DS103-05-12

Lesson 5 Hands-On

Heather Walker - 2022-11-23

Table of Contents

01 Data Wrangling

02 Factor Analysis

03 Calculating Reliability

04 Conclusions

01 Data Wrangling

Data Wrangling

- Loading Data
- Viewing Data
- Removing null values*

*I performed the first few steps of Factor Analysis and my correlation matrix was full of NA values; this tipped me off that I needed to look at my data a little more before doing calculations.

```
# ----------------------------------------------------
# Load in Data
# ----------------------------------------------------
studentSurvey = read.csv("data/studentSurvey.csv")

# ----------------------------------------------------
# View Data
# ----------------------------------------------------
View(studentSurvey)

# ----------------------------------------------------
# Data Wrangling
# ----------------------------------------------------
studentSurvey1 <- studentSurvey[, 31:42]
studentSurvey1

# ----------------------------------------------------
#  | Remove null values
# ----------------------------------------------------
studentSurvey1edit <- drop_na(studentSurvey1)
View(studentSurvey1edit)
```

02 Factor Analysis

Factor Analysis

Exploratory Factor Analysis

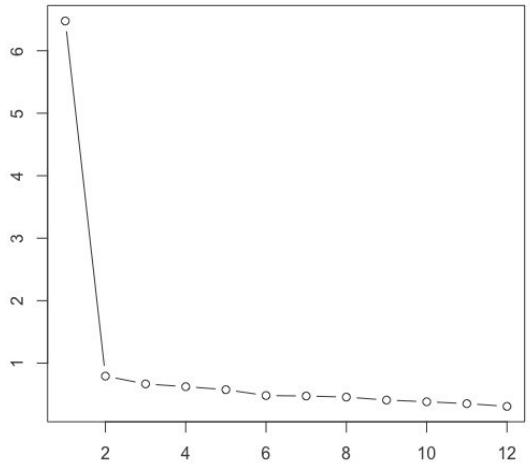
- Sample size
- Absence of Multicollinearity

```
# ----------------------------------------------------
#  | Sample Size
# ----------------------------------------------------
# Minimum sample size: 300+
# Current sample size: 464 rows after dropping null values

# ----------------------------------------------------
#  | Absence of Multicollinearity
# ----------------------------------------------------
# create correlation matrix
  studentSurveyMatrix <- cor(studentSurvey1edit)
  studentSurveyMatrix

  View(round(studentSurveyMatrix, 2)) # view the matrix
# There's nothing with a correlation HIGHER than 0.9, so
# there is no mutlicollinearity

# ----------------------------------------------------
#  | Some Relationship between Survey Items
# ----------------------------------------------------
# There's nothing with correlation LOWER than 0.3, so
# there is some relationship between survey items.

# ----------------------------------------------------
#  | |  Barlett's Test (confirm correlation matrix findings)
# ----------------------------------------------------
cortest.bartlett(studentSurvey1edit)
# Chi-Square value (`chisq`) and a _p_ value.
# You want this test to be significant, and if it is,
# this means that you have suitable correlations
# (not too high, not too low) to proceed with a factor analysis.
```

Factor Analysis

Exploratory Factor Analysis

- Determinants
- Determine Number of Factors
- Examining the Scree Plot




```
# ------------------------------------------------------
#  | |  Check your Determinants
#       (confirm correlation matrix findings)
# ------------------------------------------------------
det(studentSurveyMatrix)
# If this value is greater than .00001 (yes, that's 4 zeros),
# then again, you have a sufficient relation between your
# variables to proceed with a factor analysis.

# ------------------------------------------------------
#  Initial Pass to Determine Approximate Number of Factors
# ------------------------------------------------------
pcModel1 <- principal(studentSurvey1edit, nfactors = 10, rotate =
"none")
pcModel1

# ------------------------------------------------------
#  | Examine the Scree Plot
# ------------------------------------------------------
plot(pcModel1$values, type="b")
```

Factor Analysis Results:

0.4545455 result means that ~45% of residuals are large.

This is under 50%, so having only **1 factor** is a good model fit for the data.

03 Calculating Reliability

Reliability

- Question setup
- Calculating Reliability

Interpreting Output:

- Scale is reliable
- Inter-item reliability is good
- Items 4 and 5 are causing respondents to all answer the same way, vs. items 1, 2, 3.

```
# ----------------------------------------------------
# Question Setup
# ----------------------------------------------------
# Is my survey reliable?
# Does it measure the same thing every time?

# ----------------------------------------------------
# | Subsetting Data
# ----------------------------------------------------
# Since my model found only 1 factor, I only need to subset once
# keep subset `studentSurvey1edit`

# ----------------------------------------------------
# Calculating Reliability in R
# ----------------------------------------------------
# use `alpha` with data frame as argument

# ----------------------------------------------------
# | Interpreting output for `studentSurvey1edit`
# ----------------------------------------------------
alpha(studentSurvey1edit)
# `Chronbach's alpha` (raw alpha) = 0.87
#     = [scale is reliable]
# Corrected-item totals (r.drop) ~ 0.6+)
#     = [Inter-item reliability is good]
# `Non missing response frequency` seems even
#     = [no item is causing respondents to all answer
#       the same way]
```

04 Conclusions

Conclusions

A — Scale is reliable

B — Inter-item reliability is good

C — Items 4 and 5 are causing respondents to all answer the same way, vs. items 1, 2, 3.

This survey has good *inter-rater reliability* and *inter-item reliability*;
However, reliability testing indicates that item 4 and 5 need to be revised, as more people are answering the question the same way.

Thanks!

For questions on this project, contact **Heather Walker**.

For more information about editing slides, please read our FAQs or visit Slidesgo School: https://slidesgo.com/faqs and https://slidesgo.com/slidesgo-school