

Lesson 5 Hands-On

Directions

This Hands-On will be graded. The best way to become a data scientist is to practice!

Using [this data from a student survey](#) on the questions labeled Area1 through Area12, determine the following:

- Whether this data is suitable for factor analysis
- The appropriate number of factors for this data
- The best model fit for this data
- How the items group together into factors, if there is more than one factor indicated
- Whether the scale is considered reliable through inter-rater reliability
- Whether the scale is considered reliable through inter-item reliability

Please create a report or presentation that discusses your findings in these six areas.

```
# specifying the path
```

```
path <- "/Users/music/Desktop/studentSurvey.csv"
```

```
# reading contents of csv file
```

```
Survey <- read.csv(path)
```

```
# contents of the csv file
```

```
print (Survey)
```

```
library("corpcor")
```

```
library("GPArotation")
```

```
library("psych")
```

```
library("IDPmisc")
```

```
Survey1 <- Survey[, 1:42]
```

```
Survey2 <- Survey1[, 31:42]
```

```
Surveymatrix <- cor(Survey2)
```

```
View(round(Surveymatrix, 2))
```

```
# Has 1 down the diagonal
```

```
cortest.bartlett(Survey2)
```

```
# 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.
```

```
det(Surveymatrix)
```

```
# NaN so no sufficient relation between variables.
```

```
# This data is not suitable for factor analysis.
```

```
pcModel1 <- principal(Survey2, nfactors = 10, rotate = "none")
```

```
pcModel1
```

The factors are the appropriate number of factors for this data, PC! & SS loadings is 6.54 which is something real to examine.

```
plot(pcModel1$values, type="b")
```

```
pcModel2 <- principal(Survey2, nfactors = 2, rotate = "none")
```

```
residuals <- factor.residuals(Surveymatrix, pcModel2$loadings)
```

```
residuals <- as.matrix(residuals[upper.tri(residuals)])
```

```
largeResid <- abs(residuals) > .05
```

```
sum(largeResid)
```

```
sum(largeResid/nrow(residuals))
```

NA is not a pretty good model fit for the data.

```
pcModel3 <- principal(Survey2, nfactors = 2, rotate = "oblimin")
```

```
pcModel3
```

High loadings are h2, com,

Low loadings are u2

High and Low loadings are TC1, TC2

```
print.psych(pcModel3, cut = .3, sort=TRUE)
```

same results

```
pcModel4 <- principal(Survey2, nfactors = 2, rotate = "varimax")
```

```
print.psych(pcModel4, cut=.3, sort=TRUE)
```

```
alpha(Survey2)
```

inter-rater reliability is raw_alpha approx .8 for good, we have .91 which is good

inter-item reliability is raw.r is also because it is above .3