

Hoorcollege week 5. 05-10-15

Analyseplan

1. Heb ik missings in mijn dataset?
2. Beschrijving dataset?
3. Betrouwbaarheidsanalyse per concept
4. Nieuwe somvariabele(n) maken?
5. Welke techniek: method the Big Three
6. H0/H1, parameters en resultaten rapporteren?

Missings

Missings zijn niet ingevulde vragen. Wat kan je hiermee doen?:

Listwise deletion: je verwijdert de gegevens uit de dataset van de deelnemer en analyseert alleen de gegevens van de overige deelnemers. Een reden om dit te doen is omdat de deelnemer wellicht de rest van de vragen ook niet heeft ingevuld of de vragenlijst niet serieus heeft ingevuld. Een nadeel is dat je veel data verliest, zoals vragen waarop ze wel geantwoord hebben. Wordt gebruikt bij een onderzoek met een grote dataset.

Pairwise deletion: dat betekent dat je de deelnemer niet helemaal verwijdert, maar alleen de punten die niet zijn ingevuld. Een voordeel is dat je geen data verliest. Een nadeel is dat je bij verschillende analyses een verschillende hoeveelheid deelnemers hebt.

Vervangen: het vervangen van de lege velden, in plaats van het lege veld vul je het gemiddelde op de schaal in. Hiermee kan je dus alle deelnemers kan gebruiken. Het nadeel is dat je bepaalt wat voor antwoord de deelnemer geeft en manipuleert hiermee dus de uitkomst.

Beschrijven dataset

Het beschrijven van de deelnemers in de dataset ofwel: wat zijn de eigenschappen van de deelnemers van het onderzoek? Dat is belangrijk om aan te kunnen geven over wie je onderzoek gaat. Mag je generaliseren of niet? Typische variabelen zijn:

- Response rate?
- Mannen/vrouwen
- Leeftijd
- Opleidingsniveau
- Etc.

Betrouwbaarheidsanalyse

Een betrouwbaarheids analyse gaat over de vraag: meten de vragen het concept? Met de betrouwbaarheids analyse kijk je achteraf of je een betrouwbare en valide maat van mijn concept? Dat doe je door middel van Cronbach's alpha als je alle vragen op interval/ratio niveau hebt gemeten.

- $\text{Alpha} < .7$ → Betrouwbaarheid niet oke:
 - kunnen er nog items uit?
 - aparte analyses voor X1, X2, etc
- $\text{Alpha} > .7$ → Betrouwbaarheid is oke:

- somvariabele kan gemaakt worden

Het uiteindelijke doel van de betrouwbaarheidsanalyse is dat je drie vragen kunt samennemen tot één somvariabele waarmee je kan rekenen. Zodat je alle analyses maar één keer hoeft te doen, i.p.v. alle drie de items apart.

Noteer: voor de opdracht hanteren we Cronbach's alpha = .6 vanwege simulatiedata.

Als er drie vragen op verschillende meetniveaus zijn gemeten is de betrouwbaarheidsanalyse wat moeilijker.

Nieuwe somvariabele(n) maken?

De somvariabele volgt uit een Cronbach's alpha die hoog genoeg is. De vraag is, mag je altijd een somvariabele maken?

De waarden van nominale data zijn geen waarden maar alleen namen van bijvoorbeeld een groep en hiermee kan je dus niet rekenen. Nominale data mag je dus niet optellen.

Ordinale data mag je ook niet tot een somvariabele maken. De schaal geeft een rangorde aan en is meer een categorisatie dan betekenisvolle getallen. Twee ordinale variabelen kan je dus niet optellen.

Interval data/quasi-interval data, hiermee mag je de Cronbach's alpha berekenen en ook tot een somvariabele maken. Hierbij is het belangrijk dat je goed moet kijken naar de schaalgrootte, om te bepalen of je ze samen kan nemen.

Als je twee interval/quasi-interval variabelen die een ongelijke schaalgrootte hebben wilt sommeren moet je deze twee schalen eerst standaardiseren. Dan kan je ze wel tot een somvariabele maken.

Welke techniek: methode the Big Three

Om de techniek te kiezen zijn er drie vragen:

1. Hoeveel variabelen betrokken?

Hierop zijn drie antwoorden mogelijk:

- Univariate analyse; hoe gemotiveerd zijn deze mensen?; Beschrijvende statistiek.
- Bivariate analyse; toetsende statistieken

2. Wat is het meetniveau van de betrokken variabele(n)?

- Nominaal: man/vrouw; laagste meetniveau
- Ordinaal schaal: nooit soms vaak altijd
- Interval of ratio-schaal; hoogste meetniveau

3. Asymmetrisch vs. symmetrisch?

Alleen van belang voor bivariate technieken

Asymmetrisch wanneer:

- Variabelen verschillend meetniveau hebben
- DV willen voorspellen uit IV ("causale" relatie)

Anders is het symmetrisch

Tabellen met verschillende technieken worden bij het tentamen gevoegd en hoef je dus niet uit je hoofd te leren.

Regressie vs. Correlatie m.b.t. intervaldata:

- Als we alleen willen vast stellen of er een lineaire relatie (samenhang) is
 - Correlaties

- Als we ook de waarde van een variabele willen voorspellen op basis van een andere variabele
 - regressie analyse

Over belang van $k = 2$ of $k > 2$

Als je kijkt naar een nominale onafhankelijke variabele en een ordinale of interval afhankelijke variabele wordt het ingewikkelder. Dan moet je ook naar k kijken, is $k > 2$?

What's wrong with t-tests?

Nothing, except ...

*If you want to **compare three or more groups** using t-tests with the usual 0.05 level of significance, you would have to compare the three groups pairwise (A to B, A to C, B to C), so the chance of getting the wrong result would be:*

$$1 - (0.95 \times 0.95 \times 0.95) = 14.3\%$$

*If you wanted to compare four or more groups, the chance of getting the wrong result would be $(0.95)^6 = 26\%$, and for five groups, 40%. **Not good, is it?** So we use ANOVA.*

Never perform multiple t-tests: Anyone on this module discovered performing multiple t-tests when they should use ANOVA will be shot!

Related vs. unrelated sample

Als je een nominale onafhankelijke variabele hebt, moet je rekening houden met de related/unrelated samples.

Related sample:

- Afhankelijke metingen tussen groepen die je wilt vergelijken
- Bijvoorbeeld: voor- en nameting bij zelfde steekproef
- Within-subjects design (binnenproefpersonendesign)

Unrelated sample:

- Onafhankelijke metingen tussen groepen
- Bijvoorbeeld: een groep met interventie en een andere groep zonder.
- Between-subjects design (tussenproefpersonendesign)

De volgende technieken worden tijdens dit vak niet behandeld en mogen ook niet gebruikt worden tijdens de opdrachten:

- Logistic regression analysis^a
- Discriminant analysis^a
- Repeated measures ANOVA^a

Testen en rapporteren

Hoe kom je tot een verhaal en hoe rapporteer je de resultaten.

Rapportage resultaten

ABCD/oma-methode

- A: Wat is het doel?

Je onderzoeksvraag

- B: Hoe heb je dit gedaan?

Wat voor toets je hebt gebruikt

- C: Wat was het resultaat?

Kwam er een significant resultaat uit en met welke de belangrijkste parameters zijn

- D: Welke conclusie?

Het antwoord op je onderzoeksvraag

Het heet een oma-methode, zodat je het duidelijk antwoord geeft op je onderzoeksvraag door middel van de voorgenoemde stappen dat door een leek goed te begrijpen is.

WEL doen:

- Zelf tabellen maken met gegevens die nodig zijn, maar het liefst getallen opnemen in de tekst
- Variabelen en categorienamen uitschrijven (mannen gaven aan meer tevreden te zijn dan vrouwen)

NIET doen:

- SPSS-tabellen en/of figuren overnemen
- Afkortingen gebruiken (groep 1 had een hoger gemiddelde op W1 dan groep 2) de lezer kent ze waarschijnlijk niet