

Homework Assignment 2

Claes Kock, Yuchong Wu, Mayara Latrech

1 Caution: Never copy the whole thing into the latest version from your own workspace

2 Example for making table:

```
table = data.frame(column1 = c(1, 2, "text"),
                    column2 = c(4, 5, 6))

# Then for PDF:
kable(table, caption = "title here", digits = 2)

# For presentation:
datatable(table)
```

3 Reminders:

1. Titles for all the plots and tables
-

Contents

1	Caution: Never copy the whole thing into the latest version from your own workspace	1
2	Example for making table:	1
3	Reminders:	1
4	Introduction	3
5	Theory	4
5.1	Measurement part	4
5.2	Structural Part	4
6	Dataset	5
7	Task 1	5
8	Task 2	6
8.1	One factor model:	6
8.2	Three factor models	7
8.2.1	Model 1	7
8.2.2	Model 2	7
8.2.3	Model 3	8
8.2.4	Model 4	8
8.2.5	Models evaluation	10
8.3	Fitted model	11
9	Conclusion	14
10	Appendix	15

4 Introduction

The purpose of this assignment is to perform a factor analysis. The assignment is divided into two parts: one theoretical, where we try to make a theoretical model of an equally theoretical latent variable. The second is more practical - we use an actual dataset. The purpose of the second part of the assignment is to get some hands-on experience performing factor analysis, including estimation, interpretation and model checking. It is here that we will build structural equation models with which we will perform factor analysis. Our dataset for this task is from Gierl and Rogers (1996). The latent variable we want to model is test anxiety. We model this by using the variables from the dataset and successively build a structural equation model.

5 Theory

Using factor analysis, we try to solve a problem: there are variables that we want to measure, but which we can not observe. These are called latent variables. These can typically be rather complex variables, such as quality of life or general mental ability. Structural equation models (SEM) are a tool that can be used to try to measure these latent variables. We try to explain the variation in the unexplained, latent variables (factors), using observed, correlated variables.

As an example we have IQ-tests. Here, IQ is the unknown factor that we want to measure. The actual measurement is done by a series of tests, which serve as the observable variables.

The structural equation models consist of two parts:

Measurement part: - Factor model for the endogenous variables - Factor model for the exogenous variables

Structural part: - Model implied by theory

5.1 Measurement part

Factor model for the endogenous variables - y as a function of latent variables η (eta):

$$y_q = \alpha_{q1}^y \eta_1 + \alpha_{q2}^y \eta_2 + \dots \alpha_{qs}^y \eta_s + \epsilon_q$$

Factor model for the endogenous variables - x as a function of latent variables ξ (xi):

$$y_q = \alpha_{p1}^x \xi_1 + \alpha_{p2}^x \xi_2 + \dots \alpha_{ps}^x \xi_s + \epsilon_p$$

Interpretation of factor loadings - standardize $V(y_i) = 1$ and $V(\eta) = 1$, then the OLS estimator

$$\hat{\alpha} = (Cov(y, \eta) / V(\eta)) = \rho_{y, \eta}$$

and

$$V(\eta) = \hat{\alpha}^2 V(\eta) + V(\eta)$$

$$1 = \hat{\alpha}^2 + V(\eta)$$

meaning, $\hat{\alpha}^2$ is the proportion of the variance of y originating from the factor η and $1 - \hat{\alpha}^2$ from the error.

5.2 Structural Part

The structural part consists of a regression model for each dependent latent variable as a function of: - Other dependent latent variables, endogenous, η - Latent exogenous variables, ξ

$$\eta_s = \beta_{s1} \eta_1 + \beta_{s2} \eta_2 + \dots + \beta_{ss-1} \eta_{s-1} + \gamma_{s1} \xi_1 + \gamma_{s2} \xi_2 + \gamma_{sr} \xi_r + \zeta_s$$

6 Dataset

Our dataset is from Gierl, M.J. and Rogers, T. (1996) A Confirmatory Factor Analysis of the Test Anxiety Inventory Using Canadian High School Students, Educational and Psychological Measurement 56(2).

The variables of the dataset measure test-related anxiety of Canadian high school students. Our data set contains observations from 335 male and 389 female high school students, of 20 variables each. However, we only use 7 variables and we do not use any gender variable. For this study we only look at male students, disregarding the female students.

7 Task 1

Our hypothetical occupation is being employed by the ministry of education. Our objective is to measure the degree of test anxiety among students, in order to learn more about why students do not like to take exams. We can assume this is a part of some long term study to to measure the mental health of students.

We want to measure test anxiety, which is difficult since it can not be seen or captured using a single method. We decide to divide the latent variable “test anxiety” into three new latent variables. These new variables have the advantage of being more clearly defined.

Table 1: Latent variables and Measurements

latent	measurement
physical symptoms	jittery
	heart_rate
intrusive thoughts	think_get_through
	thoughts_interfere
	Think_failing
worrying	Worry_before
	Uneasy

Our new latent variables are: - physical symptoms (1) physical symptoms - intrusive thoughts (2) mental symptoms - worrying (3) long term symptoms?

Physical symptoms cover intrusive physical stress symptoms, such as jitteriness or increased heart rate. Intrusive thoughts cover mental disturbances during the test. Worrying covers long term doubts of students, such as worrying about getting through school or passing tests.

Thus, we have broken “test anxiety” down into three different ways to measure test anxiety.

To measure the latent variables, we can use observable variables like the following:

- Being jittery during tests (1)
- increased heart rate during a test (1)
- thinking about finishing school (2)
- Thoughts interfere with concentration (2)
- thoughts about failing (2)
- Worrying about taking tests (3)
- General unease (3)

Being jittery or having an increased heart rate during tests are easily measurable physical symptoms, so they can be used to explain physical symptoms of anxiety. The occurrence of intrusive thoughts regarding academic performance is a measure for mental anxiety. We expect that general unease and worrying about tests will explain the “worrying” in the long term. Thus, we have broken down the latent variable “test anxiety” into physical, mental and long term stress parts.

8 Task 2

In this task, we are presented with many datasets. We choose the “Anxiety_Data” dataset to fit our model. It includes 335 observations of 7 variables: 1. Uneasy, upset feeling (3) 2. Jittery when taking tests (1) 3. Worry before important tests (3) 4. Heart beating fast during tests (1) 5. Thinking about getting through school (2) 6. Thoughts interfere with concentration (2) 7. Think about failing (2)

Table 2: Latent variables and Measurements

latent	measurement
physical symptoms	jittery
	heart_rate
intrusive thoughts	think_get_through
	thoughts_interfere
	Think_failing
worrying	Worry_before
	Uneasy

For people experiencing a high amount of stress, tests and performance can be scary and threatening. This results in students afflicted by both mental and physical symptoms. We expect all of these variables to be somewhat correlated with each other, as they are all expressions of test-related anxiety. Some of these variables are defined as occurring before exams and some occur during exams. Some variables are not defined in their relation to time, so we assume these variables to be constant, or at least occurring often. In general, we assume that all of our variables contribute to more test anxiety, so all of the variables should have a positive effect on their respective latent variables. Scoring a high score in all 3 latent categories would, in turn, mean that the individual is suffering a high amount of test-related anxiety.

8.1 One factor model:

For a first attempt at factor modeling, we start with a one factor model: We choose one latent variable: Anxiety and we fit it to being Uneasy, Jittery when taking tests, Worry before important tests, Heart beating fast during tests, Thinking about getting through school, Thoughts interfere with concentration, Think about failing.

The measures for the latent variable Anxiety are all significant.

Holding all variables constant, a unit change in Uneasy would affect the Anxiety level by 1.

Holding all variables constant, a unit change in Think_get_through would affect the Anxiety level by 0.8583135.

Holding all variables constant, a unit change in Thoughts_interfere would affect the Anxiety level by 1.0929684.

Holding all variables constant, a unit change in Jittery would affect the Anxiety level by 1.1111209.

Holding all variables constant, a unit change in Worry_before would affect the Anxiety level by 1.1427412.

Holding all variables constant, a unit change in Think_failing would affect the Anxiety level by 1.0189077.

Holding all variables constant, a unit change in Heartbeat_fast would affect the Anxiety level by 1.0609877.

We can calculate the R-squared value for each variable. We notice that the model captures 0.3841371, 0.2829946, 0.4588824, 0.4742515, 0.5016282, 0.3988008, 0.4324212.

Overall the data does not really support the model. In fact, our measurements are more aligned with being in 3 different categories than just one category called Anxiety. We can see that being Jittery and having a

fast Heartbeat can be categorized into physical symptoms. Thinking about getting through school, Thoughts interfere with concentration and Think about failing are Intrusive thoughts. Being Uneasy and Worrying before an exam can be categorized as worrying. Hence, we obtain three latent variables.

8.2 Three factor models

8.2.1 Model 1

We now try having three 3 different categories: physical symptoms, Intrusive thoughts and worrying. Where being Jittery and having a fast Heartbeat are measurements of physical symptoms. Thinking about getting through school, Thoughts interfere with concentration and Think about failing are measurements of Intrusive thoughts. Being Uneasy and Worrying before an exam are measurements of worrying.

The measures for the latent variable Physical symptoms are all significant.

Holding fast heartbeat constant, a unit change in Jitteriness would affect the Physical symptoms by 1.

Holding fast Jitteriness, a unit change in heartbeat constant would affect the Physical symptoms by 0.9537536.

The measures for the latent variable Intrusive thoughts are all significant.

Holding Thoughts interfere with concentration and Think about failing constant, a unit change in Thinking about getting through school would affect the Intrusive thoughts by 1.

Holding Thinking about getting through school and Think about failing constant, a unit change in Thoughts interfere with concentration would affect the Intrusive thoughts by 1.1044528.

Holding Thinking about getting through school and Thoughts interfere with concentration , a unit change in Think about failing constant would affect the Intrusive thoughts by 1.1049222.

The measures for the latent variable Worrying are all significant.

Holding Worrying before exam constant, a unit change in Uneasiness would affect the Physical symptoms by 1.

Holding Uneasiness constant, a unit change in Worrying before exam would affect the Physical symptoms by 1.

8.2.2 Model 2

Similarly to the first model, we try having three 3 different categories: physical symptoms, Intrusive thoughts and worrying. Where being Jittery and having a fast Heartbeat are measurements of physical symptoms. Thinking about getting through school, Thoughts interfere with concentration and Think about failing are measurements of Intrusive thoughts. Being Uneasy and Worrying before an exam are measurements of worrying. We now add regressions to the latent variables: physical symptoms is explained by uneasiness and worrying.

The measures for the latent variable Physical symptoms are all significant.

Holding fast heartbeat constant, a unit change in Jitteriness would affect the Physical symptoms by 1.

Holding fast Jitteriness, a unit change in heartbeat constant would affect the Physical symptoms by 0.9537536.

The measures for the latent variable Intrusive thoughts are all significant.

Holding Thoughts interfere with concentration and Think about failing constant, a unit change in Thinking about getting through school would affect the Intrusive thoughts by 1.

Holding Thinking about getting through school and Think about failing constant, a unit change in Thoughts interfere with concentration would affect the Intrusive thoughts by 1.1044528.

Holding Thinking about getting through school and Thoughts interfere with concentration , a unit change in Think about failing constant would affect the Intrusive thoughts by 1.1049223.

The measures for the latent variable Worrying are all significant.

Holding Worrying before exam constant, a unit change in Uneasiness would affect the Physical symptoms by 1.

Holding Uneasiness constant, a unit change in Worrying before exam would affect the Physical symptoms by 1.

For the regressions, only worrying is significant for explaining physical symptoms. Thoughts before the exam are not significant.

8.2.3 Model 3

Similarly to the second model, we try having three 3 different categories: physical symptoms, Intrusive thoughts and worrying. Where being Jittery and having a fast Heartbeat are measurements of physical symptoms. Thinking about getting through school, Thoughts interfere with concentration and Think about failing are measurements of Intrusive thoughts. Being Uneasy and Worrying before an exam are measurements of worrying. We remove thoughts before the exam from the regression since it was not significant in the last model.

The measures for the latent variable Physical symptoms are all significant. Holding fast heartbeat constant, a unit change in Jitteriness would affect the Physical symptoms by 1.

Holding fast Jitteriness, a unit change in heartbeat constant would affect the Physical symptoms by 0.9537406.

The measures for the latent variable Intrusive thoughts are all significant.

Holding Thoughts interfere with concentration and Think about failing constant, a unit change in Thinking about getting through school would affect the Intrusive thoughts by 1.

Holding Thinking about getting through school and Think about failing constant, a unit change in Thoughts interfere with concentration would affect the Intrusive thoughts by 1.1043475.

Holding Thinking about getting through school and Thoughts interfere with concentration , a unit change in Think about failing constant would affect the Intrusive thoughts by 1.1057598.

The measures for the latent variable Worrying are all significant.

Holding Worrying before exam constant, a unit change in Uneasiness would affect the Physical symptoms by 1.

Holding Uneasiness constant, a unit change in Worrying before exam would affect the Physical symptoms by 1.

For the regressions, worrying is still significant for explaining physical symptoms.

8.2.4 Model 4

Similarly to the third model, we try having three 3 different categories: physical symptoms, Intrusive thoughts and worrying. Where being Jittery and having a fast Heartbeat are measurements of physical symptoms. Thinking about getting through school, Thoughts interfere with concentration and Think about failing are measurements of Intrusive thoughts. Being Uneasy and Worrying before an exam are measurements of worrying. We now have two regressions. Physical symptoms are explained with worrying and thoughts before the exam are also explained with worrying.

The measures for the latent variable Physical symptoms are all significant.

Holding fast heartbeat constant, a unit change in Jitteriness would affect the Physical symptoms by 1.

Holding fast Jitteriness, a unit change in heartbeat constant would affect the Physical symptoms by 0.9537533.

The measures for the latent variable Intrusive thoughts are all significant.

Holding Thoughts interfere with concentration and Think about failing constant, a unit change in Thinking about getting through school would affect the Intrusive thoughts by 1.

Holding Thinking about getting through school and Think about failing constant, a unit change in Thoughts interfere with concentration would affect the Intrusive thoughts by 1.104453.

Holding Thinking about getting through school and Thoughts interfere with concentration , a unit change in Think about failing constant would affect the Intrusive thoughts by 1.1049222.

The measures for the latent variable Worrying are all significant.

Holding Worrying before exam constant, a unit change in Uneasiness would affect the Physical symptoms by 1.

Holding Uneasiness constant, a unit change in Worrying before exam would affect the Physical symptoms by 1.

For the regressions, worrying still significant for explaining physical symptoms but it is not significant for physical symptoms.

8.2.5 Models evaluation

We use a variety of methods to evaluate our models. We use AIC (Aikake information criterion) and BIC (Bayesian information criterion), to determine the strength of the models relative to each other. Additionally, we use CIF (Confirmatory factor analysis), RMSEA (root mean square error of approximation) and SRMR (Standardized Root Mean Square Residual) to determine model fit. We want all of these measurements to be as small as possible. The model with the smallest values is the model that we consider to be the best.

Table 3: Evaluation of models

	One-factor models	Model1	Model2	Model3	Model4
aic	5.998616e+03	5.933012e+03	5.933012e+03	5.931029e+03	5.933012e+03
bic	6.052014e+03	5.997853e+03	5.997853e+03	5.992055e+03	5.997853e+03
cfi	8.847601e-01	9.753281e-01	9.753281e-01	9.766258e-01	9.753281e-01
rmsea	1.364278e-01	7.121482e-02	7.121482e-02	6.636564e-02	7.121482e-02
srmr	6.824745e-02	3.587922e-02	3.587922e-02	3.592432e-02	3.587921e-02

As we can see in the above table, model 3 has the smallest values in the categories AIC, BIC and RMSEA, the most out of any model. Thus we consider it to be our best model. For this model, physical symptoms and intrusive thoughts are not correlated. We expected all of the latent variables to be somewhat correlated, so this is not what we expected. However, both of these latent variables are correlated with worrying.

8.3 Fitted model

1. Our final model contains three different latent variables: physical symptoms, Intrusive thoughts and worrying. Where being Jittery and having a fast Heartbeat are measurements of physical symptoms. Thinking about getting through school, Thoughts interfere with concentration and Think about failing are measurements of Intrusive thoughts. Being Uneasy and Worrying before an exam are measurements of worrying. We now have one regression to the latent variables: physical symptoms is explained by worrying. Hence, the model as a path diagram is:

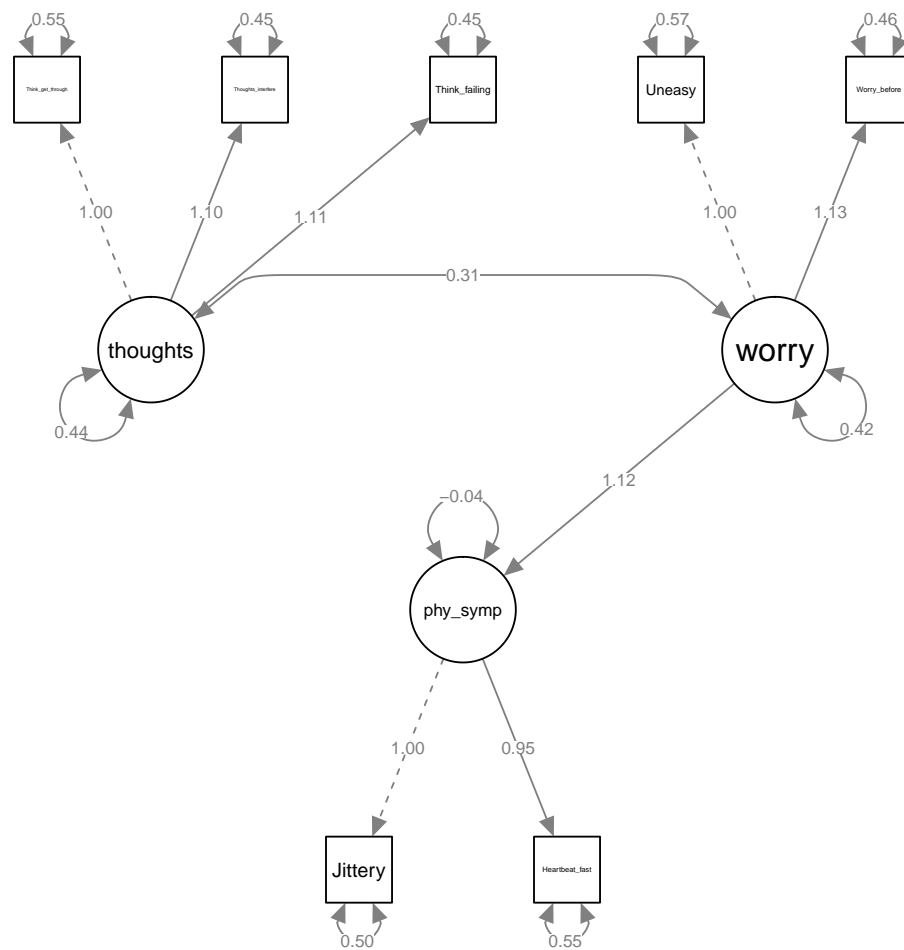


Figure 1: Fitted model

2. Model in equation form:

- Measurement Part:

$$phy.sym = 1Jittery + 0.9537533Heartbeat.fast + \epsilon_1$$

$$thoughts = 1Think.gt.thrgh + 1.104453Thoughts.ntrfr + 1.1049222Think.failing + \epsilon_2$$

$$worry = 1Uneasy + 1.1306737worry.before + \epsilon_3$$

- Structural Part:

$$phy.symp = 1.1157455worry + \zeta_1$$

3. -The latent variable phy_sym can be interpreted as the true score that can not be observed stemming from the observed measurements Jittery and Heartbeat_fast.

-The latent variable thoughts can be interpreted as the true score that can not be observed stemming from the observed measurements Think_gt_thrgh, Thoughts_ntrfr and Think_failing.

-The latent variable worry can be interpreted as the true score that can not be observed stemming from the observed measurements Uneasy and worry_before.

4. In order to assign a scale for the latent variables, we calculated the estimated latent variables using the model. Then, for each latent variable we calculated the maximum and minimum values. We chose a scale from one to 10 for the three latent variables.

For each value of the latent variable we attribute a score from 1 to 10. Where 1 is starting from the minimum value until the minimum value plus a step. 10 is starting from the minimum value plus 9*step until the maximum value. The step is calculated by:

$$step = \frac{max - min}{10}$$

Table 4: Factor Loadings

	phy_symp	thoughts	worry
Jittery	0.71	0.00	0.00
Heartbeat_fast	0.67	0.00	0.00
Think_get_through	0.00	0.67	0.00
Thoughts_interfere	0.00	0.74	0.00
Think_failing	0.00	0.74	0.00
Uneasy	0.00	0.00	0.65
Worry_before	0.00	0.00	0.74

5. Jiteeriness and Fast heartbeat have respectively 0.7053425 and 0.6727229 lodgings which indicates that the factor physical symptoms strongly influences those variables. all other variables have null lodgings which indicates that the factor physical symptoms has no influence on those variables.

Thinking about getting through school for Thoughts, Thoughts interfere with concentration and Think about failing have respectively 0.6680356 , 0.7378138 and 0.7381274 lodgings which indicates that the factor Intrusive thoughts strongly influences those variables. all other variables have null lodgings which indicates that the factor Intrusive thoughts has no influence on those variables.

Uneasiness and Worrying before exam have respectively 0.6533862 , 0.7387667 and 0.7381274 lodgings which indicates that the factor Worrying strongly influences those variables. all other variables have null lodgings which indicates that the factor Worrying has no influence on those variables.

6. Jitteriness seems to be a slightly stronger indicator than Faster heartbeat for physical symptoms. However they are very close. Thoughts interfere with concentration and Think about failing are slightly stronger indicators than Thinking about getting through school for Thoughts. Worrying before an exams is a slightly stronger indicator than uneasiness for Worrying. Overall, we can say that all observed variables are equally strong indicators of the latent variables.

9 Conclusion

The purpose of this assignment has been to gain some experience using factor modeling. Our dataset formed the basis for our latent variables which, in turn, formed our models. Our goal was to map student anxiety regarding exams. This was represented by our latent variables, consisting of physical symptoms, intrusive thoughts and worry. When creating the models, we worked our way up from using only one latent variable and adding more latent variables once the first one was rejected, due to not fitting our data. The full models, i.e. the ones that used all three latent variables, were evaluated using five different measures of model fit. We chose the model that was the best in the most amount of categories. This model did not completely match our expectations, as not all of the latent variables were correlated. However, the latent variable Worry was correlated with the other two latent variables. All our observable variables had a positive effect on their respective latent variables, and all of them were significant.

10 Appendix