PDR Clothing and Item Response Theory

Analysis

Robert Chapman

12/05/2021

**Data Prep and Description**

#osf\_download(osf\_ls\_files(osf\_retrieve\_node("https://osf.io/eyn3x/"))[3,], path="data/", conflict="overwrite")  
  
source("data/Generate\_Codebook.R")  
  
#Codebook\_list is a list object with this structure:   
#"Description" - string of description of the variable  
#"Value" - vector of numeric values for the data, e.g., 1-5  
#"Labelled" - vector of labelled response options, e.g., "never", "sometimes", "always"  
  
#osf\_download(osf\_ls\_files(osf\_retrieve\_node("https://osf.io/eyn3x/"))[1,], path="data/", conflict="overwrite")  
Data=read.csv("data/engeln\_zola\_2021.csv")

1. Data Processing and Cleaning
   * Visual inspection of data for clear errors or empty rows/columns
     + *Base R: View()*

View(Data) #12/5: reviewed - no obvious issues with data

+ Checking that variables are correctly formatted   
 - \*Base R: class()\*

###Are all items and response formatted correctly - e.g., numeric v factor v value labels  
 sapply(Data, class) #need to manually review for factor and labels, etc.

## gender obcs1 obcs2 obcs3 obcs4 obcs5   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## obcs6 obcs7 obcs8 bas1 bas2 bas3   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## bas4 bas5 bas6 bas7 bas8 bas9   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## bas10 shoepain shoepainfreq itchy itchyfreq adjust   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## adjustfreq arms armsfreq walk walkfreq bend   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## bendfreq welts weltsfreq breaths breathsfreq skirts   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## skirtsfreq breasts breastsfreq hot hotfreq cold   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## coldfreq stand standfreq drag dragfreq catch   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## catchfreq obcs1r obcs2r obcs3r obcs4r obcs7r   
## "integer" "integer" "integer" "integer" "integer" "integer"   
## obcs8r bodysurv bastotal sample   
## "integer" "numeric" "numeric" "integer"

#12/5: all integer or numeric, which should be correct

+ Cross-check data variables with data dictionary  
 - \*Base R: names(Data) %in% names(Codebook)\*

###Are all items present?  
 if(any(!(names(Data) %in% names(Codebook\_list)))){  
 print(paste("names in Data, but not in Codebook:",names(Data)[which(!(names(Data) %in% names(Codebook\_list)))]))  
 print(paste("names in Codebook, but not in Data:",names(Codebook\_list)[which(!(names(Codebook\_list) %in% names(Data)))]))  
 }else{print("All names in Data and Codebook match")}

## [1] "All names in Data and Codebook match"

#12/5: "All names in Data and Codebook match"

+ Checking that variables meet expected range/values  
 - \*Base R: unique(Data[,var]) %in% Codebook[Var]$Values\*

#Out of Range values   
 sapply(names(Codebook\_list), function(Var){  
 if(!all(is.na(Codebook\_list[[Var]]$Value))){  
 unique(Data[,Var])[!(unique(Data[,Var]) %in% Codebook\_list[[Var]]$Value)]}})

## $gender  
## integer(0)  
##   
## $obcs1  
## [1] NA  
##   
## $obcs2  
## [1] NA  
##   
## $obcs3  
## [1] NA  
##   
## $obcs4  
## [1] NA  
##   
## $obcs5  
## [1] NA  
##   
## $obcs6  
## [1] NA  
##   
## $obcs7  
## [1] NA  
##   
## $obcs8  
## [1] NA  
##   
## $bas1  
## [1] NA  
##   
## $bas2  
## [1] NA  
##   
## $bas3  
## [1] NA  
##   
## $bas4  
## [1] NA  
##   
## $bas5  
## [1] NA  
##   
## $bas6  
## [1] NA  
##   
## $bas7  
## [1] NA  
##   
## $bas8  
## [1] NA  
##   
## $bas9  
## [1] NA  
##   
## $bas10  
## [1] NA  
##   
## $shoepain  
## [1] NA  
##   
## $shoepainfreq  
## [1] NA 6  
##   
## $itchy  
## [1] NA  
##   
## $itchyfreq  
## [1] NA 6  
##   
## $adjust  
## [1] NA  
##   
## $adjustfreq  
## [1] 6 NA  
##   
## $arms  
## [1] NA  
##   
## $armsfreq  
## [1] NA 6  
##   
## $walk  
## [1] NA  
##   
## $walkfreq  
## [1] NA 6  
##   
## $bend  
## [1] NA  
##   
## $bendfreq  
## [1] NA 6  
##   
## $welts  
## [1] NA  
##   
## $weltsfreq  
## [1] NA 6  
##   
## $breaths  
## [1] NA  
##   
## $breathsfreq  
## [1] NA 6  
##   
## $skirts  
## [1] NA  
##   
## $skirtsfreq  
## [1] NA 6  
##   
## $breasts  
## [1] NA  
##   
## $breastsfreq  
## [1] NA 6  
##   
## $hot  
## [1] NA  
##   
## $hotfreq  
## [1] NA 6  
##   
## $cold  
## [1] NA  
##   
## $coldfreq  
## [1] NA 6  
##   
## $stand  
## integer(0)  
##   
## $standfreq  
## [1] NA 6  
##   
## $drag  
## [1] NA  
##   
## $dragfreq  
## [1] NA 6  
##   
## $catch  
## [1] NA  
##   
## $catchfreq  
## [1] NA 6  
##   
## $obcs1r  
## NULL  
##   
## $obcs2r  
## NULL  
##   
## $obcs3r  
## NULL  
##   
## $obcs4r  
## NULL  
##   
## $obcs7r  
## NULL  
##   
## $obcs8r  
## NULL  
##   
## $bodysurv  
## NULL  
##   
## $bastotal  
## NULL  
##   
## $sample  
## integer(0)

#12/5: All variables look fine, but freq items (e.g., shoepainfreq), contain 6 levels, not 5. Will alert authors, but no changes needed here- since I won't be using the frequency items.

+ Visualization  
 - \*Base R: Boxplots(), pairs()\*

#reducing for the variables used in analysis  
PDRitems=names(Data)[!grepl("obcs|freq|bas|gender|sample|bodysurv",names(Data))]  
  
# Barplots  
source("data/Outlier\_Detection.R")  
Barplots(Data[,PDRitems])  
#12/5: Done with real data  
  
#Pairs Plots  
jpeg("pairs.jpg", width=2200, height=2200)  
par(mfrow=c(length(PDRitems),length(PDRitems)))  
for(i in PDRitems){for(j in PDRitems){  
barplot(table(Data[,i], Data[,j]), main=paste(i,j))}}  
dev.off()

## png   
## 2

#12/5: Done with real data  
  
#Need to closely review all these plots and make notes!!

1. Missing Data
   * Evaluate missingness patterns/Littles MNAR test/T test comparisons
     + *‘Missing Data’ script*

source("data/Missing\_Data.R")  
  
Missing\_Patterns(Data[,PDRitems[!grepl("skirts|breasts",PDRitems)]])

## shoepain itchy adjust arms walk bend welts breaths hot cold stand drag catch  
## 1 X   
## 2 X   
## 3 X   
## 4 X   
## 5 X   
## 6 X   
## 7 X   
## 8 X   
## 9 X   
## 10 X   
## 11 X  
## 12 X X X X X X  
## Freq  
## 1 3  
## 2 2  
## 3 4  
## 4 2  
## 5 2  
## 6 4  
## 7 2  
## 8 4  
## 9 1  
## 10 1  
## 11 2  
## 12 1

Missing\_Patterns(Data[which(Data$gender==1),PDRitems])

## shoepain itchy adjust arms walk bend welts breaths skirts breasts hot cold  
## 1 X X   
## 2 X X X   
## 3 X X X   
## 4 X X X   
## 5 X X X   
## 6 X X X   
## 7 X X X   
## 8 X X X   
## 9 X X X   
## 10 X X X  
## 11 X X   
## stand drag catch Freq  
## 1 387  
## 2 1  
## 3 1  
## 4 2  
## 5 2  
## 6 1  
## 7 4  
## 8 1  
## 9 2  
## 10 1  
## 11 X 2

Missing\_Patterns(Data[which(Data$gender==2),PDRitems])

## shoepain itchy adjust arms walk bend welts breaths skirts breasts hot cold  
## 1 X   
## 2 X   
## 3 X   
## 4 X   
## 5 X   
## 6 X   
## 7 X   
## 8 X   
## 9 X X X X X  
## stand drag catch Freq  
## 1 2  
## 2 1  
## 3 2  
## 4 1  
## 5 1  
## 6 2  
## 7 2  
## 8 X 1  
## 9 X 1

#12/5: not enough missingness (~1-4 cases for each missingness pattern distributed across items) to do gender by missingness comparison- it is worth noting that 4 males are missing the 'bend' item and no females are missing that item. Maybe worth taking a look at descriptive stats by gender for that item

1. Normality
   * Skew/Kurt/QQplot
     + *‘Assumption Checking’ script*

source("Assumption\_Checking.R")  
SK=sapply(Data[,PDRitems], SkewKurtosis)  
which(abs(SK["Skew",])>2)

## named integer(0)

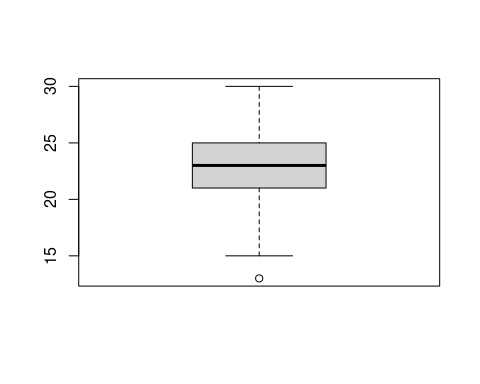
which(abs(SK["Kurtosis",])>2)

## breasts   
## 10

#12/5:only the 'breasts' items violates the kurtosis recommendations (>2), and by a hair (-2.0029). Unlikely to cause an issue, and will be retained.  
  
QQplots(Data[,PDRitems])  
#12/5: QQplots were generated, need to be reviewed!

+ Outlier detection  
 - Boxplot/zRange/plausible meanSD’s  
 + \*‘Outlier Detection’ script\*

source("data/Outlier\_Detection.R")  
  
boxplot(apply(Data[,PDRitems],1,sum,na.rm=T)) #12/5: 1 value of sum score 13 is outside the IQR



zRange(apply(Data[,PDRitems],1,sum,na.rm=T)) #12/5: z range = -3.1 to 2.26, but the sum score of 13 is inside of the z score range of |3.29|, 3.1- retain this case for now. May delete later if it appears that this case isn't a part of the population, but it doesn't appear that way now.

## [1] "-3.1:2.26"

apply(Data[,PDRitems],2, meanSD)#12/5: all item-level means/SDs appear reasonable mean 1.32:1.85 & sd 0.36:0.50

## shoepain itchy adjust arms walk bend   
## "1.57(0.50)" "1.74(0.44)" "1.47(0.50)" "1.77(0.42)" "1.77(0.42)" "1.69(0.46)"   
## welts breaths skirts breasts hot cold   
## "1.70(0.46)" "1.85(0.36)" "1.32(0.47)" "1.51(0.50)" "1.52(0.50)" "1.55(0.50)"   
## stand drag catch   
## "1.57(0.49)" "1.84(0.37)" "1.48(0.50)"

1. Descriptives
   * Mean, Median, Standard Deviation, Frequency, Range
     + *Base R: mean(), median(), sd(), table(), range())*

rbind(  
"All"=apply(Data[,PDRitems], 2, meanSD),  
"Male"=apply(Data[which(Data$gender==1),PDRitems], 2, meanSD),  
"Female"=apply(Data[which(Data$gender==2),PDRitems], 2, meanSD))

## shoepain itchy adjust arms walk   
## All "1.57(0.50)" "1.74(0.44)" "1.47(0.50)" "1.77(0.42)" "1.77(0.42)"  
## Male "1.78(0.42)" "1.82(0.38)" "1.64(0.48)" "1.83(0.37)" "1.83(0.37)"  
## Female "1.35(0.48)" "1.66(0.47)" "1.28(0.45)" "1.70(0.46)" "1.71(0.45)"  
## bend welts breaths skirts breasts   
## All "1.69(0.46)" "1.70(0.46)" "1.85(0.36)" "1.32(0.47)" "1.51(0.50)"  
## Male "1.84(0.37)" "1.86(0.34)" "1.94(0.24)" " NaN( NA)" " NaN( NA)"  
## Female "1.53(0.50)" "1.53(0.50)" "1.75(0.43)" "1.32(0.47)" "1.51(0.50)"  
## hot cold stand drag catch   
## All "1.52(0.50)" "1.55(0.50)" "1.57(0.49)" "1.84(0.37)" "1.48(0.50)"  
## Male "1.51(0.50)" "1.68(0.47)" "1.80(0.40)" "1.91(0.29)" "1.63(0.48)"  
## Female "1.52(0.50)" "1.43(0.50)" "1.34(0.47)" "1.76(0.43)" "1.32(0.47)"

#12/5:Done  
  
rbind(  
"All"=apply(Data[,PDRitems], 2, median, na.rm=T),  
"Male"=apply(Data[which(Data$gender==1),PDRitems], 2, median, na.rm=T),  
"Female"=apply(Data[which(Data$gender==2),PDRitems], 2, median, na.rm=T))

## shoepain itchy adjust arms walk bend welts breaths skirts breasts hot  
## All 2 2 1 2 2 2 2 2 1 2 2  
## Male 2 2 2 2 2 2 2 2 NA NA 2  
## Female 1 2 1 2 2 2 2 2 1 2 2  
## cold stand drag catch  
## All 2 2 2 1  
## Male 2 2 2 2  
## Female 1 1 2 1

#12/5:Done  
  
sapply(Data[,PDRitems], table)#12/5:Done

## shoepain itchy adjust arms walk bend welts breaths skirts breasts hot cold  
## 1 343 204 423 186 180 248 241 121 268 191 384 354  
## 2 451 591 369 609 615 544 554 675 124 200 408 441  
## stand drag catch  
## 1 340 130 415  
## 2 457 666 379

sapply(Data[,PDRitems], table, Data$gender)#12/5:Done

## shoepain itchy adjust arms walk bend welts breaths skirts breasts hot cold  
## [1,] 90 72 144 67 67 65 55 24 0 0 195 129  
## [2,] 313 331 258 335 336 335 348 380 0 0 207 274  
## [3,] 253 132 279 119 113 183 186 97 268 191 189 225  
## [4,] 138 260 111 274 279 209 206 295 124 200 201 167  
## stand drag catch  
## [1,] 81 36 150  
## [2,] 323 368 252  
## [3,] 259 94 265  
## [4,] 134 298 127

range(apply(Data[,PDRitems], 1, sum, na.rm=T))#12/5:Done, sum-score range= 13-30

## [1] 13 30

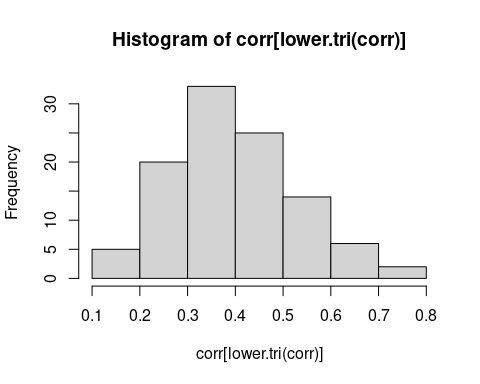
+ Correlation matrix   
 - \*psych package: tetrachoric()\*

#install.packages("psych")  
library(psych)  
corr=round(tetrachoric(Data[,PDRitems])$rho,2)#12/5:Done

range(corr[lower.tri(corr)]) #12/5:rho's range from 0.11:0.73, seems reasonable

## [1] 0.11 0.73

hist(corr[lower.tri(corr)]) #12/5: most corr coefs are around 0.3-0.5, again, reasonable



+ Standardized Cronbach’s Alpha  
 - \*psych package: alpha()\*

alpha(Data[,PDRitems])$total$std.alpha #12/5:Done - 0.83 alpha internal consistency

## [1] 0.8305561

## Replication of paper analyses

1. Dimensionality
   * Scree plots
     + *psych package: fa.parallel()*

plot(fa.parallel(Data[,PDRitems])$fa.values)

+ ICLUST  
 - \*psych package: ICLUST()\*

ICLUST(tetrachoric(Data[,PDRitems])$rho)

+ Bi-factor modeling  
 - \*psych package: omega()\*

omega(tetrachoric(Data[,PDRitems])$rho)

*review of dimensionality analyses results alongside items and literature review to qualitative guide decision on number of factors to included or modelled*

1. IRT Assumption Checks  
   *Will need to be further fleshed out for which model is selected*
   * Monotonicity - *assumed for dichotomous data*
     + Unidimensionality
       - Omega hierarchical/total & Explained Common Variance
         * *psych package: omega()*

omega(tetrachoric(Data[,PDRitems])$rho)[c("omega\_h", "omega.tot", "ECV")]

- Local dependence   
 + \*Rasch: Q3\*  
 + \*2pl: jackknife slope index – Q3 for rasch\*  
 - Item invariance (DIF)  
 + \*Rasch: Mantel-Haenszel\*  
 + \*2PL: logistic ordinal regression for 2pl?\*

1. IRT Modeling
   * Models to be evaluated:
     + Rasch
       - *psych package: rasch()*
     + Rasch w/guessing
       - *???psych package: rasch()*
     + 2PL
       - *psych package: ltm()*
     + Multidimensional IRT
       - *mirt package: mirt()*
     + IRTree Modeling
       - *???? package: ???()*
     + Explanatory IRT Modeling
       - *???? package: ???()*
   * Model fit
     + Goodness of fit/deviance
       - *Package: AIC,BIC, Chi-Sq?*
   * Item fit
     + Infit/outfit/signed chisq
       - *Package: TBD*

## What are the criteria for confirming and disconfirming the hypotheses?

### Hypothesis 1

There is either a unidimensional or multidimensional construct of ‘discomfort/distraction for appearance’ as measured by the PDR clothing items?

**Confirming**  
1. There is additional literature supporting a construct of ‘discomfort/distraction for appearance’ as indicated by wearing PDR clothing.  
2. Scree plots, ICLUST and Bifactor analyses indicate a reasonable number of factors (1 to 5) in the PDR clothing set

**Disconfirming**  
1. There is additional literature contradicting a construct of ‘discomfort/distraction for appearance’ as indicated by wearing PDR clothing.  
2. Scree plots, ICLUST and Bifactor analyses indicate an unreasonable or untestable number of factors (e.g., 6+) in the PDR clothing set

### Hypothesis 2

An Item Response Theory model can be fit to the dimensions in the PDR clothing item data, and will allow further evaluation and comparison of PDR clothing items in relation to the central construct(s) of ‘discomfort/distraction for appearance’.

**Confirming**  
1. Model assumptions will be met for each IRT model to be fit, including monotonicity, local dependence and item invariance  
2. Goodness of fit indices for the model is appropriate (e.g., CFI/TLI >0.95, RMSEA <0.08) and for item fit (e.g., non-significant signed chi-square)

**Disconfirming**  
1. Model assumptions are not met for each IRT model to be fit, including monotonicity, local dependence and item invariance  
2. Goodness of fit indices for the model are not appropriate (e.g., CFI/TLI <0.95, RMSEA >0.08) and item fit (e.g., significant signed chi-square)

## Have the analyses been validated on a subset of the data?

The analyses have not been validated on a subset of the data

## What is known about the data that could be relevant for the tested hypotheses?

Original publication introducing the PDR item displayed high internal consistency (cronbach’s alpha >0.9) and group-level differences between gender.

## Please provide a brief timeline for the different steps in the preregistration?

Planning and writing for the preregistration started in October, 2021. Data was simulated based on the codebook published on OSF <https://osf.io/ajv5z/> in November, 2021 and published on GitHub under a branch of the “Prereg\_Function” project, called “ClothingIRT” - (<https://github.com/zenit125/Prereg_Functions/tree/Clothing_IRT>). Set-up and hypothesis building will completed in December with results of additional literature review and simulated data/coding embedded in this parent RMarkdown document. The actual research below is planned for completion in December 2021.

Hypothesis 1 will be tested after completion of the set-up and hypothesis building. The results of Hypothesis 1 will be posted in an expanded version of this RMarkdown document.

Based on the results of Hypothesis 1 and any requisite additional background research, revisions to Hypothesis 2, coding and results will occur and will be documented in this parent RMarkdown document.