SEM Analysis

Alexander James Ryan r0693898 3rd June 2018

Contents

Abstract

Using a personality dataset from Goldberg et. al. (2000), a series of Structural Equation Models were built to estimate the latent variables Extroversion, Neuroticism, Agreeableness, Conscientiousness and Openness. The model indicated that there is a large amount of cross loading between various items and different latent factors, and strong correlations between items.
This model showed that on average there was a positive correlation between the trait Extroversion and the traits Agreeableness, Consciousnesses, and Openness, while there was a negative correlation with the trait Neuroticism. It was also found that on average there was a negative correlation between the trait Neuroticism and the traits Consciousnesses and Openness. In addition, it was found on average there was a positive correlation between the trait Agreeableness and Consciousnesses and Openness.
It was found that Women on average have higher levels of the traits Extroversion, Neuroticism, Agreeableness and Consciousnesses, while Men were found to be higher on average in the trait Openness. The data also indicated that age on average has a mild negative effect on the trait Neuroticism, which implies that as one ages, the amount of negative emotion experienced should decrease on average. The psychological structure of a number of different countries were also explored, indicating which countries tended to be higher or lower on average in the psychological traits.

Introduction

There have been a lot of attempts to model the human psychological structure using statistics. These modelling techniques fall under the category of Confirmatory Factor Analysis or Structural Equation Modelling. These are techniques designed to estimate factors, called a latent variable, that are either impossible or very difficult to measure, using other data sources. These data sources may be in the form of surveys, or questionnaires, or more recently, in the form of 'Likes' on Facebook, or 'Followers' and 'Following' lists on Twitter. Research in this field is still in its infancy, and its insidious possibilities have only been brought to light in the last eighteen months, see Cambridge Analytica. In this paper, a dataset from a psychological study conducted by Goldberg et. al. in 2012 will be used to reveal the psychological traits, known as the 'Big Five Factors'. The list of questions in the survey have already been systemically tailored to derive a good estimate of an individual's latent personality structure. Ignoring the extensive literature already developed on the model, I hope to show how all these factors relate to one another, including the variables such gender and country of origin.

Research Questions

Some of the research questions to be developed in this paper include: The relationship between the Latent variables:
Extroversion, Openness, Neuroticism, Agreeableness, Consciousnesses
Which items load onto different factors
What are the differences in personality structure between Men and Women
What effect does age have on the distribution of an individual's personality structure
How does one's country affect one's personality distribution
Are some countries higher or lower in personality factors than others

Dataset explanation

The dataset is an online questionnaire conducted in 2012 by Goldberg et al., with 19719 participants, 12111 men and 7608 women. 50 different survey questions (called items) relating to personality using Likert scaling were asked and each respondent had to rank their response from 1, they disagree, to 5, they agree (0 meant the question was missed). The questions were not asked in any particular order. Other questions that were asked, were the race of the person, a categorical value, the age of the participant, a binary variable if the

participant was fluent in English, the gender of the person, and finally their dominant writing hand. The IP address of the participant was also recorded, so there is also a country of origin variable.

Each item is designed to reveal a different latent variable, the personality traits, Extroversion, Neuroticism, Agreeableness, Conscientiousness, and Openness (shortened to ENACO).

Dataset Limitations

Some criticisms of the validity of the data include the presence of acquiesce bias, where survey participants have a tendency to agree with a disproportionate number of questions, irrespective of the content of the question. This is managed by using items with opposite meanings, for example, a person who answers agree to question "E5 I start conversations" is not going to also write agree to "E6 I have little to say".

Another criticism is that the nature of the responses is self reported, that is, the participants are answering personality questions about themselves. This opens up the possibility of bias in the results. Participants may answer the questions in how they wish to be, rather how they actually feel.

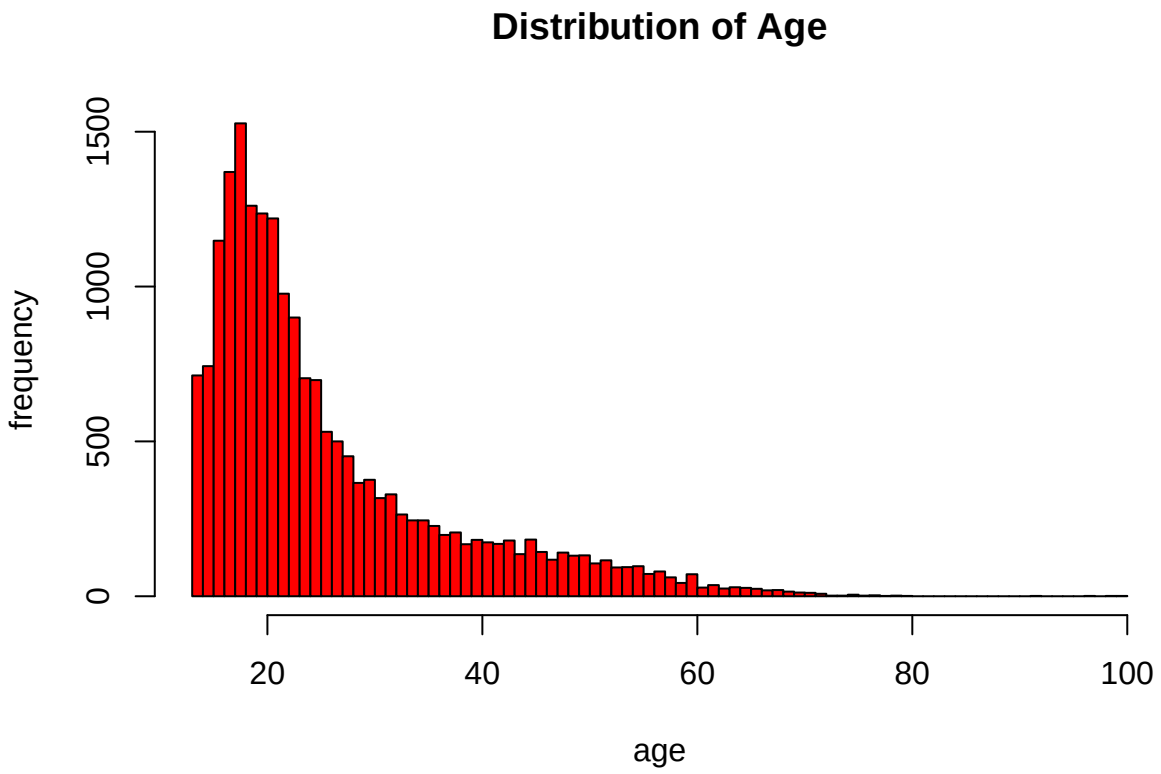
There may be bias present in who decides to conduct an online survey. A participant will need access to a computer, be reasonably computer literate, and have access to the internet. This enough would bias any results, in trying to estimate personality structures in a population.

Dataset Exploration

The table below shows the disproportion between survey participants, and their country of origin.

	country	freq
1	(nu	369
12	AU	974
24	BR	175
29	CA	924
39	DE	191
40	DK	122
52	FR	129
53	GB	1531
68	ID	172
69	IE	107
72	IN	1464
76	IT	277
107	MY	247
111	NL	133
112	NO	147
114	NZ	157
119	PH	649
120	PK	222
127	RO	135
133	SE	169
134	SG	133
149	US	8753
156	ZA	179

The mean age is approximately 26, and the median age is 22. The histogram below clearly shows that age is not normally distributed, and is skewed to the right. This might have an impact on estimating the effect of age upon one's psychological structure.



Model Specification

Restrictions are placed on E1, N1, A2, C1, O1, the items that positively correlated with each personality trait. This means that positive loadings relate to stronger values for that particular personality trait, whereas negative loadings imply there is a negative relation. For example, if Extroversion correlates negatively (loading=-0.45) onto Neuroticism, this means that those higher in the trait Extroversion, will on average have lower levels of the trait Neuroticism.

Model 1: Five Factor Model

Each personality latent variable has its own items relating to that latent variable. This is the model 1, and can be contrasted against with all further models.

```
E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10
```

Model 2: Five Factor Model (with Age, Gender and Hand)

The factors age, gender and hand are used as regression coefficients for the personality latent variables. The regressions include age and gender loading onto ENACO latent variables, with hand loading onto Agreeableness and Openness. It was found that the variable hand on the latent factors ENC, was not statistically significant.

```
E ~ age + gender
N ~ age + gender
```

```
A ~ age + gender + hand
C ~ age + gender
O ~ age + gender + hand
```

Model 3: Five Factor Model (with Country)

Binary variables, representing the country where each survey participant is from. Countries that were not statistically significant were removed.

```
E ~ US + GB + AU + IE + NL + NO + SE + IN + DK + CA + BE + BR
N ~ US + GB + AU + DE + NL + NO + IN + DK + CA + BE + BR
A ~ US + FR + AU + DE + IE + ID + IN + DK + CA + BE + BR
C ~ US + GB + FR + IT + AU + DE + AT + ID + ES + SE + BR
O ~ US + FR + IT + AU + DE + AT + ID + NL + NO + ES + RO + IN + DK + CA + BE + BR
```

Model 4: Five Factor Model (with correlated residuals and cross loadings)

After using the Modifications indices function, it was found that a lot of items were strongly correlated. For example, item N7 'I change my mood a lot' is very similarly worded to item N8 'I have frequent mood swings', and thus survey participants will probably answer in a similar way to both questions. The cross loadings include items that appear to have a strong relation to other latent factors. For example, item E3 'I feel comfortable around people', can be thought of us a measure of Neuroticism.

```
Residual Correlations
  N7 ~~ N8
  O1 ~~ O8
  O2 ~~ O4
  O5 ~~ O10
  O3 ~~ O6
  E8 ~~ E9
  A2 ~~ A7
  A4 ~~ A9
  C2 ~~ C6
  E2 ~~ E6
  N1 ~~ N3
  N4 ~~ N10
  A5 ~~ A7
  N9 ~~ A3
```

```
Cross Loadings
    N =~ E3
    O =~ C3
    C =~ A3
    E =~ A7
    E =~ A4
    E =~ A2
    E =~ A10
    N =~ C4
    O =~ C10
    O =~ E6
    N =~ O9
    N =~ A3
    N =~ A6
    A =~ E3
    N =~ O3
    E =~ N10
    A =~ N9
```

Model 5: All the above combined into one model.

Model Assessment and Model Fit

	m1	m2	m3	m4	m5
npar	110.000	125.000	184.000	141.000	217.000
chisq	96502.971	100390.091	105355.798	46083.189	59172.942
df	1165.000	1300.000	2341.000	1134.000	2458.000
pvalue	0.000	0.000	0.000	0.000	0.000
nfi	0.744	0.739	0.728	0.878	0.850
cfi	0.746	0.741	0.732	0.880	0.855
tli	0.733	0.727	0.717	0.871	0.845
rmsea	0.064	0.062	0.047	0.045	0.034
rmsea.ci.lower	0.064	0.062	0.047	0.044	0.034
rmsea.ci.upper	0.065	0.062	0.047	0.045	0.034
srmr	0.074	0.071	0.051	0.053	0.037

Chi squares test is statistically significant in all models, which means there is not a good fit in any of the models. The chi square statistic is very sensitive to sample size, and according to Schlermelleh-Engel et al. (2003), this statistic should not be the sole judgement for acceptance or rejection of a model.

Models 1-4 appear to not have a good fit based on the CFI index. The CFI value is well below the range where it is considered good (>0.90), and for Model 4 and Model 5, it is just under the threshold.

According to MacCallum, Browne and Sugawara (1996), an RMSEA value of (0.01), (0.05), and (0.08) indicates excellent, good, and mediocre fits respectively. All models have an acceptable RMSEA, which is in the bounds of an acceptable model. Hu & Bentler (1999) deem an SRMR value of less than (0.08) to be a good fit. All models presented have an SRMR value lower than this threshold.

The choice comes down to Model 4 and Model 5. Model 4 has a marginally better CFI value, but Model 5 has a superior RMSEA and SRMR. There is more parameters estimated in Model 5, and thus we can learn more about the personality structures, thus Model 5 will be analysed.

Model results

Covariances between Personality Factors

Extroversion loads negatively onto Neuroticism (-0.252), and positively onto Agreeableness (0.253), Conscientiousness (0.101) and Openness (0.224). This implies those higher in trait Extroversion on average have lower levels of Neuroticism, and higher levels of Agreeableness, Conscientiousness and Openness.

Neuroticism loads negatively onto Conscientiousness (-0.223) and Openness (-0.169). It's loading on Agreeableness is not significant (-0.028). This implies that those higher in trait Neuroticism, have on average lower levels of Conscientiousness and Openness.

Agreeableness loads positively onto Consciousnesses (0.128) and Openness (0.160). This implies those with higher levels of the trait Agreeableness have on average higher levels of Consciousness and Openness.

Conscientiousness loads marginally positively onto Openness (0.095). This implies that those with higher levels of Consciousness on average may have higher levels of Openness.

Effect of Gender on Personality Factors

Gender, an exogenous variable, is loaded onto the latent factors ENACO as regression coefficients. Women were coded with 0, while men were coded with 1. Gender has a statistically significant loading onto each

personality factor. The standardized latent variable for Extroversion loading is (coefficient=-0.109), for Neuroticism (coefficient=-0.370), for Agreeableness (coefficient=-0.233), for Conscientiousness (coefficient=-0.085), and Openness (coefficient=0.158). These results imply that women on average have a higher level of Extroversion, Agreeableness, Neuroticism, and Conscientiousness than men, and a lower level of Openness than men. This result is roughly congruent with the study done by Schmitt et. al. in 2008, except they found that there was negligible average differences in trait Openness between men and women.

Effect of Age on personality factors

Age was statistically significant for all the latent factors ENACO. Age appears to have a positive effect on Extroversion (coefficient=0.008), Agreeableness (coefficient=0.006), Consciousnesses (coefficient=0.014) and Openness (coefficient=0.005), and a negative effect on Neuroticism (coefficient=-0.012). This result implies that as people age, on average they will experience lower levels of negative emotion.

Effect of dominate hand on personality factors

Dominate hand, an exogenous variable is loaded onto the latent factors ENACO as regression coefficients. Right hand coded as 1, Left handed coded as 2, ambidextrous coded as 3. For Extroversion, Neuroticism, and Consciousness, the loading was not statistically significant. This implies that being left or right handed does not affect these personality characteristics. For Agreeableness, the coefficient is (-0.030) and for Openness it is (0.090). This implies that those who are left handed, or ambidextrous, are on average marginally lower in the trait Agreeableness, as well as marginally higher in the trait Openness.

Effect of country of origin factors

As well as the self rated race of the survey participant, the geographical location was also recorded, including the IP address of the country in which the test was conducted. The results were not proportionately distributed amoungst the different countries, with the majority of survey participants coming from the US, Britain, Ireland, and Australia, all native English speaking countries. Many of the country factors were not statistically significant, and these were removed from the model.

Below is a list of the Country and their effect upon each latent factor.

Extroversion

As can be seen in the table, those countries higher on average in the trait Extroversion include Belgium, Denmark, Ireland, Norway, while countries lower on average seem to be Brazil, Philippines and Singapore.

Factor	Country	Estimate	Std.Err	P(>|z|)
E~	US	0.065	0.015	0.000
E~	GB	0.103	0.024	0.000
E~	AU	0.117	0.029	0.000
E~	IE	0.220	0.081	0.007
E~	NO	0.182	0.070	0.010
E~	IN	0.069	0.025	0.006
E~	DK	0.323	0.079	0.000
E~	BE	0.377	0.094	0.000
E~	BR	-0.326	0.066	0.000
E~	PH	-0.131	0.035	0.000
E~	SG	-0.243	0.072	0.001

Neuroticism

Countries higher on average in trait Neuroticism include Pakistan, Philippines, England, India and Brazil, while those countries lower in trait Neuroticism include Denmark, Norway, Belgium and Germany.

Factor	Country	Estimate	Std.Err	P(>|z|)
N~	US	-0.093	0.016	0.000
N~	GB	0.136	0.027	0.000
N~	DE	-0.138	0.067	0.038
N~	NO	-0.255	0.076	0.001
N~	IN	0.228	0.027	0.000
N~	DK	-0.378	0.084	0.000
N~	CA	-0.094	0.032	0.003
N~	BE	-0.291	0.099	0.003
N~	BR	0.251	0.071	0.000
N~	PH	0.123	0.039	0.001
N~	PK	0.353	0.061	0.000

Agreeableness

Revealed in the table, the countries higher on average in trait Agreeableness include Belgium, Ireland, and India, while those countries lower in Agreeableness include Brazil.

Factor	Country	Estimate	Std.Err	P(>|z|)
A~	US	0.072	0.009	0.000
A~	AU	0.055	0.018	0.002
A~	IE	0.140	0.050	0.005
A~	IN	0.145	0.015	0.000
A~	DK	0.102	0.047	0.031
A~	CA	0.042	0.018	0.018
A~	BE	0.179	0.056	0.001
A~	BR	-0.156	0.040	0.000
A~	PK	0.132	0.035	0.000
A~	NZ	0.091	0.041	0.025
A~	ZA	0.093	0.038	0.015

Conscientiousness

The table shows those countries higher in the trait Conscientiousness on average include the US, Indonesia, and the Philippines, while countries lower include Austria, Brazil, Germany and France.

Factor	Country	Estimate	Std.Err	P(>|z|)
C~	US	0.071	0.012	0.000
C~	GB	-0.121	0.020	0.000
C~	FR	-0.172	0.063	0.006
C~	IT	-0.147	0.043	0.001
C~	AU	-0.116	0.024	0.000
C~	DE	-0.172	0.052	0.001
C~	AT	-0.346	0.157	0.028
C~	ID	0.233	0.054	0.000
C~	ES	-0.158	0.078	0.043
C~	SE	-0.217	0.055	0.000

Factor	Country	Estimate	Std.Err	P(>|z|)
C~	BR	-0.259	0.055	0.000
C~	PH	0.150	0.030	0.000

Openness

The table reveals that those higher in Openness on average are countries like Austria, Germany, Netherlands, Denmark and Brazil, while those countries lower in Openness on average include Malaysia, Pakistan, Philippines, Indonesia and India. According to Butler et al. (2000) higher levels of Openness are associated with liberal ideas, and those lower in Openness tend to support Right or Left-wing Authoritarianism.

Factor	Country	Estimate	Std.Err	P(>|z|)
O~	US	0.060	0.012	0.000
O~	FR	0.121	0.058	0.037
O~	IT	0.094	0.040	0.020
O~	DE	0.205	0.048	0.000
O~	AT	0.430	0.146	0.003
O~	ID	-0.208	0.051	0.000
O~	NL	0.131	0.057	0.022
O~	NO	0.136	0.055	0.014
O~	ES	0.146	0.073	0.044
O~	RO	0.128	0.057	0.024
O~	IN	-0.185	0.020	0.000
O~	DK	0.197	0.061	0.001
O~	CA	0.069	0.023	0.003
O~	BE	0.143	0.073	0.048
O~	BR	0.174	0.051	0.001
O~	PH	-0.165	0.028	0.000
O~	PK	-0.364	0.045	0.000
O~	MY	-0.400	0.043	0.000
O~	SG	-0.113	0.058	0.051
O~	ZA	0.121	0.050	0.015

Conclusion

Despite the limitations with the dataset, there are a number of revealing ideas. It is roughly congruent with previous studies done on differences in personality structure between nations, as well as gender. Also revealed is the way the Big Five Factors relate to one another, for example, that higher levels of Neuroticism correlate with lower levels of Consciousness, Extroversion, and Openness.

As stated in the introduction, this is a rapidly evolving field, and with improvements in sourcing data, models of personality structure will become more sophisticated.

Bibliography

Goldberg, Lewis R. "The development of markers for the Big-Five factor structure." Psychological assessment 4.1 (1992): 26. http://dx.doi.org/10.1037/1040-3590.4.1.26

Donnellan, M. B., & Lucas, R. E. (2008). Age Differences in the Big Five Across the Life Span: Evidence from Two National Samples. Psychology and Aging, 23(3), 558–566. http://doi.org/10.1037/a0012897

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130-149.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55.

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003), "Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodnessof-Fit Measures", Methods of Psychological Research Online, Vol. 8, No. 2, pp.23-74.

Butler, J. C. (2000). "Personality and emotional correlates of right-wing authoritarianism". Social Behavior and Personality. 28: 1–14. doi:10.2224/sbp.2000.28.1.1.

David P. Schmitt; Realo, A; Voracek, M., & Allik, J. (2008). "Why can't a man be more like a woman? Sex differences in big five personality traits across 55 cultures". Journal of Personality and Social Psychology. 94 (1): 168–182. doi:10.1037/0022-3514.94.1.168. PMID 18179326.

Apendix

Questions asked in Survey

Personality Questions

E1 I am the life of the party.
E2 I don't talk a lot.
E3 I feel comfortable around people.
E4 I keep in the background.
E5 I start conversations.
E6 I have little to say.
E7 I talk to a lot of different people at parties.
E8 I don't like to draw attention to myself.
E9 I don't mind being the center of attention.
E10 I am quiet around strangers.
N1 I get stressed out easily.
N2 I am relaxed most of the time.
N3 I worry about things.
N4 I seldom feel blue.
N5 I am easily disturbed.
N6 I get upset easily.
N7 I change my mood a lot.
N8 I have frequent mood swings.
N9 I get irritated easily.
N10 I often feel blue.
A1 I feel little concern for others.
A2 I am interested in people.
A3 I insult people.
A4 I sympathize with others' feelings.
A5 I am not interested in other people's problems.

A6 I have a soft heart.
A7 I am not really interested in others.
A8 I take time out for others.
A9 I feel others' emotions.
A10 I make people feel at ease.
C1 I am always prepared.
C2 I leave my belongings around.
C3 I pay attention to details.
C4 I make a mess of things.
C5 I get chores done right away.
C6 I often forget to put things back in their proper place.
C7 I like order.
C8 I shirk my duties.
C9 I follow a schedule.
C10 I am exacting in my work.
O1 I have a rich vocabulary.
O2 I have difficulty understanding abstract ideas.
O3 I have a vivid imagination.
O4 I am not interested in abstract ideas.
O5 I have excellent ideas.
O6 I do not have a good imagination.
O7 I am quick to understand things.
O8 I use difficult words.
O9 I spend time reflecting on things.
O10 I am full of ideas.

Other Questions

race

1 = Mixed Race
2 = Arctic (Siberian, Eskimo)
3 = Caucasian (European)
4 = Caucasian (Indian)
5 = Caucasian (Middle East)
6 = Caucasian (North African, Other)
7 = Indigenous Australian
8 = Native American
9 = North East Asian (Mongol, Tibetan, Korean Japanese, etc)
10 = Pacific (Polynesian, Micronesian, etc)
11 = South East Asian (Chinese, Thai, Malay, Filipino, etc)
12 = West African, Bushmen, Ethiopian
13 = Other (0=missed)

engnat Response to "is English your native language?". 1=yes, 2=no (0=missed)

gender Chosen from a drop down menu. 1=Male, 2=Female, 3=Other (0=missed)

hand "What hand do you use to write with?". 1=Right, 2=Left, 3=Both (0=missed)

R Code

```
library(plyr)
nrow(ds) # number of subjects = 19719
```

```r
ncol(ds) # number of variables = 57
mean(ds$age) # mean age is ~26
median(ds$age) # median age is 22
count(ds, 'gender')
count(ds, 'country')
```

```r
# Model 1:  Five Factor Model
# (A2 needs to go first, so agreeableness factor is positive for high agree)

models$m1 <- '
    # measurement model
      E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
      N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
      A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
      C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
      O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10'

  fits$m1 <- sem(models$m1, data=ds)
    summary(fits$m1, fit.measures = TRUE, standardized = TRUE)


# Model 2:  Five Factor Model with age and gender and hand

models$m2 <- '
    # measurement model
      E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
      N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
      A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
      C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
      O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10
    # regressions
      E ~ age + gender + hand
      N ~ age + gender + hand
      A ~ age + gender + hand
      C ~ age + gender + hand
      O ~ age + gender + hand
'
  fits$m2 <- sem(models$m2, data=ds)
    summary(fits$m2, fit.measures = TRUE, standardized = TRUE)


# Model 3:  Five Factor Model (with Country)

models$m3 <- '
    # measurement model
      E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
      N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
      A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
      C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
      O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10
    # regressions
      E ~ US+GB+AU+IE+NL+NO++IN+DK+CA+BE+BR+PH+SG+ZA
      N ~ US+GB+AU+DE+NL+NO+IN+DK+CA+BE+BR+PH+PK+ZA
      A ~ US+FR+AU+DE+IE+IN+DK+CA+BE+BR+PK+NZ+ZA
```

```
      C ~ US+GB+FR+IT+AU+DE+AT+ID+ES+SE+BR+PH+ZA
      O ~ US+FR+IT+DE+AT+ID+NL+NO+ES+RO+IN+DK+CA+BE+BR+PH+PK+MY+SG+ZA
'
  fits$m3 <- sem(models$m3, data=ds)
    summary(fits$m3, fit.measures = TRUE, standardized = TRUE)


# Bonus Model:  Five Factor Model (with Race)

# models$mb <- '
#     # measurement model
#       E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
#       N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
#       A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
#       C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
#       O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10
#     # regressions
#       E ~ r2 + r11
#       N ~ r3 + r4 + r5 + r11 + r12
#       A ~ r1 + r3 + r7 + r9 + r11
#       C ~ r1 + r3 + r4 + r6 + r7 + r9
#       O ~ r1 + r3 + r11 + r12
# '

  # fits$mb <- sem(models$mb, data=ds)
  #   summary(fits$mb, fit.measures = TRUE, standardized = TRUE)


# Model 4:  Five Factor Model (with correlated residuals)

models$m4 <- '
    # measurement model
      E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
      N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
      A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
      C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
      O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10
    # residual correlations
      N7 ~~ N8
      O1 ~~ O8
      O2 ~~ O4
      O5 ~~ O10
      O3 ~~ O6
      E8 ~~ E9
      A2 ~~ A7
      A4 ~~ A9
      C2 ~~ C6
      E2 ~~ E6
      N1 ~~ N3
      N4 ~~ N10
      A5 ~~ A7
      N9 ~~ A3
    # cross loadings
```

```
      N =~ E3
      O =~ C3
      C =~ A3
      E =~ A7
      E =~ A4
      E =~ A2
      E =~ A10
      N =~ C4
      O =~ C10
      O =~ E6
      N =~ O9
      N =~ A3
      N =~ A6
      A =~ E3
      N =~ O3
      E =~ N10
      A =~ N9
'
  fits$m4 <- sem(models$m4, data=ds)
    summary(fits$m4, fit.measures = TRUE, standardized = TRUE)

# Model 5:  Final model

models$m5 <- '
    # measurement model
      E =~ E1 + E2 + E3 + E4 + E5 + E6 + E7 + E8 + E9 + E10
      N =~ N1 + N2 + N3 + N4 + N5 + N6 + N7 + N8 + N9 + N10
      A =~ A2 + A1 + A3 + A4 + A5 + A6 + A7 + A8 + A9 + A10
      C =~ C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10
      O =~ O1 + O2 + O3 + O4 + O5 + O6 + O7 + O8 + O9 + O10
    # regressions
      E ~ age + gender
      N ~ age + gender
      A ~ age + gender + hand
      C ~ age + gender
      O ~ age + gender + hand
      E ~ US+GB+AU+IE+NO+IN+DK+BE+BR+PH+SG
      N ~ US+GB+DE+NO+IN+DK+CA+BE+BR+PH+PK
      A ~ US+AU+IE+IN+DK+CA+BE+BR+PK+NZ+ZA
      C ~ US+GB+FR+IT+AU+DE+AT+ID+ES+SE+BR+PH
      O ~ US+FR+IT+DE+AT+ID+NL+NO+ES+RO+IN+DK+CA+BE+BR+PH+PK+MY+SG+ZA
    # residual correlations
      N7 ~~ N8
      O1 ~~ O8
      O2 ~~ O4
      O5 ~~ O10
      O3 ~~ O6
      E8 ~~ E9
      A2 ~~ A7
      A4 ~~ A9
      C2 ~~ C6
      E2 ~~ E6
      N1 ~~ N3
```

```
      N4 ~~ N10
      A5 ~~ A7
    # cross loadings
      N =~ E3
      O =~ C3
      C =~ A3
      E =~ A7
      E =~ A4
      E =~ A2
      E =~ A10
      N =~ C4
      O =~ C10
      O =~ E6
      N =~ O9
      N =~ A3
      N =~ A6
      A =~ E3
      N =~ O3
      E =~ N10
      A =~ N9
'
  fits$m5 <- sem(models$m5, data=ds)
    summary(fits$m5, fit.measures = TRUE, standardized = TRUE)


v <- list()

v$fitindicies <- c("npar", "chisq", "df", "pvalue", "nfi", "cfi", "tli", "rmsea", "rmsea.ci.lower", "rms

# Core fit function
core_fitmeasure <- function(fit = fits$m1, fitindicies = v$fitindicies, digits = 3 ) {
  x <- fitMeasures(fit)
  round(x[fitindicies], digits)
}
```