The Personality Structures of the 50 States

Anisha Babu, Diana DeWald, Ian Shryock, Dillon Welindt, Futing Zou

Contents

```r
SAPA <- SAPA %>%
  # filter in US
  filter(country == "USA") %>%
  # filter 50 states
  filter(!state %in% c('District of Columbia', 'Guam', 'Palau','Puerto Rico', 'Virgin Islands',
                        'Northern Mariana Islands', 'American Samoa', 'Marshall Islands', NA))

# filter out rows with all NA
SAPA <- SAPA[rowSums(is.na(SAPA)) < 99, ]

# in list form select only used Q's
usedQ <- colnames(SAPA[8:106])
IPIPkeys <- map(IPIPkeysList, function(x) {
    x[match(usedQ, x)]
    na.omit(x)
})
# select only IPIP 100
IPIPkeys <- IPIPkeys[1:4]
```

```
# idk if this is what we want here - tableby package not available
map((SAPA[c(2,3,5,6,7)]), table)
```

```
## $gender
##
## Female    Male
##   56901   24465
##
## $age
##
##    14    15    16    17    18    19    20    21    22    23    24    25    26    27    28    29
##   767  1113  3129  6462  7707  6709  5896  4885  3621  2999  2569  2300  2193  2085  1887  1649
##    30    31    32    33    34    35    36    37    38    39    40    41    42    43    44    45
##  1590  1453  1345  1230  1212  1240  1152  1159  1069   958  1010   840   866   847   815   791
##    46    47    48    49    50    51    52    53    54    55    56    57    58    59    60    61
##   688   691   681   626   665   542   537   500   433   389   309   301   246   217   212   126
##    62    63    64    65    66    67    68    69    70    71    72    73    74    75    76    77
##   132    88    68    82    50    47    29    27    27    22    21     6    12     7     5     4
##    78    79    80    81    82    83    84    85    86    87    88    89    90
##     5     3     4     1     1     2     1     3     2     2     1     2     1
##
## $state
##
##          Alabama          Alaska         Arizona        Arkansas      California
##              643             555             866             577            9709
##         Colorado     Connecticut        Delaware         Florida         Georgia
##             1097             986             592            2936            2414
##           Hawaii           Idaho        Illinois         Indiana            Iowa
##              292             340            5520            1707             982
##           Kansas        Kentucky       Louisiana           Maine        Maryland
##              808             820            2030             356            1772
##    Massachusetts        Michigan       Minnesota     Mississippi        Missouri
##             1935            2549            2104             604            1611
##          Montana        Nebraska          Nevada   New Hampshire      New Jersey
##              243             580             274             389            2495
##       New Mexico        New York  North Carolina    North Dakota            Ohio
##             1199            4942            1454             190            3600
##         Oklahoma          Oregon    Pennsylvania    Rhode Island  South Carolina
##              771            1203            4758             422            1010
##     South Dakota       Tennessee           Texas            Utah         Vermont
##              172            1133            4662             487             161
##         Virginia      Washington   West Virginia       Wisconsin         Wyoming
##             2787            1742             384            2377             126
##
## $race
##
## African American         Chinese Indian/Pakistani        Japanese
##             6108            1129             469             257
##           Korean          Latino         Mexican Native American
##              500            2079            2166             728
##            Other     Other Asian Pacific Islander        Philipino
##             3067             566             305             615
##     Puerto Rican White/Caucasian
```
```
                                   2
```

```
##              512             62859
##
## $education
##
##          College graduate     Currently attending college
##                    12381                           32469
## Graduate or professional degree         High school graduate
##                    10338                            6145
##          Less than 12 years    Some college did not graduate
##                    11759                            8274
```

```r
# score items
scores <- psych::scoreItems(keys = IPIPkeys, items = SAPA, min=1, max=6, totals = FALSE, impute = 'none
```

```r
library(arsenal)

# demographic table
demog_tab <- summary(tableby(~ age + gender + race + education,
                      data = SAPA, test = FALSE),
                  title = "Full Sample Demographics")
demog_tab
```

```
##
##
## Table: (\#tab:#3 descriptives statistics)Full Sample Demographics
##
## |                                                |  Overall (N=81366) |
## |:-----------------------------------------------|:------------------:|
## |**age**                                         |                    |
## |   Mean (SD)                      |   27.177 (11.343)  |
## |   Range                          |   14.000 - 90.000  |
## |**gender**                                      |                    |
## |   Female                         |     56901 (69.9%)  |
## |   Male                           |     24465 (30.1%)  |
## |**race**                                        |                    |
## |   N-Miss                         |           6        |
## |   African American               |      6108 (7.5%)   |
## |   Chinese                        |      1129 (1.4%)   |
## |   Indian/Pakistani               |       469 (0.6%)   |
## |   Japanese                       |       257 (0.3%)   |
## |   Korean                         |       500 (0.6%)   |
## |   Latino                         |      2079 (2.6%)   |
## |   Mexican                        |      2166 (2.7%)   |
## |   Native American                |       728 (0.9%)   |
## |   Other                          |      3067 (3.8%)   |
## |   Other Asian                    |       566 (0.7%)   |
## |   Pacific Islander               |       305 (0.4%)   |
## |   Philipino                      |       615 (0.8%)   |
## |   Puerto Rican                   |       512 (0.6%)   |
## |   White/Caucasian                |     62859 (77.3%)  |
## |**education**                                   |                    |
## |   College graduate               |     12381 (15.2%)  |
## |   Currently attending college     |     32469 (39.9%)  |
```

```
## |   Graduate or professional degree |   10338 (12.7%)   |
## |   High school graduate            |    6145 (7.6%)    |
## |   Less than 12 years              |   11759 (14.5%)   |
## |   Some college did not graduate   |    8274 (10.2%)   |
```

```
# to add for final: demographic table grouped by state

# to add for final:  improve correlation matrix format/names (below), include other personality traits


res <- cor(scores, use = "complete.obs")
round(res, 2)
```

```
##                       IPIP100agreeableness IPIP100conscientiousness
## IPIP100agreeableness                  1.00                     0.21
## IPIP100conscientiousness              0.21                     1.00
## IPIP100extraversion                   0.38                     0.13
## IPIP100intellect                      0.16                     0.08
##                       IPIP100extraversion IPIP100intellect
## IPIP100agreeableness                 0.38             0.16
## IPIP100conscientiousness             0.13             0.08
## IPIP100extraversion                  1.00             0.22
## IPIP100intellect                     0.22             1.00
```

```
library(apaTables)

apa.cor.table(scores, filename="Corr_table.doc", show.conf.interval=F)
```

```
## The ability to suppress reporting of reporting confidence intervals has been deprecated in this vers
## The function argument show.conf.interval will be removed in a later version.
```

```
##
##
## Means, standard deviations, and correlations with confidence intervals
##
##
##   Variable                  M    SD   1         2         3
##   1. IPIP100agreeableness   4.67 0.77
##
##   2. IPIP100conscientiousness 4.14 0.92 .21**
##                                         [.21, .22]
##
##   3. IPIP100extraversion      3.92 1.02 .38**       .13**
##                                         [.37, .38] [.13, .14]
##
##   4. IPIP100intellect         4.59 0.73 .16**       .08**       .22**
##                                         [.15, .16] [.07, .08] [.21, .23]
##
##
## Note. M and SD are used to represent mean and standard deviation, respectively.
## Values in square brackets indicate the 95% confidence interval.
## The confidence interval is a plausible range of population correlations
```

```
## that could have caused the sample correlation (Cumming, 2014).
##   * indicates p < .05. ** indicates p < .01.
##
```

```r
# average scores on included survey questions
average_surveyscores <- SAPA %>%
  summarize_at(vars(q_76:q_1989), mean, na.rm = TRUE)


# average scores on survey questions by state (for final: can group by other variables as well)
average_statescores <- SAPA %>%
  group_by(state) %>%
  summarize_at(vars(q_76:q_1989), mean, na.rm = TRUE)


# combine demographics & traits
factorSAPA <- cbind(SAPA[1:7], scores)

# nest data by state
by_state <- split(factorSAPA, factorSAPA$state)

# descriptive stats - THERE SHOULD BE A BETTER WAY TO DO THIS (SOME KIND OF NESTED MAP FXN?)  ## summar
stateAgree <- map(by_state, ~summarize(.x, meanFactor = mean(IPIP100agreeableness, na.rm = TRUE),
                                 sdFactor = sd(IPIP100agreeableness, na.rm = TRUE),
                                 nFactor = length(IPIP100agreeableness)))
stateCons <- map(by_state, ~summarize(.x, meanFactor = mean(IPIP100conscientiousness, na.rm = TRUE),
                                 sdFactor = sd(IPIP100conscientiousness, na.rm = TRUE),
                                 nFactor = length(IPIP100conscientiousness)))
stateExtra <- map(by_state, ~summarize(.x, meanFactor = mean(IPIP100extraversion, na.rm = TRUE),
                                 sdFactor = sd(IPIP100extraversion, na.rm = TRUE),
                                 nFactor = length(IPIP100extraversion)))
stateIntel <- map(by_state, ~summarize(.x, meanFactor = mean(IPIP100intellect, na.rm = TRUE),
                                 sdFactor = sd(IPIP100intellect, na.rm = TRUE),
                                 nFactor = length(IPIP100intellect)))
# group all factors in list
allFactor = list(stateAgree, stateCons, stateExtra, stateIntel)
names(allFactor) = c('Agreeableness', 'Conscientiousness', 'Extraversion', 'Intellect')

# function to create table - SHOULD FIND A WAY TO CAPTION EACH TABLE BY STATE & FACTOR
makeTable <- function(x) {
  kable(x, booktabs = TRUE, longtable = TRUE, col.names = c("Mean", "SD", "N"),
        caption = paste('State of ', x, ' Descriptives')) %>%
    landscape() %>%
    kable_styling(font_size = 12, latex_options = c("scale_down", "repeat_header")) %>%
    kable_classic()
}

# create tables for each factor & state
map(allFactor, ~map(.x, ~makeTable(.x)))
```

```r
stateFactordf
```

```
##   Alabama Alaska Arizona Arkansas California Colorado Connecticut Delaware
## 1       5      6       5        5          5        5           5        6
```

5

```
##    Florida Georgia Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana
## 1       5      5       6     6       5       5    5      5        5         5
##    Maine Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana
## 1     5        5             5        5         5           6        5       7
##    Nebraska Nevada New.Hampshire New.Jersey New.Mexico New.York North.Carolina
## 1        5      7             6          5          5        5              5
##    North.Dakota Ohio Oklahoma Oregon Pennsylvania Rhode.Island South.Carolina
## 1            7    5        5      5            5            7              5
##    South.Dakota Tennessee Texas Utah Vermont Virginia Washington West.Virginia
## 1            7         5     5    5       8        5          5             6
##    Wisconsin Wyoming
## 1         5       8
```

1 Introduction

1.1 Big Five

One of the most widely replicated findings within the field of personality psychology is the Big Five structure of personality. With roots in the 1800's, personality psychology sought to determine the best way to represent the large number of personality traits in a concise structure. This research initially involved researchers providing participants with large numbers of trait descriptive adjectives and asking them to rate the extent to which those adjectives characterize themselves or someone they knew. Dimension reduction analyses were then used to create a simpler structure from those responses.

Multiple research groups began converging on the five factor structure as early as the 1960's, with an increasing consensus by the late 1980's. Most of the recent work on the big five has been conducted through a combination of confirmatory factor analysis and theory driven selection of survey items based on previous findings about the structure.

1.2 Geographical Personality

In recent years, there has been increasing focus on regional variation of personality traits within the United States. Work has examined the extent to which regions of the US differ on the Big Five domains and can be said to have distinct and characteristic combinations of trait levels. For example, Rentfrow and colleagues (2013) show that the south and midwest are best characterized as friendly and conventional, whereas the west is relaxed and creative, and the northeast is temperamental and uninhibited.

A limitation of this work is that it examines the extent to which the five factor structure captures each region and what differences in the levels of each factor are due to regional variation. This research utilizes confirmatory factor analyses that assume that the five factor structure is the ideal level of dimensionality to characterize all regions.

1.3 Cross Cultural Studies

Much of the cross-cultural work on personality structure has found some support for the notion that the five factor structure has applicability in a number of cultures. However, these studies typically are conducted from an etic perspective that translate the items used in western samples.

However, when studies are conducted from an emic perspective – that is, using trait descriptive adjectives from the language of the culture, rather than translations of items used in the big five framework – different structures emerge. A varying number of factors have been found to best fit different cultures, ranging from one to seven in many cases.

1.4 Geographical Factor Structure within US

Within the US, the regional variation in factor structures has not been an extensively studied topic. Because most research operates within a framework that utilizes confirmatory factor analysis, there is little information on the extent to which regions differ in their factor structure.

In the current study, we use exploratory factor analyses to provide estimates of the optimal factor structures for each of the fifty states.

2 Brief Methods

2.1 Measures

The International Personality Item Pool is an open-source repository of personality trait items that have been researched extensively in the big five tradition. The current study uses ninety nine of one hundred items from the IPIP-100. Participants rated themselves on a number of personality traits from 1- not at all like me to 6- very much like me.

2.2 Data Collection

Data were obtained from the Harvard Dataverse. Data were initially collected using the Synthetic Aperture for Personality Assessment (Revelle et al., 2016; Condon and Revelle, 2014; Wilt et al., 2011) which utilizes a massively missing completely at random design, wherein each participant only provides responses to a fraction of items.

3 Analyses

First, we provide descriptive norms for the entire US sample, and then by state.

Next, we use parallel analysis to determine the optimal number of factors in the whole sample. Our hypothesis is that five factors will provide an optimal fit.

The main analyses are fifty parallel analyses, one for every state, that estimates the optimal number of personality dimensions for each state. We hypothesize that there will be variation in the number of ideal dimensions across states.

List of Tables

List of Figures