan komen aan bod in het laatste deel. Merk op dat de procedure er gelijkaardig uitziet als de Anova-procedure. Het enige verschil is het feit dat er ook "Covariates" (continue predictoren) worden ingevoerd. Zoals eerder gezien [1.1.2D; p.6] kan je ook multipele regressie (Analyze → Regression → Linear…) gebruiken als je zowel categorische als continue predictoren hebt. Het nadeel daar is echter dat je dan je categorische predictoren moet hercoderen (omdat SPSS die niet beschouwd als zijnde categorisch). Meer duidelijkheid over welk van de twee procedures je nu best kan gebruiken in welke situatie komt aan bod in het laatste deel.

A.) Statistische controle: door toevoeging van continue predictoren ('covariaten') een voorspelling op basis van categorische predictoren statistisch gevoeliger maken.

Procedure

Analyze → General Linear Model → Univariate…
Dependent Variable: Afhankelijke variabele
Fixed Factor(s): Categorische predictor(en)
Covariate(s): Continue predictor(en)
[Ok]

Output

a.) Significantie van de predictor(en) na statistische controle voor de continue predictor(en): Test of Between-Subjects Effects → Sig. [naam van de categorische predictor(en)]

B.) Toetsing homogeniteit binnencel regressies: Heeft de regressie van de continue predictor op de afhankelijke variabele de zelfde helling in iedere cel. (Cellen: de verschillende waarden van de categorische predictor)

Procedure

Analyze → General Linear Model → Univariate…
Dependent Variable: Afhankelijke variabele
Fixed Factor(s): Categorische predictor(en)
Covariate(s): Continue predictor(en)
[Model…]
⦿ Custom
Model: [naam van de categorische predictor]
[naam van de continue predictor]
[categorische] * [continue] (selecteer hiervoor beide en druk op het pijltje)
[Continue]
[Ok]

Output

a.) Zijn de binnencelregressies homogeen:
Test of Between-Subjects Effects → Sig. [categorische] * [continue]
Indien significant ($p < 0.05$): Binnencelregressies zijn NIET homogeen en verschillen dus van elkaar.
Indien niet significant ($p > 0.05$): Binnencelregressies zijn WEL homogeen.

1.2 Veralgemeend lineair model (Afhankelijke variabele categorisch)

1.2.1 Binaire logistische regressie (Afhankelijke is categorisch)

De afhankelijke variabele is dichotoom → Kan enkel waarden 0 of 1 aannemen.

A.) Voorspellen op basis van toeval (Model 0) vs. voorspellen met predictoren (Model 1)

Opgelet: Je kan zowel categorische als continue predictoren invoeren, maar als het categorische predictoren zijn moet je ze eerst hercoderen. Uitzondering: als de categorische predictor enkel de waarden 0 en 1 heeft mag je hem zo ingeven (aangezien dit overeenkomt met Dummy-codering). [1.2.1C; p.17]

Procedure

Analyze → Regression → Binary Logistic…

Dependent: Afhankelijke variabele

Covariates: Onafhankelijke variabelen

[Options…]

☑ Iteration history

[Continue]

[Ok]

Output Block 0: Beginning Block

a.) Verlies geassocieerd met model 0: Iteration History →
-2 Log likelihood [onderste]

b.) De kans op basis van toeval op waarde 0 op de afhankelijke variabele:
Classification Table → Percentage Correct [Overal Percentage]
(De kans op waarde 1 op de afhankelijke variabele is dan uiteraard 1-… aangezien het om een dichtome afhankelijke variabele gaat)

c.) De kans op basis van toeval op waarde 0 gedeeld door de kans op basis van toeval op de waarde 1: Variables in the Equation → Exp(B)

Output Block 1: Method = Enter

a.) Verlies geassocieerd met model 1: Iteration History →
-2 Log likelihood [onderste]

b.) Likelihood Ratio: Omnibus Tests of Model Coefficients → Chi-square
(= Het verschil tussen verlies geassocieerd met model 0 en met model 1)
De significantie (Sig.) zegt ons of model 1 beter is dan model 0.
Met andere woorden: Hebben de predictoren enig nut bovenop een voorspelling op basis van toeval.

c.) Percentage correcte voorspellingen op basis van de onafhankelijke variabelen: Classification Table → Percentage Correct [Overall Percentage]

d.) Belang van de predictoren: Variables in the Equation → Exp(B)
* B kleiner dan 1: Hoe hoger de waarde op die predictor, hoe kleiner de kans op waarde 1 op de afhankelijke variabele.
* B groter dan 1: Hoe hoger de waarde op die predictor, hoe groter de Kans op waarde 1 op de afhankelijke variabele.

B.) Model vergelijken met beperkter model.

Procedure

Analyze → Regression → Binary Logistic…

Dependent: Afhankelijke variabele

Covariates: Onafhankelijke variabelen MODEL 1

[Next]

Covariates: Onafhankelijke variabelen die wel in MODEL 2 maar niet in MODEL 1 zitten.

[Options…]

☑ Iteration history

[Continue]

[Ok]

Output Block 1: Method = Enter

a.) Significantie van Model 1 ten opzichte van Model 0:

Omnibus Tests of Model Coefficients → Sig.

Output Block 2: Method = Enter

a.) Significantie van Model 2 ten opzichte van Model 1:

Omnibus Tests of Model Coefficients → Sig. [Step of Block]

b.) Significantie van Model 2 ten opzichte van Model 0:

Omnibus Tests of Model Coefficients → Sig. [Model]

C.) Categorische predictoren met meer dan twee waarden.

Je kan zowel Dummy-, Effect- en Contrastcodering gebruiken, maar aangezien sommige categorische variabelen reeds in Dummy staan (Als ze enkel de niveaus 0 en 1 hebben) wanneer je de data ingeeft is het het gemakkelijkste om ineens alles in Dummy te doen (Dummy noemt "indicator in SPSS).

Procedure

Analyze → Regression → Binary Logistic…

Dependent: Afhankelijke variabele

Covariates: Onafhankelijke variabelen (zowel interval als categorische)

[Categorial…]

Categorial Covariates: De onafhankelijke variabele die gehercodeert dient te worden

[Continue]

Voor interactietermen tussen categorische variabelen: Selecteer ze links (Ctrl ingedrukt houden) en druk: [>a*b>]

[Ok]

Output Block 1: Method = Enter

a.) Impact van de niveaus van de gehercodeerde categoriale variabele:
Variables in the Equation → Exp(B)

Opgelet: Aangezien het laatste niveau van de variabele referentieniveau is betreft de rij [naam van de variabele] op het laatste niveau. [naam van de variabele](1) betreft dan het 1e niveau van de variabele en zo verder. Exp(B) geeft telkens de impact van het niveau van de variabele weer ten opzichte van het referentieniveau. Als je ten opzichte van het eerste niveau wil vergelijken doe dan de procedure opnieuw, maar:

[Categorial…]

Selecteer de variabele in de kolom "Categorical Covariates"
⦿ First
[Change]
[Continue]

Wil je de impact van een variabele vergelijken met de gemiddelde impact, dan moet je effect codering gebruiken ("deviation" in SPSS):

[Categorial…]

Selecteer de variabele in de kolom "Categorical Covariates"
Contrast: Deviation
[Change]
[Continue]

1.2.2 Loglineaire modellen (Afhankelijke is frequentie of percentage)

Geen leerstof

1.2.3 Poisson Regressie (Afhankelijke is een 'count')

Geen leerstof

2. Bilineaire modellen

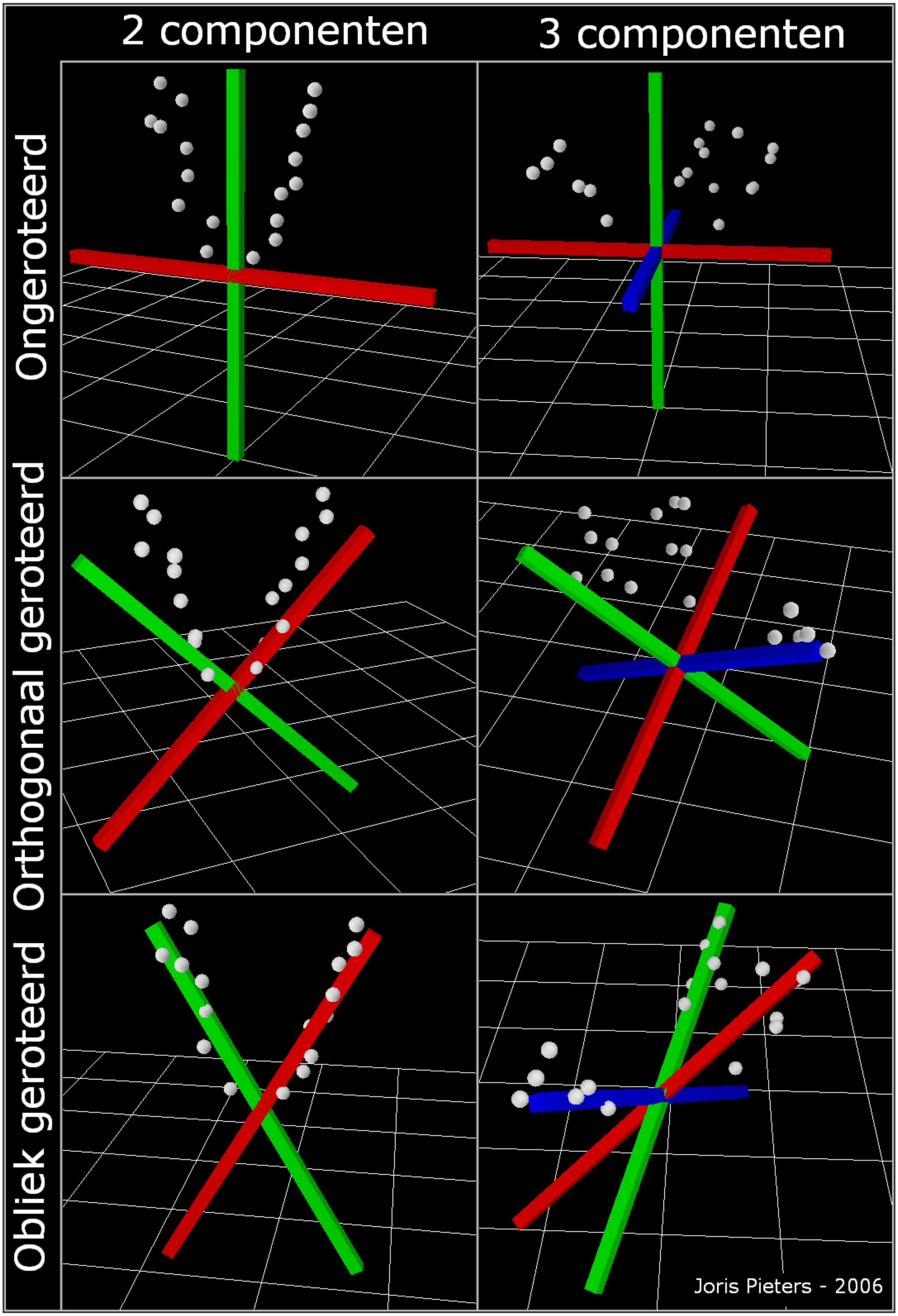
2.1 Verschil tussen PCA en FA

Zowel principale componentanalyse als factoranalyse hebben als doel een groot aantal variabelen te reduceren tot een kleiner aantal. Dit gebeurt op basis van de correlaties tussen de variabelen. Hoog gecorreleerde variabelen zullen samen één component of factor vormen. Het grote voordeel hiervan is dat je daarna een regressie kan doen met veel minder variabelen.
Principale componentanalyse ga je gebruiken wanneer je verschillende variabelen hebt die allemaal <u>éénzelfde observeerbaar construct</u> meten, bijvoorbeeld verschillende soorten metingen van gewicht ("gewicht gemeten door een balans", "gewicht gemeten met een elektronische weegschaal", "buikomtrek") kunnen samen worden beschouwd als één nieuwe variabele.
Factoranalyse daarentegen gebruik je wanneer de gecorreleerde variabelen <u>éénzelfde onderliggend construct</u> meten. Typevoorbeeld hiervan zijn vragenlijsten: de vragen "ik spreek regelmatig onbekenden aan op straat", "op feestjes ben ik altijd diegene die zorgt voor de ambiance", en "ik durf nooit een vraag te stellen tijdens de les" kunnen beschouwd worden aan variabelen onderliggend aan de factor "Extraversie". Merk op dat de laatste vraag juist het tegenovergestelde leek te meten. Dit doet totaal niet ter zake aangezien het gaat om hoge absolute correlaties (dus dicht bij -1 of dicht bij 1).
Soms is het onderscheid tussen observeerbare en onderliggende constructen misschien niet zo duidelijk, daarom een vuistregeltje: heb je fysieke variabelen gebruik PCA, heb je psychologische variabelen gebruik FA.

2.2 Rotatie

Na een PCA of FA zonder rotatie krijg je als resultaat een eerste component/factor waar alle variabelen zo hoog mogelijk op laden. Op de overige factoren laden de variabelen in afnemende mate. Een voorbeeld waarbij zo een ongeroteerde factoranalyse van pas kan komen is een IQ-test: hoewel sommigen van mening zijn dat IQ niet één construct is maar uit twee of meerdere factoren bestaat (Voorbeeld: verbaal en performaal) pleiten anderen ervoor om IQ te vatten in één cijfer; met andere woorden één factor waar alle subtesten van de IQ-test hoog op laden.
Rotatie daarentegen zorgt ervoor dat er een meer gelijke verdeling komt. Iedere component/factor zal een aantal variabelen krijgen die hoog op deze component/factor laden. We maken een onderscheid tussen orthogonale en oblieke rotaties.

- Bij orthogonale rotaties blijven de verschillende componenten/factoren loodrecht op elkaar staan. Het nadeel is dat meerdere variabelen toch nog een redelijk grote lading kunnen hebben op een andere component/factor. Het voordeel is dat de componenten/factoren ongecorreleerd zijn. Bij de grafische voorstelling zal duidelijk worden waarom ongecorreleerd ook wel "loodrecht" wordt genoemd.
- Bij oblieke rotaties roteren de componenten/factoren vrij: correlaties tussen de componenten/factoren zijn met andere woorden wel toegestaan.



2 componenten
3 componenten
Ongeroteerd
Orthogonaal geroteerd
Obliek geroteerd
Joris Pieters - 2006

Op bovenstaande tekening zie je de twee types rotaties voor twee en voor drie componenten/factoren. Een visuele weergave van een rotatie van één component/factor zit er niet bij om de eenvoudige reden dat één component/factor roteren zinloos is. Een visuele weergave van meer dan drie componenten/factoren zit er niet bij omdat dit onmogelijk is. Mocht je er toch in slagen je vier of meer componenten/factoren orthogonaal voor te stellen dan ben je waarschijnlijk psychotisch.

De keuze voor orthogonale of oblieke rotatie ligt in je hypothese. Voorbeeld: twee onderzoekers nemen een persoonlijkheidsvragenlijst af. De ene is van mening dat de vijf factoren die hij vindt (Big-V) volledig onafhankelijk zijn van elkaar. Hij zal een orthogonale rotatie gebruiken. De andere vermoedt dat er wel verbanden tussen de factoren kunnen zijn, bijvoorbeeld dat neurotische mensen minder extravert zijn. Hij zal kiezen voor een oblieke rotatie.

De verschillende rotatietechnieken in SPSS: In onderstaande procedures staat Varimax aangeduid. Afhankelijk van de onderzoeksvraag kan een andere rotatie soms meer van toepassing zijn. Voor het examen: kies Varimax tenzij er duidelijk aanwijzingen zijn dat je een oblieke rotatie moet gebruiken. In dat laatste geval gebruik je Oblimin.

- Varimax: Orthogonale rotatie. Elke factor heeft slechts enkele hoge ladingen (in absolute zin) en de overige ladingen zijn nagenoeg 0.
- Quartimax: Orthogonale rotatie. Elke variabele heeft op slechts een beperkt aantal factoren (liefs één) een hoge lading. De overige ladingen zijn nagenoeg 0.
- Equamax: Orthogonale rotatie. Combinatie van Varimax en Quartimax
- Oblimin: Oblieke rotatie
- Promax: Oblieke rotatie

Opgelet: Vergeet niet in je rapport te zetten welke rotatie je hebt gebruikt en de GEROTEERDE component/factor-scores te rapporteren en dus niet de ongeroteerde!

2.3 Principale component analyse

Procedure

Analyze → Data Reduction → Factor…

Variables: Alle items

[Extraction…]

Method: Principal components (standaard staat dit op hierop)

[Continue]

[Rotation…]

☑ Varimax

[Continue]

[Scores…]

☑ Save as variables

☑ Display factor score coefficient matrix

[Continue]

[Ok]

Output

a.) Componenten: Total Variance Explained → Total

(Het aantal componenten = het aantal getallen boven 1 in deze kolom.)

b.) Verklaarde variantie component: Total Variance Explained

→ Rotation Sums of Squared Loadings (Meest rechtse kolommen!!!)

→ % of Variance

c.) Cumulatieve verklaarde variantie: Total Variance Explained

→ Rotation Sums of Squared Loadings (Meest rechtse kolommen!!!)

→ Cumulative %

(De Cumulative % van de laatste component noemt men ook wel de 'pasmaat' van de componentenanalyse)

d.) Lading van een item op een (geroteerde) component:

(Rotated) Component Matrix →

Kolom: Nummer van de component

Rij: Naam van het item

(Opgelet: NIET de tabel 'Component Score Coefficient Matrix', ze lijken nogal sterk op elkaar!)

e.) Communaliteit van een item (= de proportie verklaarde variantie door dat item): Communalities → Extraction

(Merk op dat dit de som is van de kwadraten van de waarden per rij van de tabel 'Component Matrix'.)

f.) Gewichten waarmee de scores op de variabelen lineair gecombineerd moeten worden om tot de score op de 1^e (2^e, 3^e,…) component te komen: Component Score Coefficient Matrix (NIET '(Rotated) Component Matrix') → 1^e (2^e, 3^e,…) kolom.

Voorbeeld: Score Subject 1 op Component 1

= [(1^e kolom, 1^e rij) * score op 1^e item]

+ [(1^e kolom, 2^e rij) * score op 2^e item]

+ … + [(1^e kolom, laatste rij) * score op laatste item]

Uiteraard kan SPSS dit ook zelf berekenen: Dankzij '☑ Save as variables' heb je nu in het datascherm (dus niet bij de output) per component een extra kolom die aangeeft hoeveel een subject op die component scoort. Eén probleem: SPSS gebruikt deviatiescores. Deze zullen dus verschillen van wat je zelf uitrekent en je kan er niets mee doen.

2.4 Factoranalyse

Procedure

Analyze → Data Reduction → Factor…
Variables: Alle items
[Extraction…]
Method: Principal axis factoring (standaard staat dit op Principal components, zie voorgaande)
[Continue]
[Rotation…]
☑ Varimax
[Continue]
[Scores…]
☑ Save as variables
☑ Display factor score coefficient matrix
[Continue]
[Ok]

Output

a.) Kijk eerst of SPSS wel een zinvol resultaat heeft bereikt: Factor Matrix
→ Onderaan: … factors extracted. … iterations required.
(indien dit er niet staat is het resultaat niets waard!)
b.) Factoren: Total Variance Explained (Rechterdeel!!!) → Total
(Het aantal factoren = het aantal getallen boven 1 in deze kolom.)
c.) Verklaarde variantie factor: Total Variance Explained
→ Rotation Sums of Squared Loadings (Meest rechtse kolommen!!!)
→ % of Variance
d.) Cumulatieve verklaarde factor: Total Variance Explained
→ Rotation Sums of Squared Loadings (Meest rechtse kolommen!!!)
→ Cumulative %
(De Cumulative % van de laatste factor noemt men ook wel de 'pasmaat' van de factoranalyse)
e.) Lading van een item op een (geroteerde) factor: (Rotated) Factor Matrix →
Kolom: Nummer van de factor
Rij: Naam van het item
(Opgelet: NIET de tabel 'Factor Score Coefficient Matrix', ze lijken nogal sterk op elkaar!)

f.) Communaliteit van een item (= de proportie verklaarde variantie door dat item): Communalities → Extraction

(Merk op dat dit de som is van de kwadraten van de waarden per rij van de tabel 'Factor Matrix'.)

g.) Gewichten waarmee de scores op de variabelen lineair gecombineerd moeten worden om tot de score op de 1^{e} (2^{e}, 3^{e},…) factor te komen:

Factor Score Coefficient Matrix (NIET '(Rotated) Factor Matrix') → 1^{e} (2^{e}, 3^{e},…) kolom.

Voorbeeld: Score Subject 1 op Factor 1

= [(1^{e} kolom, 1^{e} rij) * score op 1^{e} item]

+ [(1^{e} kolom, 2^{e} rij) * score op 2^{e} item]

+ … + [(1^{e} kolom, laatste rij) * score op laatste item]

Uiteraard kan SPSS dit ook zelf berekenen: Dankzij '☑ Save as variables' heb je nu in het datascherm (dus niet bij de output) per factor een extra kolom die aangeeft hoeveel een subject op die factor scoort. Eén probleem: SPSS gebruikt deviatiescores. Deze zullen dus verschillen van wat je zelf uitrekent en je kan er niets mee doen.

3. Clusteranalyse

De bedoeling is om een beeld te krijgen van het aantal groepen.
Opmerking: De variabelen die je ingeeft moeten allemaal volgens dezelfde schaal zijn (Voorbeeld: Likertschaal variërend van 1 tot 5). Is dit niet zo dan moet je de waarden op alle variabelen standaardiseren. Als je twijfelt standaardiseer ze dan maar om zeker te zijn (een reeds gestandaardiseerde variabele standaardiseren maakt toch geen verschil).
Indien je niet weet hoeveel clusters je verwacht doe dan een hiërarchische clusteranalyse. Als je wel op voorrand weet hoeveel clusters je verwacht (bijvoorbeeld omdat dat in de opgave van het examen staat) gebruik dan K-Means clusteranalyse.

3.1 Hiërarchische clusteranalyse

Procedure

Analyze → Classify → Hierarchical Cluster…
- Variable(s): Alle variabelen die je wil betrekken
- [Plots…]
  - ☑ Dendrogram
  - [Continue]
- [Ok]

Output

a.) Bekijk het dendrogram en beslis op basis daarvan wat het meest logische aantal clusters is.

Nu je beslist hebt hoeveel clusters je wilt doe je de clusteranalyse opnieuw maar nu laat je SPSS weten dat je een nieuwe variabele wilt met een clusternummer voor ieder subject:

Procedure

Analyze → Classify → Hierarchical Cluster…
- Variable(s): Alle variabelen die je wil betrekken
- [Plots…]
  - ☑ Dendrogram
  - [Continue]
- [Save…]
  - ⦿ Single solution (maakt een nieuwe variabele aan)
  - Number of clusters: [aantal clusters dat je verwacht]
  - [Continue]
- [Ok]

Output

a.) In het datascherm is een nieuwe variabele toegevoegd met daarin het nummer van de cluster waarin het subject of het object zich bevindt.

3.2 K-Means clusteranalyse

Procedure

Analyze → Classify → K-Means Cluster…

Variable(s): Alle variabelen die je wil betrekken

Number of clusters: [aantal clusters dat je verwacht]

[Save…]

☑ Cluster membership

[Continue]

[Ok]

Output

a.) In het datascherm is een nieuwe variabele toegevoegd met daarin het nummer van de cluster waarin het subject of het object zich bevindt.

4. Uitgewerkte oefeningen

4.1 Ancova of meervoudige regressie???

Zoals eerder vermeld kan je als je zowel categorische als interval variabelen hebt als predictoren zowel meervoudige regressie [1.1.2; p.4] als Ancova [1.1.4; p.14] gebruiken. De reden dat je beide hebt in SPSS is eerder historisch. Meervoudige regressie werd oorspronkelijk enkel voor interval predictoren gebruikt terwijl Anova (Ancova zonder interval predictoren) voor het vergelijken van groepen (Een groepsnummer is een categorische predictor!) werd ontwikkeld.
Als het aan mij lag was het verleden al lang over boord gegooid en zouden we in programma's als SPSS nog maar één procedure overhouden. Helaas heb ik daar niets over te zeggen dus moeten we beide procedures kunnen gebruiken. Hieronder volgt een overzicht. Dit is wel beperkt tot wat we in de lessen hebben gezien, in feite kan je zowat alle functies met beide procedures doen.

	Meervoudige regressie	**Ancova**
SPSS-Procedure	Analyze → Regression → Linear…	Analyze → General Linear Model → Univariate…
Wat te doen met je categorische predictoren?	Moeten eerst gehercodeerd worden en dan bij *Independent(s)* gezet worden. Vergeet ook de interactietermen niet als je meer dan één categorische predictor hebt.	Zet bij *Fixed Factor(s)*
Wat te doen met je interval predictoren?	Zet bij *Independent(s)*	Zet bij *Covariate(s)*
Modelvergelijkingen	Door middel van de verschillende *Blocks* bij *Independent(s)*. Enkel "eenvoudige modelvergelijkingen". [1.1.2B; p.5]	Iets ingewikkelder, want hier gebeurt het via de *Syntax Editor*, maar wel veel meer mogelijkheden: "gecompliceerde modelvergelijkingen". [1.1.2F; p.8]
Contrasten toetsen	Niet mogelijk	[1.1.3B; p.11]

Wat kan je nu best gebruiken op het examen? Lees eerst je opgave eens goed door. Als er een meer gecompliceerde modelvergelijking of een contrasttoets nodig zal zijn zal je zeker Ancova nodig hebben. Zit er een eenvoudige modelvergelijking bij dan kan je kiezen: meervoudige regressie is gemakkelijker en biedt dus iets meer zekerheid (geen gebruik van de Syntax Editor nodig) maar je moet hercoderen.

4.2 Oefening

4.2.1 Opgave

Datafile: Voorbeeld.sav

De universiteit van Gent stelt zicht de vraag welke variabelen de punten van Data-analyse I het best voorspellen. Dit is de beschikbare data:

- [PuntenData1]: Het aantal punten dat een student op het examen Data-analyse I heeft behaald op 20.

- [PuntenStat2]: Het aantal punten dat een student op het examen Statistiek II heeft behaald op 20.

- [AttitudeData] & [AttitudeSPSS] De ingesteldheid die een student heeft ten opzichte van het vak Data-analyse I en ten opzichte van SPSS. Dit werd gemeten door middel van een vragenlijst met 5 items die beantwoord konden worden door middel van een Likertschaal met een range van 1 (helemaal niet mee eens) tot 5 (volledig mee eens). De items zijn:
  - Item 1: "Ik vind professor De Corte nen toffe"
  - Item 2: "In mijn vrije tijd bereken ik voor de lol singuliere matrixdecomposities"
  - Item 3: "De lessen van Data-analyse I waren leuk"
  - Item 4: "Iedere avond voor het slapen gaan zeg ik het SPSS-gebed op"
  - Item 5: "Ik heb de bijbel der SPSS procedures onder mijn hoofdkussen liggen"
  - Item 6: "Ik oefen iedere dag enkele uurtjes SPSS"

- [Geaardheid]: 1 = hetero; 2 = homo/lesbisch (Sorry, maar ik ben die standaardvoorbeeldjes met geslacht als predictor beu, dus hier eens iets anders.)

- [Kledingmaat]: 1 = Small; 2 = Medium; 3 = Large; 4 = Extra Large

A.) Beschouw een model met de volgende predictoren: PuntenStat2, AttitudeData, AttitudeSPSS, Geaardheid, en Kledingmaat. Welke variabelen spelen een rol in het voorspellen van de punten op Data-analyse I? In welke zin kunnen we de significante effecten interpreteren?

B.) Is er met betrekking tot de punten op Data-Analyse I een verschil tussen: (gezamenlijk beschoud)

- Dikke hetero's (Geaardheid = 1; Kledingmaat = 4)
- Homo's en lesbiennes ongeacht de kledingmaat (Geaardheid = 2)
- Mensen met een small kledingmaat ongeacht de geaardheid (Kledingmaat = 1)

C.) Stel dat je enkel de hetero's in beschouwing neemt, is er dan een verschil tussen mensen met een small of een medium enerzijds en mensen met een large en een extra large anderzijds?

D.) Kunnen de punten op Data-analyse I even goed voorspeld worden zonder de predictor punten op Statistiek II?

4.2.2 Oplossing

A.) Bekijk de datafile en je zal merken dat de variabelen AttitudeData en AttitudeSPSS niet bestaan. Wel staan er zes variabelen (Item1 t.e.m. Item6) die de scores zijn op de vragenlijst die peilde naar de ingesteldheid ten opzichte van het vak Data-analyse I en ten opzichte van SPSS. Om zes variabelen te reduceren tot twee variabelen hebben we een bilineair model nodig: Principale component analyse of factoranalyse. Aangezien een attitude geen extern observeerbaar iets is maar een onderliggende trek kiezen we voor factoranalyse. (Opgelet: op het examen kan dit veel minder expliciet vermeld staan. Als je een variabele die je nodig hebt niet vindt moet er dus een belletje gaan rinkelen.)

Analyze → Data Reduction → Factor…
Variables: Item1 t.e.m. Item6
[Extraction…]
Method: Principal axis factoring
[Continue]
[Rotation…]
☑ Varimax
[Continue]
[Ok]

We zien dat SPSS twee overhoudt. Op basis van de opgave hadden we niet anders verwacht. We hebben onze output nu wel, maar wat we eigenlijk nodig hebben is onze predictoren AttitudeData en AttitudeSPSS. Sluit daarom je output af en doe de factoranalyse opnieuw, maar doe nu dit erbij:

[Scores…]
☑ Save as variables
[Continue]

Om de tabel "Rotated Factor Matrix" die we nodig hebben gemakkelijk te kunnen interpreteren is het aan te raden het volgende er ook bij te doen:

[Options…]
☑ Suppress absolute values less than: 0,40
[Continue]

In je datascherm zie je nu twee nieuwe variabelen, namelijk FAC1_1 en FAC2_1. Op basis van de tabel "Rotated Factor Matrix" in de output (NIET de tabel "Factor Matrix" gebruiken want dat is de ongeroteerde oplossing. Ook al is die in dit voorbeeld bijna gelijk, het is fout!) Kan je zien dat op de eerste factor (FAC1_1) vooral items 4, 5, en 6 laden. Op basis van de items (zie opgave) kan je concluderen dat deze factor overeen komt met AttitudeSPSS. Analoog komt de factor FAC2_1 overeen met AttitudeData. Hernoem beide factoren. Vergeet niet in je rapport te vermelden dat je deze factoranalyse hebt gedaan. Dit was maar voorbereiding, nu kunnen we verder. Stel je nu even deze vragen:

- Wat is de afhankelijke variabele? Is deze categorisch of interval?
- Wat zijn de onafhankelijke variabelen? Welke zijn categorisch, welke interval?

Analyze → General Linear Model → Univariate…
Dependent Variable: PuntenData1
Fixed Factor(s): Geaardheid, Kledingmaat
Covariate(s): PuntenStat2, AttitudeData, AttitudeSPSS
[Options…]
Display: ☑ Parameter estimates
[Ok]

Voor de resultaten hiervan: zie rapport. Voor de vrijheidsgraden van de *F*-toets in het rapport: de eerste vrijheidsgraad is diegene die in de kolom "df" ter hoogte van de predictor staat, de tweede vrijheidsgraad is steeds die in de kolom "df" ter hoogte van "Error" staat. Vergeet ook niet om te vermelden in welke richting de significante effecten wijzen! Gebruik hiervoor kolom "B" in de tabel "Parameter Estimates". Voor diegenen die de laatste jaren op een andere planeet hebben doorgebracht: Significant zijn die predictoren met een *p*-waarde kleiner of gelijk aan 0.050.

B.) Dit is een voorbeeld van contrasten toetsen. Het eerste wat we moeten bepalen is welke van de twee categorische variabelen α en welke β is. De keuze is vrij maar wel belangrijk aangezien SPSS de eerste die je invoegt als α beschouwt en de tweede als β. Hier zullen we α gebruiken voor Geaardheid en β voor Kledingmaat. Ter recapitulatie: de achterste index van de γ's (die voor de β's staat dus) loopt het snelste op!

De opgave komt overeen met volgende vergelijking:

$\mu_{14} = \mu_{2\cdot} = \mu_{\cdot 1}$

Dit is één grote hypothese, maar om deze te toetsen in SPSS gaan we ze opsplitsen in kleinere contrasten: slechts 1 gelijkheidsteken per contrast:

$\mu_{14} = \mu_{2\cdot}$
$\mu_{14} = \mu_{\cdot 1}$

Merk op dat we ook $\mu_{2\cdot} = \mu_{\cdot 1}$ hadden kunnen invoegen; dit is echter zinloos aangezien deze vergelijking redundant is. De regel is dus: evenveel kleine contrasten als er gelijkheidstekens in het oorspronkelijke contrast zitten.

$\mu_{14} = \mu + \alpha_1 + \beta_4 + \gamma_{14}$

$\mu_{2\cdot} = 1/4\ \mu_{21} + 1/4\ \mu_{22} + 1/4\ \mu_{23} + 1/4\ \mu_{24}$

$= 1/4(\mu + \alpha_2 + \beta_1 + \gamma_{21}) + 1/4(\mu + \alpha_2 + \beta_2 + \gamma_{22})$
$+ 1/4(\mu + \alpha_2 + \beta_3 + \gamma_{23}) + 1/4(\mu + \alpha_2 + \beta_4 + \gamma_{24})$

$= \mu + \alpha_2 + 1/4\ \beta_1 + 1/4\ \beta_2 + 1/4\ \beta_3 + 1/4\ \beta_4 + 1/4\ \gamma_{21} + 1/4\ \gamma_{22} + 1/4\ \gamma_{23} + 1/4\ \gamma_{24}$

$\mu_{\cdot 1} = 1/2\ \mu_{11} + 1/2\ \mu_{21}$

$= 1/2(\mu + \alpha_1 + \beta_1 + \gamma_{11}) + 1/2(\mu + \alpha_2 + \beta_1 + \gamma_{21})$

$= \mu + 1/2\ \alpha_1 + 1/2\ \alpha_2 + \beta_1 + 1/2\ \gamma_{11} + 1/2\ \gamma_{21}$

Het eerste contrast wordt dan:
$\mu + \alpha_1 + \beta_4 + \gamma_{14} = \mu + \alpha_2 + 1/4\ \beta_1 + 1/4\ \beta_2 + 1/4\ \beta_3 + 1/4\ \beta_4$
$+ 1/4\ \gamma_{21} + 1/4\ \gamma_{22} + 1/4\ \gamma_{23} + 1/4\ \gamma_{24}$

$\alpha_1 - \alpha_2 - 1/4\ \beta_1 - 1/4\ \beta_2 - 1/4\ \beta_3 + 3/4\ \beta_4 + \gamma_{14} - 1/4\ \gamma_{21} - 1/4\ \gamma_{22} - 1/4\ \gamma_{23} - 1/4\ \gamma_{24} = 0$

En het tweede contrast:
$\mu + \alpha_1 + \beta_4 + \gamma_{14} = \mu + 1/2\ \alpha_1 + 1/2\ \alpha_2 + \beta_1 + 1/2\ \gamma_{11} + 1/2\ \gamma_{21}$

$1/2\ \alpha_1 - 1/2\ \alpha_2 - \beta_1 + \beta_4 - 1/2\ \gamma_{11} + \gamma_{14} - 1/2\ \gamma_{21} = 0$

Dit geeft deze tabel:

	μ	α_1	α_2	β_1	β_2	β_3	β_4	γ_{11}	γ_{12}	γ_{13}	γ_{14}	γ_{21}	γ_{21}	γ_{23}	γ_{24}
Contrast1	0	1	-1	-1/4	-1/4	-1/4	3/4	0	0	0	1	-1/4	-1/4	-1/4	-1/4
Contrast2	0	1/2	-1/2	-1	0	0	1	-1/2	0	0	1	-1/2	0	0	0

Analyze → General Linear Model → Univariate…
Dependent Variable: PuntenData1
Fixed Factor(s): Geaardheid, Kledingmaat (We hadden afgesproken dat Geaardheid overeen zou komen met α en Kledingmaat met β dus let er op dat Geaardheid boven Kledingmaat staat in het venster "Fixed Factor(s)"!)
Covariate(s): PuntenStat2, AttitudeData, AttitudeSPSS
[Paste]

```
UNIANOVA
  PuntenData1  BY Geaardheid Kledingmaat  WITH PuntenStat2 AttitudeData AttitudeSPSS
  /lmatrix "Toets contrast"
  Geaardheid 1 -1
  Kledingmaat -1/4 -1/4 -1/4 3/4
```

Geaardheid*Kledingmaat 0 0 0 1 -1/4 -1/4 -1/4 -1/4 ; (Punt-komma niet vergeten!!!)
Geaardheid 1/2 -1/2
Kledingmaat -1 0 0 1
Geaardheid*Kledingmaat -1/2 0 0 1 -1/2 0 0 0
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/print = test(lmatrix)
/DESIGN = PuntenStat2 AttitudeData AttitudeSPSS Geaardheid Kledingmaat
Geaardheid*Kledingmaat .

[►]

C.) Ook dit is een contrast. Wederom komt α overeen met geaardheid en β met kledingmaat.

$1/2(\mu_{11} + \mu_{12}) = 1/2(\mu_{13} + \mu_{14})$

$1/2(\mu + \alpha_1 + \beta_1 + \gamma_{11}) + 1/2(\mu + \alpha_1 + \beta_2 + \gamma_{12})$
$= 1/2(\mu + \alpha_1 + \beta_3 + \gamma_{13}) + 1/2(\mu + \alpha_1 + \beta_4 + \gamma_{14})$

$\mu + \alpha_1 + 1/2\ \beta_1 + 1/2\ \beta_2 + 1/2\ \gamma_{11} + 1/2\ \gamma_{12}$
$= \mu + \alpha_1 + 1/2\ \beta_3 + 1/2\ \beta_4 + 1/2\ \gamma_{13} + 1/2\ \gamma_{14}$

$1/2\ \beta_1 + 1/2\ \beta_2 - 1/2\ \beta_3 - 1/2\ \beta_4 + 1/2\ \gamma_{11} + 1/2\ \gamma_{12} - 1/2\ \gamma_{13} - 1/2\ \gamma_{14} = 0$

	μ	α_1	α_2	β_1	β_2	β_3	β_4	γ_{11}	γ_{12}	γ_{13}	γ_{14}	γ_{21}	γ_{21}	γ_{23}	γ_{24}
Contrast	0	0	0	1/2	1/2	-1/2	-1/2	1/2	1/2	-1/2	-1/2	0	0	0	0

Analyze → General Linear Model → Univariate…
Dependent Variable: PuntenData1
Fixed Factor(s): Geaardheid, Kledingmaat (We hadden afgesproken dat Geaardheid overeen zou komen met α en Kledingmaat met β dus let er op dat Geaardheid boven Kledingmaat staat in het venster "Fixed Factor(s)"!)
Covariate(s): PuntenStat2, AttitudeData, AttitudeSPSS
[Paste]
UNIANOVA
PuntenData1 BY Geaardheid Kledingmaat WITH PuntenStat2 AttitudeData AttitudeSPSS
/lmatrix "Toets contrast"
Geaardheid 0 0
Kledingmaat 1/2 1/2 -1/2 -1/2
Geaardheid*Kledingmaat 1/2 1/2 -1/2 -1/2 0 0 0 0
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/print = test(lmatrix)
/DESIGN = PuntenStat2 AttitudeData AttitudeSPSS Geaardheid Kledingmaat
Geaardheid*Kledingmaat .

[►]

D.) Dit komt overeen met een modelvergelijking. In het eerste model zitten alle variabelen behalve PuntenStat2, in het tweede model zit PuntenStat2 er wel bij. Als het tweede model niet significant beter is dan het eerste model betekent dit dat de punten op Data-analyse I even goed kunnen worden voorspeld zonder de predictor PuntenStat2. Als het tweede model wel significant beter is dan het eerste model betekent dit dat de punten op Statistiek II wel degelijk bijdragen tot de predictie van de punten op Data-analyse I.

Er zijn twee manieren om dit te toetsen. De eerste is via meervoudige regressie, wat betekent dat we de categorische variabelen zullen moeten hercoderen.
Maak effectcoderingsvariabelen aan voor je categorische predictoren. Dat zijn er één voor Geaardheid en drie voor Kledingmaat, we zullen ze EfGeaard, EfKleding1, EfKleding2, en EfKleding3 noemen. Vergeet zeker niet ook drie interactievariabelen aan te maken door EfGeaard te vermenigvuldigen met EfKleding1 (dat wordt dan EfInt1), EfGeaard te vermenigvuldigen met EfKleding2 (dat wordt dan EfInt2), en EfGeaard te vermenigvuldigen met EfKleding3 (dat wordt dan EfInt3).

Analyze → Regression → Linear…

Dependent: PuntenData1
Independent(s): EfGeaard, EfKleding1, EfKleding2, EfKleding3, EfInt1, EfInt2, EfInt3, AttitudeData, AttitudeSPSS
[Next]
Independent(s): PuntenStat2
[Statistics…]
☑ R squared change
[Continue]
[Ok]

De tweede manier is doormiddel van Ancova en de Syntax Editor. We moeten nu niet hercoderen.

Analyze → General Linear Model → Univariate…

Dependent Variable: PuntenData1
Fixed Factor(s): Geaardheid, Kledingmaat
Covariate(s): PuntenStat2, AttitudeData, AttitudeSPSS
[Paste]

```
UNIANOVA
 PuntenData1  BY Geaardheid Kledingmaat  WITH PuntenStat2 AttitudeData AttitudeSPSS
 /lmatrix "Toets model"
 PuntenStat2 1
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /CRITERIA = ALPHA(.05)
 /DESIGN = PuntenStat2 AttitudeData AttitudeSPSS Geaardheid Kledingmaat
                                                   Geaardheid*Kledingmaat .
```

[▶]

4.2.3 Rapport

Een factoranalyse werd uitgevoerd op de zes items van de attitudevragenlijst. Twee factoren werden weerhouden. De bekomen oplossing werd geroteerd via varimax. De cumulatieve variantie verklaard door deze twee factoren bedraagt 68.5%. De eerste factor waarop items 4 t.e.m. 6 hoog laden en daarom hernoemd werd tot AttitudeSPSS verklaart 37.8% van de variantie. De tweede factor waarop items 1 t.e.m. 3 hoog laden en daarom hernoemd werd tot AttitudeData verklaart 30.7% van de variantie.

Een covariantie-analyse werd uitgevoerd met de score voor Data-analyse I als afhankelijke variabele. De fixed factors waren de geaardheid en de kledingmaat en de covariaten waren de punten voor Statistiek 2, de attitude ten opzichte van Data-analyse en de attitude ten opzichte van SPSS.
Van de fixed factors was enkel het hoofdeffect van geaardheid significant, $F(1, 19) = 26.000$, $p = 0.000$. Homo's en lesbiennes blijken beter te presteren op Data-analyse I dan hetero's. Het interactie-effect en het hoofdeffect van kledingmaat waren niet significant, $F(3, 19) = 0.081$, $p = 0.969$ en $F(3, 19) = 0.927$, $p = 0.447$ respectievelijk.
Van de covariaten is enkel het effect van de punten op Statistiek II significant, $F(1, 19) = 54.090$, $p = 0.000$. Het effect ligt in de verwachte richting: Hoe hoger de punten op Statistiek II, hoe hoger de punten op Data-analyse I. De attitudes ten opzichte van SPSS en ten opzichte van Data-analyse bleken geen significante predictoren te zijn, $F(1, 19) = 0.191$, $p = 0.667$ en $F(1, 19) = 0.113$, $p = 0.741$.

Wat de nominale variabelen Geaardheid en Kledingmaat betreft blijkt er een significant verschil te zijn tussen Dikke hetero's, Homo's en lesbiennes ongeacht de kledingmaat, en Mensen met een small kledingmaat ongeacht de geaardheid. $F(2, 19) = 8.764$, $p = 0.002$.

Indien men enkel de hetero's in beschouwing neemt, dan blijkt er geen verschil te zijn op punten Data-analyse I tussen diegenen met een small of medium en diegenen met een large of een extra large. $F(1, 19) = 0.609$, $p = 0.445$.

De punten op Statistiek II zorgen voor een significant betere predictie van de punten op Data-analyse I bovenop de andere predictoren. $F(1, 19) = 54.090$, $p = 0.000$.