

**Techniques:**

- univariate **AN**alysis **Of VA**riance (ANOVA)
- bivariate analysis
- **Multivariate AN**alysis **Of VA**riance (MANOVA) **Describing and analysing relationships between at least two variables**
  - Factor Analysis
  - Multiple Regression Analysis
  - Multiple Discriminant Analysis
  - Conjoint Analysis
  - Cluster Analysis
  - Multidimensional Scaling

**Decision tree**

Level of measurement	Types of variables	Causality	Structure
metric (interval, ratio)	dependent	structure techniques	interdependence
non-metric (nominaal, ordinaal)	independent	functional techniques	dependence

**General approach**

- Define research problem (research questions; fundamental relations; conceptual model; general level of variables)
- Develop analysis plan (techniques; objective and design; sample size; types of variables; acquired info)
- Evaluate assumptions (techniques; conceptual; statistical)
- Estimation of the multivariate model (apply assumptions; option within techniques; overall fit of model; acceptable level of statistical criteria)
- Interpretation of the output (technique used; underlying dimensions / separate variables; answering research questions)
- Validate the multivariate model (generalizability; 2 samples; split sample)
- Conclusions and recommendations (in terms of research questions)

**Data analysis**

- Acquire data
- Rough data files
- Evaluate data files
  - Graphically
  - Errors
    - check for logic
    - define ranges
    - more than one answer for one single question
  - Missing data (missende variabelen zijn pas onbelangrijk als ze random zijn)
    - Generalize?
    - Underlying reasons? (data entry; routing; refusal or incomplete)
    - Final sample
    - Approach missing values
      - Use only complete data
      - Delete cases or variables
      - Imputation methods (estimation; metrical data; replacement methods)
  - Outliers (uitschieters)
    - Distinguishing data (cause: data entry; value range; exceptional observation)
    - not representative for population
    - Detection (describing and/or removal)
  - Statistical assumptions
    - Normal distribution (shape of data; univariate normal distribution)
    - Graphical test (histogram; normal probability plot)
    - Homoscedasticity (variable relations; equal distribution variances of variables; graphical test)
    - Linearity (correlation or association; equal distribution variances of variables)
    - Graphical test (scatter plot)
    - Statistical test (regression)



**Communalities:** zegt iets over de waarde van variabelen binnen de 5 factoren

**Validation**

- Generalizability
  - replicability of results
  - stable factor structure
- Influence on results
  - Impact of outliers

**Subsequent (vervolg) analysis** (Bij vervolganalyses mogen de variabelen niet correleren)

- Regression analysis
- Cluster analysis

**R-type factoranalysis** voegt variabelen samen

**Q-type factoranalysis** voegt objecten of respondenten samen

---

Regression Analysis

---

**Regression Analysis:** method to analyze relationships between metric dependent variable and one or more independent variables

- Explaining
- Forecasting

**Research design of Regression Analysis**

- Basic model ( $y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + e$ )
- level of measurement (dummy variables for non-metric data with L-levels => L-1 dummy variables)
- selection of dependent variable (research objective)
- selection of independent variables (concepts; influence on dependent variable)
- sample size (1:15-20)

**Statistical Assumptions of Regression Analysis**

- Normality (variabelen zijn normaal verdeeld)
- Homoscedasticity (gelijke variantie spreiding)
- Linearity (tussen afhankelijke en on-afhankelijke variabele)
- Level of measurement (metric data)
- Little multi-collinearity (onderlinge afhankelijkheid, liever niet te hoog omdat anders het unieke van de onafhankelijke variabele niet naar voren komt) Hoe hoger  $R^2$  (verklaarde variantie), hoe hoger de kans op multicollineariteit (In hoeverre is er samenhang tussen de variabelen)
- Sample size 1 variable :15-20

**Tollerantiewaarde** =  $1 - R^2$  ; **Tolerance factor** =  $1 - R^2 < 0,1$  => multicollineariteit ; **Variance Inflation Factor** =  $1 / (1 - R^2)$

**Direct Method or Enter method:** alle onafhankelijke factoren komen terug in de vergelijking (mate van invloed van alle variabelen is relevant)

**Stepwise Methoden**

- forward inclusion (variabele met meeste toegevoegde wordt toegevoegd)
- backward inclusion (f-toets) (begint met totaal model, var met minste toegevoegde waarde wordt verwijderd)
- stepwise selection (variabelen met meeste waarde toevoegen daarna verwijderen van voorgaande variabelen als ze geen toegevoegde waarde meer hebben)

**Regressie coefficienten:** aan welke knoppen draaien om wat te veranderen ( $\Delta\%$  independent vs  $\Delta\%$  dependent)

**$\beta$ -coefficient,** gestandaardiseerde regressie coefficient en ligt tussen -1 en 1 (tussen verschillende onderzoeken)

**$R^2$**  geeft de kwaliteit (verklaarde variantie) weer van het onderzoeksmodel

**$R^2$ -adjusted** is basis  $R^2$  met correctie op het aantal respondenten en onafhankelijke variabelen (vergelijken samples)

**F-test,** Is  $R^2$  significant  $> 0$ , anders stoppen

**T-toets,** kijkt de onafhankelijke variabele significant af van 0, anders variabele weglaten

$R^2 = 0,6$  is hoog (sociaal wetenschappelijk onderzoek is nooit hoger)

$R^2$  (Sig $<0,05$ ) kijkt significant af van 0, dan pas regressie analyseren

Conclusie sheet 20, factor 1 heeft de hoogste toegevoegde waarde

sheet 21 factor 3 en 5 hadden weinig bijgedragen aangezien we  $R^2$  van 58% hebben tov oorspronkelijke 59%

sheet 23 alle onafhankelijke variabelen hebben een significante toegevoegde waarde  $< 0,05$  (Dit was ook de stepmethode)

Conjoint Analysis

**Conjoint Analysis** (Waarde die respondenten hechten aan attributen toe te dichten, op levelniveau waardes toekennen (soorten vlaai-bodems))

- rating (vlaaien, belangrijkheid verschil)
- ranking (waarde toekennen)

**attribuut**=bodem

**attribuut-level**=soort bodem, soort versiering

Combinaties van attribuut-levels=combinatie van soort bodem en soort versiering

Bij elk attribuut even veel levels kiezen

Realistische combinaties

**adaptieve conjoint analysis:** Vragenlijst wordt afhankelijk van de antwoorden dynamisch aangepast (Vlaaionderzoek)

**Methode**

	<b>Traditional</b> (pen en papier)	<b>Adaptive</b>	<b>Choice based</b>
<b>Max. attributes</b>	9	30	6
<b>Level</b>	individual	individual	Aggregated
<b>Model</b>	additive	additive	additive + interaction

**Research Design of Conjoint Analysis**

- **Composition rule**
  - Additive model (voorkeuren bij elkaar optellen, statisch)
  - Interactive (voorkeuren hebben verschillende wegen)
- **Type of presentation**
  - Trade-off (eenvoudige, alleen ranking van 2 attributen)
  - full-profile (ranking en rating on a profile card)
  - pairwise comparison methode (ranking en rating)
- **Customer preference with**
  - full-profile method
  - pairwise comparison method

**Statistical Assumptions of Conjoint Analysis**

- Understanding of weight process
- Model specifications

**Overall fit**

- hierachy of data
- rating scale of data

**Interpretation**

- Individual level
- Aggregated level
- Relative interest (Het attribuut met de hoogste som van levels is het belangrijkste)

**Validation**

- Internal (appropriate composition rule)
- External (representativeness)

**Cluster Analysis**

---

Binnen een cluster zitten respondenten met dezelfde kenmerken

Clusteranalysis = segmentation; buying behavior; new product opportunities

Variabelen die onderling correleren worden in een factor samengebracht

**Clusteranalyse**

- Data simplification
- Relationship identification

**Research design**

- Outliers
  - positief ivm niche-markt
  - negatief als je ze niet kunt identificeren
- Similarity measures (groepen van gelijke kenmerken)
- Standardized data (gecorrigeerd op dezelfde schaal)
- Correlational measures
- Distance measures (gestandaardiseerde data) Mahalanobis  $D^2$
- Association measure  $\chi^2$  toets

**Statistische aannames**

- Representatieve populatie
- Multicollineariteit (hoge samenhang tussen variabelen, je hebt dan niet de verklaarde variantie van de variabelen afzonderlijk) toewijzing van variabelen kunnen onterecht worden toegewezen aan een cluster

**Clusteranalyse:**

- Hierarchische (optimaal aantal clusters)
  - agglomeratieve methode, respondenten indelen in clusters (100 respondenten = 100 clusters), vervolgens clusters samenvoegen die hoge correlaties hebben totdat we 1 cluster overhouden (boom-structuur)
    - single linkage (minimale afstand)
    - complete linkage (maximale afstand)
    - average linkage (gem afstand tussen de variabelen van clusters)
    - Ward's method (som van de kwadraten van de variabelen tussen clusters )(meest gebruikt)
    - Centroid method (afstand tussen cluster centroids)(op een na meest gebruikt)
  - Divisive methode
- Non-hierarchische (specificeer het aantal clusters en objects worden vervolgens in clusters ingedeeld)

De methode om het optimaal aantal clusters te verifiëren

- sequential threshold (starten met 1 cluster centroid)
- parallel threshold (starten met 4 cluster seeds)
- combination

Clusters zijn homogeen en de verhouding tussen clusters dient heterogeen te zijn

**Interpretation**

- based on average scores, cluster centroids
- fill up variables with relevant variables outside cluster process
- Labeling of the clusters, profiling

**Validation (generalizability)**

- Split sample
- Use of checking variables

**Agglomeratiecoëfficiënt:** afstand waarbij clusters zijn samengevoegd

Grootste toename in afstand is **hoogste % change**

**Profielen én tabellen bekijken**

**Multiple Discriminant Analysis**

---

**discriminant-analyse:** (Bij discriminantanalyse staat de specificatie van de groepen vast)

- Voorspellen groepslidmaatschap (een afhankelijke variabele wordt ingedeeld max 2 elkaar uitsluitende groepen, 2 groepen betekend 1 model)
- Inzicht in verschillen tussen onafhankelijke variabelen (Model beschrijven dat zo goed mogelijk het verschil aangeeft tussen twee groepen (nominale schaal)

Afhankelijke techniek die lijkt op regressie-analyse

Kijken naar het discriminierend effect van groepen (onafhankelijke variabelen) op de afhankelijke variabele (onderverdeling in groepen)

**Aannames:**

- normaliteit
- lineariteit
- meetniveau
- multicollineariteit niet wenselijk, omdat je niet het discriminerende effect van onafhankelijke variabelen kunt bepalen

**research design discriminant analysis**

- variables
  - dependent non-metric
  - independent metric
- sample size 1:20

**Discriminatlading** = factorlading

**2 methoden om discriminant functie te bepalen**

- Simultaan of Enter (Standaardiseren om verschillende schalen tussen vragen te vergelijken, beta)
  - alle onafhankelijke variabelen worden meegenomen
  - totale effect van variabelen in de set
  - statistische relevantie via Wilk's lambda (Hoe hoger Wilk's Lambda des te slechter is het discriminierend effect is; Laagste Wilk's lambda correspondeert met hoogste significantie)
- Stepwise (een voor een worden onafh var toegewezen, variabele met hoogste significante discriminerende effect wordt het eerst meegenomen)
  - elke keer één variabele met de hoogste discriminantie toevoegen
  - statistische significantie via Mahalanobis  $D^2$

**Overall fit**

- Calculate Z-score (grenswaarde)
- Compare Observations to Z-score
- Determine Predicting Capacity (classification; hitratio (correct geclassificeerd); Mahalanobis  $D^2$ )
- Divide respondents into groups based on cutting score (weighted average of group centroids)
  - Same group size  $Z_{ce} = Z_a + Z_b / 2$
  - different groupsize  $Z_{cu} = (N_a Z_b + N_b Z_a) / (N_a + N_b)$

**Interpretation**

- Discriminant weights
- Discriminant loadings
- Partial F-values
- Rotation (Varimax)
- Potency index (discriminating capacity)

**Canonische correlatie:** discriminierend effect (groter dan 40% ; kleiner dan 10% is zwak model)

**Validation:**

- analysis sample (calculating D-function)
- hold-out sample (using classification matrix)

hold-out sample, niet je volledige dataset gebruiken maar een aantal om je model te toetsen

Construction of group profiles (groepen goed bekijken en of je je complete set van onafhankelijke variabelen hebt meegenomen)

**Multi-Dimensional Scaling**

---

**multidimensional scaling** ofwel **perceptual mapping**

- positionering, grafische weergave, interdependence-techniek
- relatieve positie van producten of bedrijven weergeven
- positionering of segmentatie
- clusteren van producten

**Three main decisions:**

- Selection of object (nature and relevancy)
- Similarity or preference data (paarsgewijze vergelijking)
- individual or group level analysis gelijkenis (kaartjes van biersoorten, respondent maakt groepen op basis van attributen)

**research design:**

- **decompositional** methode:
  - niet uitsplitsen naar attributen, alleen totaaloordeel
  - In hoeverre prefereert de consument product a of b
  - vb parfum, wasmiddel
- **compositional** methode:
  - totaal oordeel en attributen

Vaak combinatie van beide methoden

Belangrijk bij multidimensional scaling is dat alle producten meeneemt (wel of geen witbier meenemen in bieronderzoek)

**meetniveau:** metrisch of non-metrische info metrisch maken (chi<sup>2</sup> toets)

**Overall fit:**

- Aantal **dimensies**:
  - subjectief bekijken
  - Net als bij factoranalyse screeplot maken en zo min mogelijk factoren meenemen (knik in curve)
  - %verklaarde variantie
- **stress-maat** is het verschil tussen werkelijke afstand en berekende afstand (stresswaarde 0,02 is redelijk)
- **Overall fit-index >0,60**
- **Projectie** van het ideale punt:
  - relatieve afstand tov ideale punt
  - richting tussen object en ideale punt

**Interpretatie**

- Subjectief (decompositional of compositional methode)
- Objectief (compositional methode)

Totale set van producten bekijken en representatieve populatie

**Validation:**

- generalizability (objects;consumers)
- 2 samples
- decompositional en compositional approach

**Multivariate Data Tools in Analysis**

	<b>Factoranalyse</b>	<b>Regressie analyse</b>	<b>Discriminant analyse</b>	<b>Conjunct meten</b>	<b>Cluster analyse</b>	<b>Multidimensional scaling</b>
<b>Doel</b>	- Bepaling onderliggende structuur - Datareductie	- Bepaling verband tussen 1 afh var (criterium) & > onafh var (predictoren)	- Analyse significante verschil in groepen - Classificeren in vaststaande groepen	- Inzicht krijgen in wijze waarop consum. waarde toekennen aan prod/dienst	- <u>Objecten</u> groeperen in categorie/clusters o.b.v. gemeens kenmerken/variabelen	- Analyse dimensies v/e meerdimens ruimte en rel positie van objecten hierin
<b>Toepassing</b>	- Clusteren variabelen - Structureren groepen <u>variabelen</u> - Reductie # variabelen	- Voorspellen afh var o.b.v. onafh - Vergelijken onafh var - Verklaren rel belang onafh irt afh var	Voorspell groepsleidm - segmentatie analyse - kredietwaardigheid - succes/faal kansen Inzicht onderscheidende predictoren	- New Prod Dev. - Winstkansanalyse - Conjuncte simulatie analyse - Segmentatie analyse	Segmentatie van - consumenten - bedrijven obv bijv. productattributen	- O.b.v. oordeel resp voorkeur/gelijkenis - Imago/positioneren - Beoordelings-dimensies
<b>Onderzoeks-ontwerp</b>	- metr. meetniv - steekproefomvang 5:1	- $y=a + bX...+e$ - metr. Meetniveau - st.omv. 15:1 st.omv. step 20:1	$Z=a + WX...W_nX_n$ • onafh=metrisch • afhank=non metrisch -st.omv 20:1 (theorie) -(praktijk) #onafh var =# pers kleinst groep	- Conjoint methodol: Traditioneel/adaptive/choice based - Onderzoeksstimuli: Concreet/visuele/real combi/ # attr, levels - Model: additief/interactief - dataverz:trade off/full profile/pairwise comp.	- Outliers kunnen storend zijn - Gelijkenis coëfficiënten - Correlatie maten - Afstand maten (!) - Associatie maten (ordi, nomi)	Decompositionele - Voorkeur vs gelijkenis - Bepaal posit/ # dim/ voorkeursdata in mds Vector vs punt - Composition ben /correspond anal
<b>Aannames</b>	Normaliteit Homoscedasticiteit Lineariteit Alleen vereist indien statistische toetsen	Normaliteit Homoscedasticiteit Lineariteit Meetniveau Geen multicollineariteit (corr. Matrix/tollerantie)	Normaliteit Homoscedasticiteit Lineariteit Meetniveau Geen multicollineariteit	- Inzicht in aard beoordelingsproces - Modelspecificaties	Representativiteit Geen multicollin tussen cluster variabelen gewenst	D: metr meetniv D: obj vergelijkb D: vergel percep resp D: alle obj beoord
<b>Statistisch model</b>	Structuur: KMO (> 0,6) Barlett (signi) Correl. (>= 0,3)	$R^2$ en signi. F-toets (model) T-toets (variabelen)	# functies=# groep -1 Wilks $\lambda$ (signi) gehele model		Selectie cluster formatie procedure: - Hierarch cl proc - Non-hierar proc	
<b>Overall fit</b>	Aantal factoren - Eigenwaarde >1 - a-priori - % $R^2$ - Screetest/plot	<b>Direct:</b> -Enter method <b>Stepwise:</b> -Forward inclus -Backward inclus -Stepwise select	Bepaal discr functie: - Simultaan/Stepwise Bepaal fit: - hitratio	Schat conjoint model: - rangorde - ranking	<b>Hierarc: agglom</b> Linkage/Ward/Centroide <b>Niet hier: (cl seeds)</b> Sequential tresh/ parallel tresh/ combi	Corresp: kwanti data, nomi dus posi getal, $\chi^2$ , verband rij-kol, samenhang obj-attributen
<b>Interpretatie / conclusie</b>	Factormatrix, -lading - communaliteit <i>Rotatie:</i> Herschikken info Orthogo = varimax Oblique = oblimin	Regressie coëff (b) $\beta$ - coëff, teken en grootte antwoord op doelstelling significantie model $R^2$	Model significant Discrim gewichten Discrim ladingen Hitratio Canonische correlatie <sup>2</sup>	- Indi niveau + geaggregeerd (trad/adapt) - Geaggregeerd niveau (choice based) - grootst range levels	<b>Hierarchisch:</b> # clusters obv agg. coëff  <b>Hier / non hierarch</b> O.b.v gem score profiel Labeling clusters (naam geven)	D: subjec/object in tweedimensionale ruimte
<b>Validatie</b>	Repliceren Split sample	Repliceren Split sample	Analyse sample Holdout sample\	Interne validatie Externe validatie	Split sample Inzetten controle var	D: generalis D: 2 steekproeven D: decom en comp
<b>Vervolg analyse</b>	Regressie/ Cluster an.					