Small groups and Questionnaires (for quality control)

useR! 2008

Lucien Lemmens

Universiteit Antwerpen

Introduction

- the problem in general
- the classical approach for large groups
- a transcription for small groups

The inquiry

- A questionnaire
  - questions with answers on a Likert scale
- The inquiry
  - item q&a
  - dimension :items around the same topic
  - inquiry: collection of almost independent dimensions
  - random ordering of items

The Questionnaire

- 12 dimensions
- 3 items per dimension

Dimension:
content of
lecture notes

Items
readability
understandable
badly written

The construction of such a questionnaire is a time consuming process

Spooren P., Mortelmans D., Denekens J..- *Student evaluation of teaching quality in higher education: development of an instrument based on 10 Likert scales*.- In: Assessment and evaluation in higher education, 32:6(2007), p. 667-679

The Likert Scale

1	very bad	a	f
2	bad	b	e
3	close on bad	c	d
4	close on good	d	c
5	good	e	b
6	very good	f	a
Value	Meaning	Positive formulation	Negative formulation

The inquiry

- An independent agency
  - objectivity
- All at once (only one session → missing data)
  - independence
- Written (Standard forms: encircling a-f per item)
  - automatic reading
- Anonymity warranted
  - no drawback

Traditional analysis

- Scores on dimensions are summarized
  - location: mean
  - scale: standard deviation
- A decision tree is build on this summary
  - more than x dimension under 3.5
  - more than x dimensions under 2
- reliability : cronbach alpha
- no control on outliers

The probability model & its inverse

- Model in words
  - multivariate hypergeometric
    - sampling a box with cards (of different colors) without replacement
  - multinomial
    - a method to put the cards into the box
  - Dirichlet
    - describing the circumstances of the choice of a card
- Bayes-rule

The probability model & its inverse for an item

- Model in formulas:
  - MH

$$p(\{n_i\} \mid \{N_i\}, I) = \frac{\prod_i C_{n_i}^{N_i}}{C_n^N} \Theta(\sum_i N_i = N) \Theta(\sum_i n_i = n)$$

  - MMH

$$p(\{N_i\}\{p_i\} \mid I) = N! \prod_i \frac{p_i^{N_i}}{N_i!} \Theta(\sum_i N_i = N) \Theta(\sum_i p_i = 1).$$

  - DMMH

$N_i = n_i + e_i$

$$p(\{e_i\}_a, \{p_i\}_a, \mid D_a, I) \propto \prod_i \frac{p_i^{n_i+\alpha_i}}{n_i!} \Theta(\sum_i n_i = n) \prod_i \frac{p_i^{e_i}}{e_i!} \Theta(\sum_i e_i = N - n)$$

The probability model & its inverse for an dimension

- Model in words:
  - item 1 posterior=DMMH
  - item 2 prior =posterior(item 1)= DMMH
  - item 3 prior =posterior(item 2)= DMMH
- DMMH belongs to the exponential family
  - updating

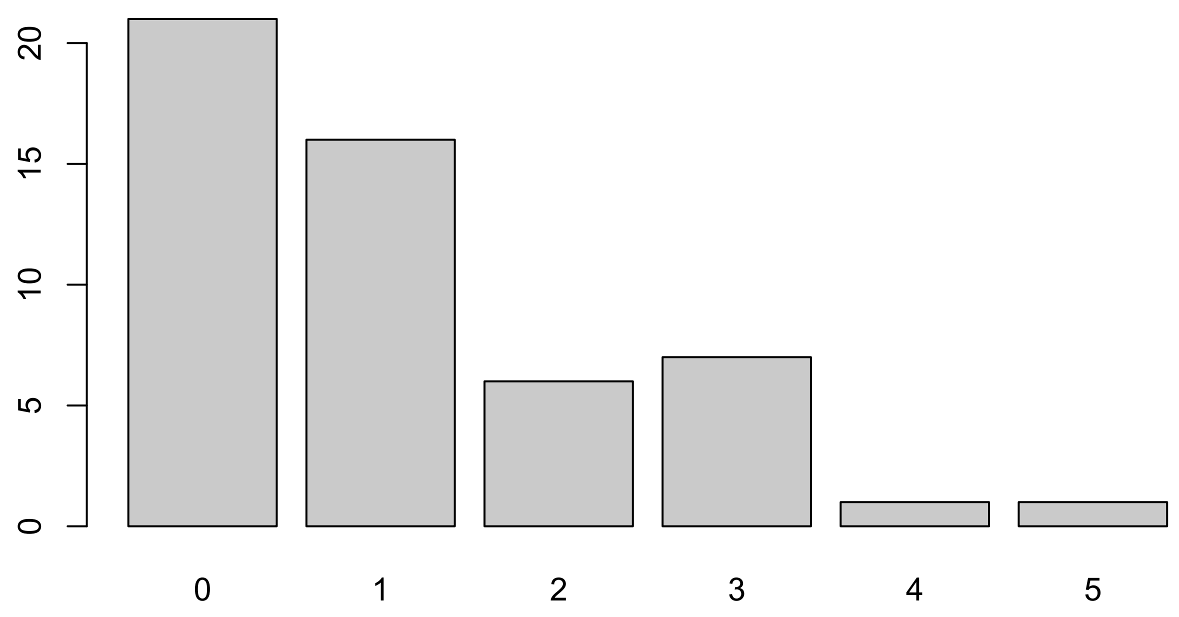
Testing the new model

- Confirmation of the analysis done for large groups from small group model
- How reliable is the model?
- How reliable are the conclusions?

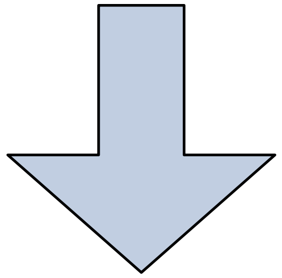
How reliable is the classical model?

- Based on the central limit theorem
  - Cronbach alpha (no direct transcription to small groups)is a measure for consistency.
- Rational argument behind this measure
  - when ranked from undesired to desired (reversing order for negatively asked questions) there is a strong correlation between items belonging to the same dimension
  - range of the ranking should be small

Range of the ranking for a dimension



a filling in at random
b interpreting a positively formulated question as negatively formulated
c filling in on position



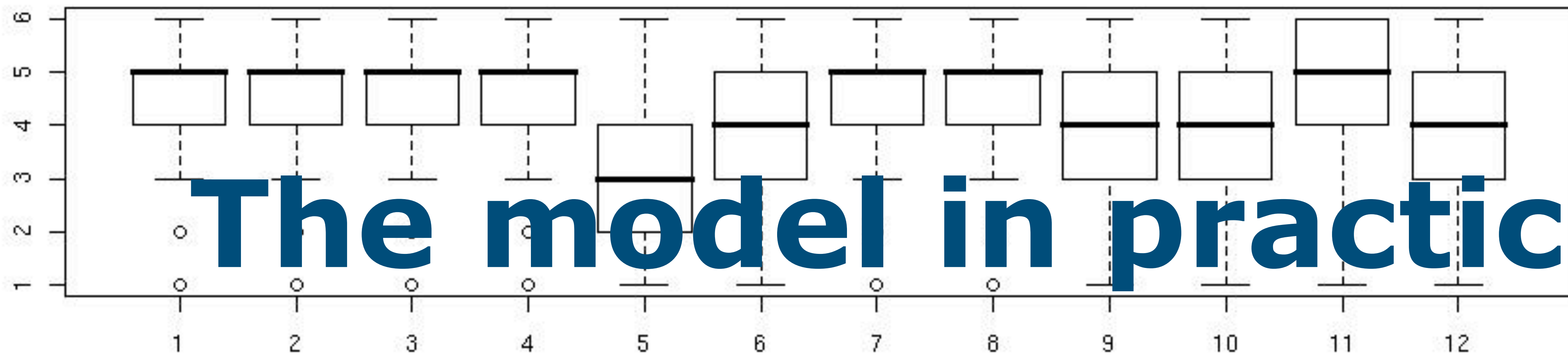
Classification of respondents

Quick & dirty

- if the range of the ordered answers in a dimension is larger than 2 then classify the dimension as non respondent
- why not 1
  - too many answers are classified as non respondent
- why not 3
  - the distinction between strongly agree and disagree a little bit should be clear

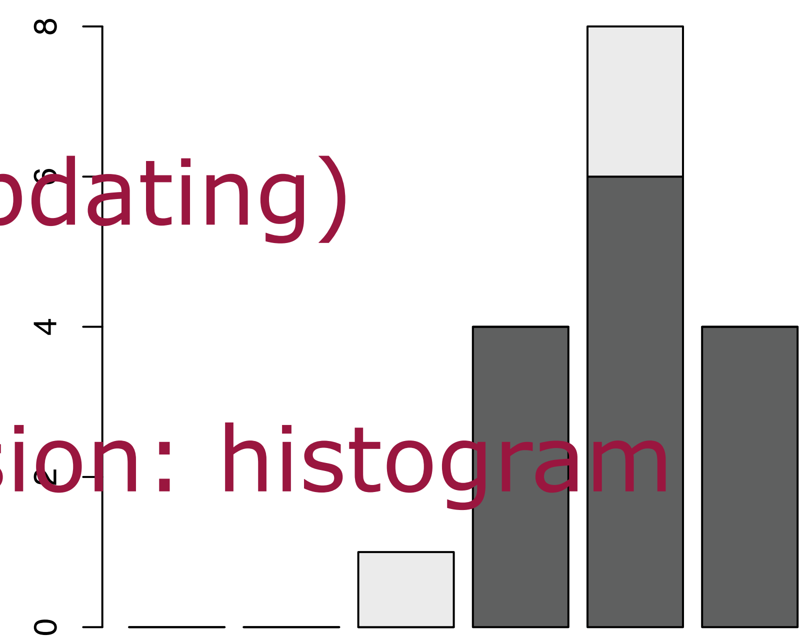
A better way to classify

- see
  - Finite Mixture and Markov Switching Models (Fruehwirth)
  - Bayesian methods for Finite Population Sampling (Ghosh & Meeden)
- adaptation to small groups is not straightforward
- going from items to dimensions is also not straightforward



The model in practice

- Determine the number of respondents for a dimension
- count n
- determine the posterior (p & e)(updating)
- calculate p(e)
- communicate this for each dimension: histogram or box and whisker plot summary



Reliability

- Simplify the statements:
  - bad---(no opinion)--- good
- Without non-respondents (no uncertainty)

$$\text{odds} = \frac{N_g}{N - N_g}$$

- With non-respondents (Odds becomes a RV)

$$N_g = n_g + e_g \qquad \text{Odds} = \frac{n_g + e_g}{N - n_g - e_g}$$

Where does R coming in ?

- Example from the faculty of science: 5 bachelor degrees: 3 years: ± 12 courses : ± 300 questionnaires
- analysis has to be automated
- only simple commands are possible
- output can be used without modifications

Automatization

Names and numbers supplied by commercial OCR software and administration

```
documenten<-c("A steekproef 8 populatie 16.csv","B steekproef 19
populatie 36.csv","C steekproef 7 populatie 15.csv","D steekproef 20
populatie 39.csv","E steekproef 5 populatie 12.csv","F steekproef 5
populatie 8.csv","G steekproef 6 populatie 8.csv","H steekproef 5
populatie 9.csv","I steekproef 5 populatie 18.csv")
aantallen<-c(16,36,15,39,12,8,8,9,18)
```

Analysis

The sequence of the questions is standard

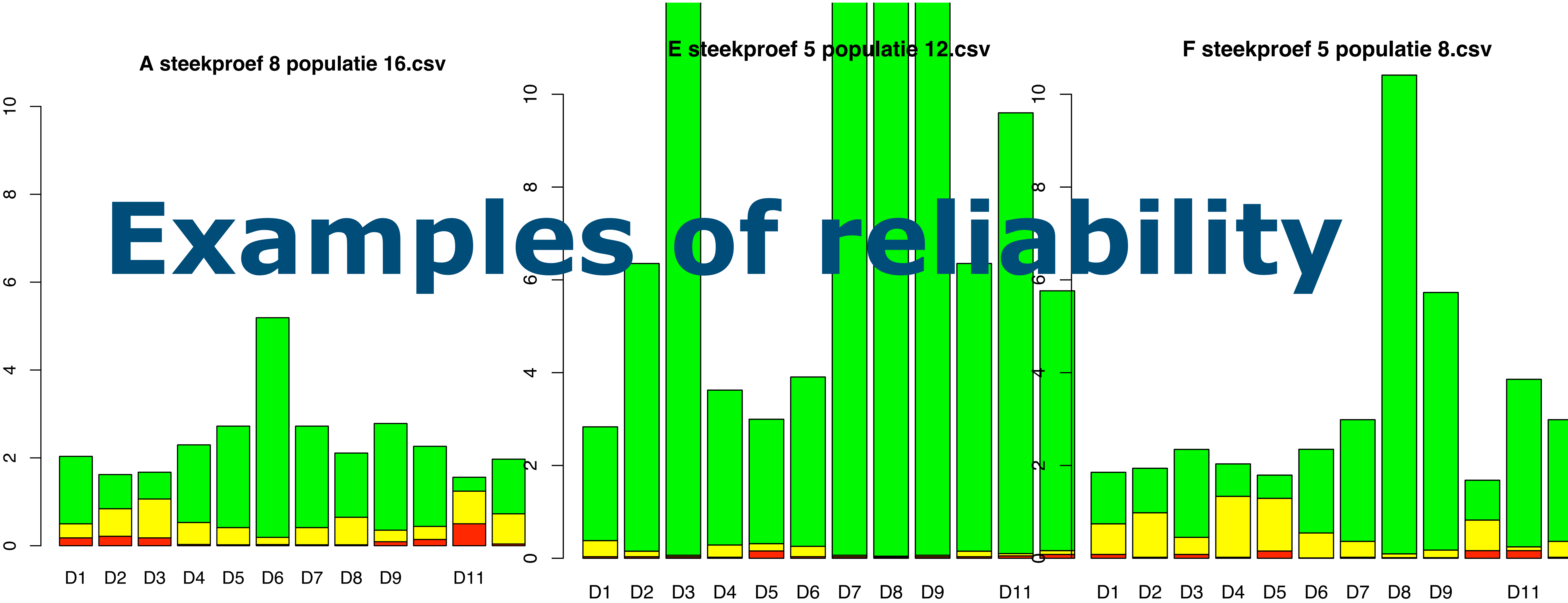
The reliability control per dimension

The figures in pdf

Comments in R on the screen

```
geg<-read.csv2(documenten[k],header=T)
 attach(geg)
 par(ask=T)
 N<-aantallen[k]
 print(doc
 DatItems<
cbind(X2A, BB,X3C,X4A,X4B,X4C,X5A,X5B,X5C,X6A,X6B,X6C,X7A,X7B,X7C,X7D,X8A,X8B,X8C,X9A,X9B,X9C,X10A,X10B,X10C,X11A,X11B,X11C,X12A,X12B,X12C,X13A,X13B,X1
 nitem<-le
 DatMatrix<-matrix(DatItems,nrow=nitem)
 itemst<-c(1,4,7,10,13,16,20,23,26,29,32,35)
 itemfn<-c(3,6,9,12,15,19,22,25,28,31,34,37)
 NOdim<-length(itemst)
 pDABC<-c()
 nDN<-c()
 require(lattice)
 for(j in 1:12){
 D2<-DatMatrix[,itemst[j]:itemfn[j]]
 ndim<-itemfn[j]-itemst[j]
 D2r<-apply(D2,1,max)-apply(D2,1,min)
 Ind<-which(D2r<=2)
 D2F<-D2[Ind,]
 D2S<-if(length(Ind)==1){median(D2F)} else {apply(D2F,1,median)}#### controle
 bpdata<-c()
 for(i in 1:6){bpdata[i]<-length(D2S[D2S==i])}
 # barplot(bpdata)
 nitem<-length(D2S)
 bpsim<-bpdata+1  ### de 1 komt van de a priori
 D2sim<-rmultinom(100,N-nitem,prob=bpsim)+bpdata
bpD2sim<-apply(D2sim,1,sum)
D2ABC<-matrix(bpD2sim,nrow=2)
 pD2ABC<-apply(D2ABC,2,sum)/sum(bpD2sim)*100
 pDABC<-c(pDABC,pD2ABC)
 nDN<-c(nDN,nitem)}
 cat("Het percentage dat tot de model A B of C behoort uit n zorgvuldige deelnemers van N studenten \n")
 OndDim<-c("D1","D2","D3","D4","D5","D6","D7","D8","D9","D10","D11","D12")
 Cat<-c("A","B","C")
 prD<-matrix(pDABC,ncol=3,byrow=T,dimnames=list(OndDim,Cat))
print(prD)
pdf(file=paste(k,".pdf",sep=""))
print(barchart(prD,col=rainbow(3),main=documenten[k]))
dev.off()
OndMax<-apply(prD,1,max)
OndOds<-OndMax/(100-OndMax)
nameMax<-function(index){if(index==1) nama<-"A" ;if(index<=2) nama<-"B" else nama<-"C";return(nama)}
print(matrix(nDN,ncol=1,dimnames=list(OndDim,c("n"))))
cat("Aantal N")
print(N)
indices<-c
for(j in 1:12
OddsInfo<-rbi
print(t(OddsI
```

Examples of reliability



No evidence 1-4
Weak evidence 4-7
Mediocre evidence 7-10
Strong evidence 10-100
Very strong evidence 100-

Discussion

- Ad hoc classification is ok for now. It was checked on large groups and it is in accordance with the construction of the questionnaire: the method should be improved for new questionnaires.
- The multi-item technique is very demanding for the author of the questions
- The Dirichlet prior is taken uniform: it contains some information (unjustified?)

Conclusions

- The expectation value of the Odds and the reference to the evidence used in model selection, gives a good indication of the reliability of the conclusion.
- After explaining the model and it consequences, it was decided to use it temporally only for feedback.
- The R-code did his job.