

Homework 3

Applied Multivariate Analysis

Emorie Beck

September 22, 2018

1 Workspace

1.1 Packages

```
library(car)
library(knitr)
library(psych)
library(lavaan)
library(semPlot)
library(kableExtra)
library(multcomp)
library(lme4)
library(plyr)
library(tidyverse)
library(MVN)
```

1.2 data

The file, Set_5.csv, contains data from a study in which college students completed the NEO-PI Personality Inventory. This 240-item scale purportedly measures the Big Five personality dimensions, assumed to be fairly independent. The inventory is scored on 6 subscales per dimension, listed below. The file contains the subscale scores, rather than the individual items, which should help reduce the impact of the small sample size.

Neuroticism: Anxiety Neuroticism: Angry_Hostility Neuroticism: Depression Neuroticism: Self_Consciousness
Neuroticism: Impulsiveness Neuroticism: Vulnerability Extraversion: Warmth Extraversion: Gregariousness
Extraversion: Assertiveness Extraversion: Activity Extraversion: Excitement_Seeking Extraversion: Posi-
tive_Emotions Openness: Fantasy Openness: Aesthetics Openness: Feelings Openness: Actions Openness:
Ideas Openness: Values Agreeableness: Trust Agreeableness: Straightforwardness Agreeableness: Altru-
ism Agreeableness: Compliance Agreeableness: Modesty Agreeableness: Tender_Mindedness Conscientious-
ness: Competence Conscientiousness: Order Conscientiousness: Dutifulness Conscientiousness: Achieve-
ment_Striving: Conscientiousness: Self_Discipline Conscientiousness: Deliberation

```
wd <- "https://github.com/emoriebeck/homeworks/raw/master/multivariate/homeworks/homework6"

dat <- sprintf("%s/Set_5(1).csv", wd) %>%
  read.csv(., stringsAsFactors = F)

head(dat)
```

##	ID	Anxiety	Angry_Hostility	Depression	Self_Consciousness	Impulsiveness
## 1	2	2.625	2.000	1.750	2.250000	2.625
## 2	3	3.625	2.875	3.000	3.500000	4.250
## 3	4	3.000	2.750	2.625	2.875000	3.000
## 4	5	4.375	3.125	4.500	4.000000	3.875
## 5	6	3.500	2.875	3.000	2.571429	3.625
## 6	7	4.000	4.125	2.875	2.375000	4.000

##	Vulnerability	Warmth	Gregariousness	Assertiveness	Activity
## 1	2.166667	4.666667	4.000	3.000000	4.833333
## 2	2.125000	4.500000	2.750	2.625000	3.000000
## 3	2.875000	3.750000	3.125	2.375000	3.250000
## 4	3.750000	3.250000	2.250	2.500000	1.875000
## 5	2.750000	3.750000	3.125	3.285714	3.500000
## 6	3.125000	3.500000	2.625	3.375000	3.125000

##	Excitement_Seeking	Positive_Emotions	Fantasy	Aesthetics	Feelings
## 1	3.500	4.750	3.857143	3.571429	4.666667
## 2	2.875	3.500	3.500000	4.125000	3.625000
## 3	3.875	3.375	3.375000	3.500000	3.250000
## 4	2.750	2.625	3.000000	3.750000	4.250000
## 5	3.750	3.625	3.125000	1.625000	3.125000
## 6	2.000	3.375	3.500000	2.000000	3.250000

##	Actions	Ideas	Values	Trust	Straightforwardness	Altruism	Compliance
## 1	2.571429	4.400	4.600	5.000	2.166667	4.833333	2.750
## 2	3.000000	3.875	3.125	3.250	3.750000	3.625000	3.125
## 3	2.375000	4.125	3.500	3.250	3.125000	4.000000	3.750
## 4	3.375000	2.750	4.125	3.000	3.428571	3.875000	4.000
## 5	2.750000	2.500	3.625	3.375	3.250000	4.125000	3.625
## 6	2.625000	1.125	3.625	2.500	2.875000	3.000000	2.250

##	Modesty	Tender_Mindedness	Competence	Order	Dutifulness
## 1	4.000	3.833333	4.50	3.625	3.285714
## 2	2.625	3.250000	3.00	2.250	3.875000
## 3	2.750	3.250000	3.75	3.250	3.750000
## 4	4.125	3.750000	2.75	3.000	2.875000
## 5	3.375	3.375000	3.75	4.000	3.750000
## 6	2.625	3.375000	3.00	3.625	2.625000

##	Achievement_Striving	Self_Discipline	Deliberation
## 1	4.333333	4.250	2.875
## 2	2.750000	3.750	3.500
## 3	3.375000	3.375	3.125
## 4	2.875000	2.625	3.250
## 5	3.375000	2.875	3.375
## 6	3.000000	2.625	2.625

```
source <- tribble(
  ~Factor, ~Facet,
  "Neuroticism", "Anxiety",
  "Neuroticism", "Angry_Hostility",
  "Neuroticism", "Depression",
  "Neuroticism", "Self_Consciousness",
  "Neuroticism", "Impulsiveness",
  "Neuroticism", "Vulnerability",
  "Extraversion", "Warmth",
  "Extraversion", "Gregariousness",
```

```

"Extraversion", "Assertiveness",
"Extraversion", "Activity",
"Extraversion", "Excitement_Seeking",
"Extraversion", "Positive_Emotions",
"Openness", "Fantasy",
"Openness", "Aesthetics",
"Openness", "Feelings",
"Openness", "Actions",
"Openness", "Ideas",
"Openness", "Values",
"Agreeableness", "Trust",
"Agreeableness", "Straightforwardness" ,
"Agreeableness", "Altruism",
"Agreeableness", "Compliance",
"Agreeableness", "Modesty",
"Agreeableness", "Tender_Mindedness",
"Conscientiousness", "Competence",
"Conscientiousness", "Order",
"Conscientiousness", "Dutifulness",
"Conscientiousness", "Achievement_Striving",
"Conscientiousness", "Self_Discipline",
"Conscientiousness", "Deliberation"
)

dat <- dat %>% select(ID, source$Facet)

```

Use confirmatory factor analysis to answer the following questions.

2 Question 1

First, test the hypothesis that the structure of personality is best described by five independent factors. How well does this model fit the data? Base your decision on the χ^2 goodness of fit test along with the goodness-of-fit index of your choice.

```

b5.base <- '
# define the measurement model
E =~ Warmth + Gregariousness + Assertiveness + Activity + Excitement_Seeking + Positive_Emotions
A =~ Trust + Straightforwardness + Altruism + Compliance + Modesty + Tender_Mindedness
C =~ Competence + Order + Dutifulness + Competence + Order + Dutifulness
N =~ Anxiety + Angry_Hostility + Depression + Self_Consciousness + Impulsiveness + Vulnerability
O =~ Fantasy + Aesthetics + Feelings + Actions + Ideas + Values
,

b5.uncorr <-
,

# uncorrelated factors
E ~~ 0*A
E ~~ 0*C
E ~~ 0*N
E ~~ 0*O

A ~~ 0*C
A ~~ 0*N
A ~~ 0*O

```

```

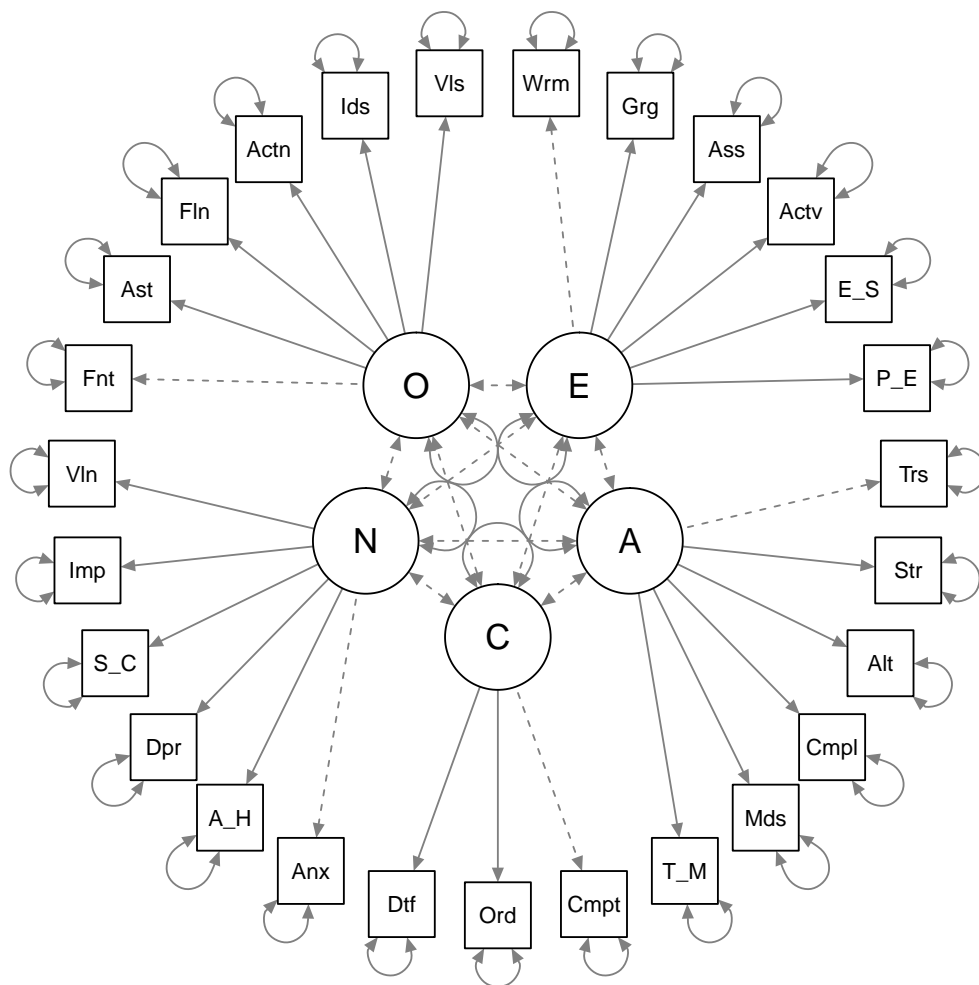
C ~~ 0*N
C ~~ 0*0
'

b5.mod <- paste(b5.base, '\n\n', b5.uncorr, sep = '', collapse = '')

fit1 <- cfa(b5.mod, dat, orthogonal = T)

semPaths(fit1, layout = "circle2")

```



```

summary(fit1)

## lavaan 0.6-3 ended normally after 49 iterations

```

```

##
## Optimization method NLMINB
## Number of free parameters 54
##
## Number of observations 201
##
## Estimator ML
## Model Fit Test Statistic 1467.470
## Degrees of freedom 324
## P-value (Chi-square) 0.000
##
## Parameter Estimates:
##
## Information Expected
## Information saturated (h1) model Structured
## Standard Errors Standard
##
## Latent Variables:
## Estimate Std.Err z-value P(>|z|)
## E =~
## Warmth 1.000
## Gregariousness 1.054 0.098 10.751 0.000
## Assertiveness 0.789 0.096 8.242 0.000
## Activity 0.691 0.076 9.029 0.000
## Excitemnt_Skng 0.635 0.085 7.484 0.000
## Positive_Emtns 1.096 0.095 11.494 0.000
## A =~
## Trust 1.000
## Strghtfrwrdsnss 0.917 0.128 7.149 0.000
## Altruism 0.962 0.115 8.400 0.000
## Compliance 0.903 0.119 7.591 0.000
## Modesty 0.736 0.117 6.295 0.000
## Tender_Mnddnss 0.744 0.100 7.429 0.000
## C =~
## Competence 1.000
## Order 1.080 0.130 8.281 0.000
## Dutifulness 1.145 0.133 8.585 0.000
## N =~
## Anxiety 1.000
## Angry_Hostilty 0.566 0.071 7.998 0.000
## Depression 1.091 0.074 14.773 0.000
## Self_Conscsnss 0.785 0.061 12.795 0.000
## Impulsiveness 0.424 0.063 6.783 0.000
## Vulnerability 0.829 0.061 13.530 0.000
## O =~
## Fantasy 1.000
## Aesthetics 1.096 0.137 7.973 0.000
## Feelings 0.745 0.102 7.269 0.000
## Actions 0.718 0.093 7.698 0.000
## Ideas 0.824 0.121 6.818 0.000
## Values 0.752 0.091 8.221 0.000
##
## Covariances:

```

```

##               Estimate Std.Err z-value P(>|z|)
## E ~~
##   A             0.000
##   C             0.000
##   N             0.000
##   O             0.000
## A ~~
##   C             0.000
##   N             0.000
##   O             0.000
## C ~~
##   N             0.000
##   O             0.000
## N ~~
##   O             0.000
##
## Variances:
##               Estimate Std.Err z-value P(>|z|)
## .Warmth         0.151   0.022   6.979   0.000
## .Gregariousness  0.254   0.032   8.028   0.000
## .Assertiveness   0.332   0.036   9.187   0.000
## .Activity        0.196   0.022   8.934   0.000
## .Excitemnt_Skng  0.277   0.030   9.377   0.000
## .Positive_Emtns  0.202   0.028   7.299   0.000
## .Trust           0.273   0.033   8.247   0.000
## .Strghtfrwrdsnss 0.306   0.035   8.703   0.000
## .Altruism        0.147   0.021   6.964   0.000
## .Compliance      0.232   0.028   8.320   0.000
## .Modesty         0.306   0.033   9.179   0.000
## .Tender_Mnddnss  0.172   0.020   8.476   0.000
## .Competence      0.153   0.026   5.971   0.000
## .Order           0.306   0.039   7.878   0.000
## .Dutifulness     0.142   0.031   4.665   0.000
## .Anxiety         0.168   0.023   7.175   0.000
## .Angry_Hostilty  0.333   0.035   9.582   0.000
## .Depression      0.183   0.026   6.925   0.000
## .Self_Conscsnss  0.175   0.021   8.366   0.000
## .Impulsiveness   0.274   0.028   9.723   0.000
## .Vulnerability   0.159   0.020   7.972   0.000
## .Fantasy         0.260   0.034   7.605   0.000
## .Aesthetics      0.412   0.050   8.200   0.000
## .Feelings        0.270   0.031   8.747   0.000
## .Actions         0.204   0.024   8.445   0.000
## .Ideas           0.409   0.045   8.993   0.000
## .Values          0.169   0.021   7.928   0.000
## E               0.283   0.043   6.517   0.000
## A               0.213   0.044   4.811   0.000
## C               0.214   0.039   5.464   0.000
## N               0.434   0.060   7.245   0.000
## O               0.265   0.051   5.236   0.000

```

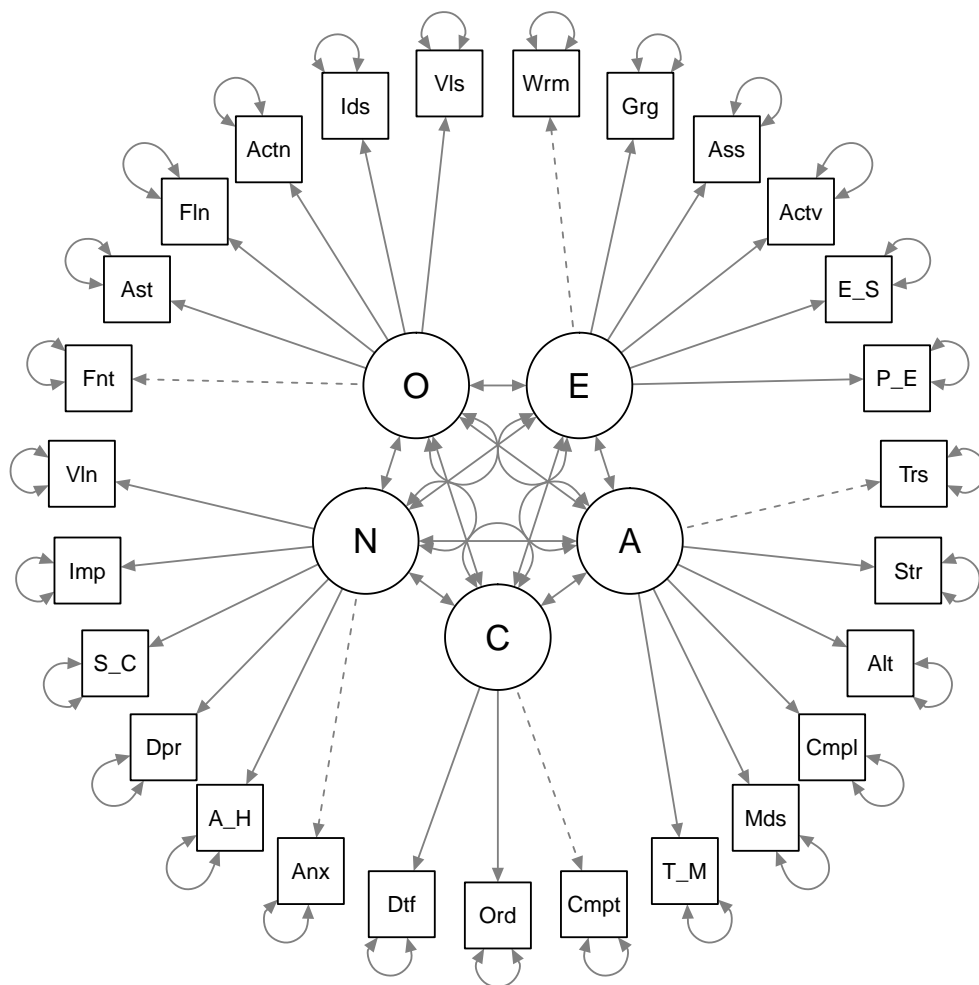
3 Question 2

Now allow the factors to correlate.

3.1 Part A

Does this model fit the data significantly better? Use a χ^2 difference test to answer the question.

```
fit2 <- cfa(b5.base, dat)
semPaths(fit2, layout = "circle2")
```



```
(c1 <- anova(fit1, fit2))

## Chi Square Difference Test
```

Table 1: Question 2B

lhs	E	A	C	N	O
E	0.29 [0.20, 0.37]	0.16 [0.11, 0.22]	0.12 [0.07, 0.17]	-0.05 [-0.11, 0.00]	0.17 [0.11, 0.23]
A		0.20 [0.12, 0.29]	0.11 [0.06, 0.15]	0.07 [0.02, 0.12]	0.13 [0.08, 0.18]
C			0.21 [0.14, 0.29]	-0.01 [-0.05, 0.04]	0.03 [-0.01, 0.07]
N				0.43 [0.32, 0.55]	0.07 [0.01, 0.13]
O					0.27 [0.17, 0.37]

```
##
##          Df      AIC      BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fit2 314 9142.4 9353.8 1224.6
## fit1 324 9365.2 9543.6 1467.5      242.88      10 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The orthogonal factor model fits the data better, $\chi^2_{diff}(10) = 242.88$.

3.2 Part B

Which of the factor correlations are statistically significant?

```
res2 <- parameterestimates(fit2, standardized = T)

res2 %>% tbl_df %>%
  filter(op == "~" & lhs %in% c("E", "A", "C", "N", "O")) %>%
  full_join(crossing(lhs = c("E", "A", "C", "N", "O"), rhs = c("E", "A", "C", "N", "O"))) %>%
  mutate(sig = ifelse(pvalue < .05, "sig", "ns")) %>%
  select(lhs, rhs, est, ci.lower, ci.upper, sig) %>%
  mutate_at(vars(est:ci.upper), funs(sprintf("%.2f", .))) %>%
  mutate_at(vars(lhs, rhs), funs(factor(., levels = c("E", "A", "C", "N", "O")))) %>%
  mutate(value = sprintf("%s [%s, %s]", est, ci.lower, ci.upper),
         value = ifelse(sig == "sig", sprintf("\\textbf{%s}", value), value),
         value = ifelse(is.na(value), "", value)) %>%
  select(lhs, rhs, value) %>%
  spread(key = rhs, value = value) %>%
  kable(., "latex", booktabs = T, escape = F,
        caption = "Question 2B") %>%
  kable_styling(full_width = F)
```

4 Question 3

Test a model that constrains all factor correlations to be equal.

4.1 Part A

Is this constraint acceptable (i.e., is it statistically different from the model tested in Question 2)?

4.2 Part B

Is the estimated latent variable correlation significant?

5 Question 4

Use the most parsimonious model from the first three steps. Constrain the loadings within each dimension to be equal. Is this simplification acceptable?

6 Question 5

Use the modification indices to diagnose the major problem with the model in Question 2. What change to that model would produce the biggest improvement in model fit?