EXPLORING SEXUAL COMPULSIVITY WITH FACTOR ANALYSIS

Group 1: Final Project Report

Vignesh Mulay

Sam Borak

Ryan Kim

Benjamin McGriff

STAT 4220

Professor Ferrara

April 22, 2019

Table of Contents

Background of Statistical Analysis

Overview of Method

Factor analysis is a multivariate statistical method and an unsupervised learning technique. It is used in statistical modelling and machine learning to determine the most effective model for a given context. Essentially, factor analysis — specifically exploratory factor analysis (EFA) — is used to a) reduce the number of relevant variables in a model and b) assess a data construct's dimensionality.[1] The foundational assumption of factor analysis is that for a given collection of measured variables, there is a smaller set of underlying factors that can explain the interrelationships between the observed variables. In abstract terms, EFA can help identify *latent constructs* that exist behind the variables included in a dataset. More concretely, EFA determines which measured variables are highly intercorrelated as these variables could very well just be reflecting one hidden factor. In this way, users of EFA can reduce the complexity of their data to generate the most precise predictive models. Confirmatory factor analysis (CFA) is another statistical method often used to validate the results of EFA. Specifically, while EFA describes the amount of latent factors needed to represent the data, CFA can specify the exact number as well as indicate, in particular, which measured variable is related to which underlying factor. In short, CFA tests whether or not a researcher's hypothesized model about a data construct is actually consistent with the aspects (dimensions and factors) of that construct.[2]

Methodology

When performing a factor analysis, a dataset must satisfy a number of assumptions. For one, all the variables used should be metric. A metric variable is a variable that is measured on a quantitative scale. The distance between points on this scale are homogeneous. For example, the distance between 1 and 4 is the same as the distance between 7 and 10. Dummy variables may be used but only in special cases. The sample size should be fairly large, usually greater than 200. The sample should also be homogeneous. If the sample is not homogeneous, the sample size increases as the number of variables increase. To check homogeneity between variables, a reliability analysis should be run. In exploratory factor analysis, multivariate normality is not a required assumption. There must also be some correlation between variables with a minimum correlation of 0.30. Lastly, there should be no outliers in the data. [3]

Once assumptions have been checked and met, the EFA method can be followed through. Firstly, the data is split in half. Half of the data will be used to create our EFA model, and the other half will be used to check the model with CFA. After the data is split, one would use the EFA data subset and create a correlation matrix which will be used to calculate the Eigenvalues. Eigenvalues are numeric indicators of the amount of variance explained by each factor. A higher eigenvalue suggests that a factor has greater variance which would mean they are substantially

---

[1] Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r

[2] Statistics Solutions. (2013). Confirmatory Factor Analysis. Retrieved from http://www.statisticssolutions.com/academic-solutions/resources/directory-of-statistical-analyses/confirmatory-factor-analysis/

[3] Statistics Solutions. (2013). Exploratory Factor Analysis. Retrieved from https://www.statisticssolutions.com/factor-analysis-sem-exploratory-factor-analysis/

different enough to be a stand-alone construct. The eigenvalues above the cutoff of 1 are considered statistically significant and relevant factors in our data. These eigenvalues above the cutoff can be seen visually in a Scree Plot. After isolating the exact number of factors, one can create the EFA model. This model can determine how each survey question (item) relates (correlates) to the underlying factors by reading the factor loadings. These factor loadings are essentially correlation results on a scale from -1 to 1. One can also see how each individual observation relates to the factors through the factor scores. This is done on a spectrum where high positive values reflect strong expression of the underlying factor, while high negative values reflect a strong lack of expression of the factor with 0 indicating neutral expression of the factor.

Once interpretations of the Exploratory Factor Analysis are obtained, one can begin to delve into how adequate (fit) the model is. This can be done by calculating Absolute Fit statistics (Chi-square test, Tucker-Lewis Index, Root Mean Square Error of Approximation). All fit statistics try to represent the discrepancy between the observed data and the expected given the constructed EFA model. Ideally, we want a non-significant Chi-Square test result, a Tucker-Lewis Index greater than .9 and a Root Mean Square Error of Approximation (RMSEA) less than .05. It is often difficult to hit all of these cutoffs so one may also cross-reference their model's adequacy against another model (a model with a different number of suspected factors). This is called Relative Model fit which compares the fit between multiple models using the Bayesian Information Criterion (BIC). The lower the BIC, the more relatively adequate the model is compared to the other model. While this concludes EFA, CFA can still be performed if needed to confirm the EFA results.

Once the assumptions for the CFA have been checked, which include multivariate normality, a sufficient sample size ($n > 200$), and a correct *a priori* model specification, one can perform CFA, using the other randomized half of our data. Most importantly, the EFA model must first be translated to CFA syntax. This transformation creates a 'theory' that specifies which items load onto each factor based on the researcher's assumptions of what the hidden factors are. Once the correct syntax is created, one can use the newly made theoretical object to perform CFA analysis (through statistical software). The analysis will produce the r-squared values of each item along with the parameter estimates for each item/factor relationship. The r-squared values indicate the proportion of variance in the underlying factor explained by a specific item, while the parameter estimates (factor loadings) represent how strongly an item loads onto a hidden factor. If both of these measures are consistently high, it would indicate that our theory seems to be correct. Ideally, if our model fits well, the factor loadings of our latent variables will be greater than 0.7. Similar to EFA, we can also use absolute fit (i.e RMSEA, Goodness of Fit Index, and Comparative Fit Index) and relative fit statistics (BIC values) to more accurately assess the validity of the theory.[4]

Practical Applications

While factor analysis is a very functional statistical tool, it only has a few niche use cases. For the most part, factor analysis is limited to data collected from surveys and self-reporting questionnaires due to the requirement of metric variables. In other words, factor analysis is

[4] Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r

employed when a scale is being used to measure a factor such as a Likert Scale (0-5). This is most commonly found in the fields of psychology, government, education and health.

Factor analysis may be used for a number of reasons including to reduce the number of variables being studied, develop theoretical constructs, and prove/disprove proposed theories. For instance, if a researcher is conducting a survey with 25 questions, factor analysis can help determine if certain questions are highly correlated and if the items are actually only measuring a handful of hidden factors, therefore allowing for the possibility to reduce the number of variables. Also, if a new data set is being studied, factor analysis can be useful for understanding the dimensionality of the data, specifically the dimensions that exist between observed and unobserved variables. Historically, the most important reason to use factor analysis is for theory development and validation. For example, when psychologists used factor analysis to evaluate personality survey data, they found that certain descriptors are repeatedly assigned to the same individuals and therefore identified the Big Five personality traits from a list of dozens of items.[5]

The main advantage of factor analysis is its ability to simplify a data set and a model. Through factor analysis, one can reduce both the number of dimensions and variables in a data set. The inclusion of irrelevant additional measures in a model leads to unnecessary noise and bias that will hinder the generalizability of the results. As the principle of Occam's Razor states, the most simplified explanation is the best as it will lead to the most parsimonious model — the one that relies on the least assumptions and variables and holds the greatest explanatory power.[6] The biggest disadvantage of factor analysis is the inherent subjectiveness that is involved. As EFA and CFA simply describe the number of factors that produce the best-fitting model, both of them cannot actually be used to identify what the hidden factors represent. Accordingly, then, the researcher is free to interpret the results of factor analysis in any way they see fit which naturally leads to researcher bias as individuals are incentivized to confirm their own theories.

Business Objectives

Business Understanding

We are analysts at KBMM HR Solutions, LLC., an American human resources consulting firm. We leverage analytics to provide clients with statistical solutions to HR problems, often dealing with employee management. Recently, many companies are becoming more aware of issues with sexual harassment and offensive sexual advances in the workplace. Due to this growing concern and the apparent need for statistical analysis, KBMM has began exploring the possibility to applying our data expertise to this relatively new area of human resources. Several of our clients have hired us to help them identify employees who may potentially be or become sexual harassers. To do this, firms can use the 10-question survey described in the next section to gauge sexual compulsivity which can be defined as an excessive preoccupation with sexual urges and behaviors that are difficult to control and cause distress. However, our clients aren't sure whether or not the questionnaire is an actual direct measure of sexual compulsivity or if there are other extraneous factors in play that they want to avoid.

---

[5] Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r

[6] Salkind, Neil J. *Encyclopedia of Research Design.* Thousand Oaks, CA: SAGE Publications, Inc., 2010. *SAGE Research Methods*. Web. 19 Apr. 2019, doi: 10.4135/9781412961288.

Accordingly, we pose the following question: **"Does Kalichman and Rompa's Sexual Compulsivity Scale, in fact, solely measure sexual compulsivity?"** This directly relates to KBMM's primary business objective which is to ensure that the SCS truly only measures sexual compulsivity so that our clients can accurately predict which employees are more likely to engage in sexual harassment and other forms of inappropriate sexual behavior. By answering the research question, we will be able determine if this is the best and most relevant survey to use to help clients identify and monitor "bad" prospective or current employees — in other words, sexually compulsive individuals.

Considering the aforementioned question and objective, we aim to employ exploratory factor analysis to determine how many factors underlie the data collected from the survey. We hypothesize that the questions are all related to sexual compulsivity only (one hidden factor) but we can use factor analysis to see if this presumptive hypothesis is reasonable. [7]

Implications

The costs of this business objective could be potentially significant. The Sexual Compulsivity Scale is based on a large body of research so if our factor analysis leads to a reduction in the number of variables (questions), we could be losing good questions. Another cost is that if our analysis is even slightly wrong, we would incorrectly identify bad employees which would be grossly unjust and could lead to serious lawsuits. On the other hand, the benefits of our work would be two-fold. First, we can help our clients and companies around the world be able to better prevent workplace harassment by proactively determining which individuals within the company are more sexually compulsive. Second, we can increase our share of the consulting market by establishing ourselves as leaders in this new field within human resources. However, even if we were to follow factor analysis methodology exactly, inherent flaws exist with the technique. Most importantly, interpretations of factor analysis results rely on assumptions and theories which can not be proven. In other words, there will always be a degree of uncertainty and bias in our conclusions.

Data Background

Data Understanding

The data is pulled from openpsychometrics.org. The dataset includes the survey results of an online version of the Sexual Compulsivity Scale questionnaire, created by Kalichman and Rompa in 1995. The data is only composed of results from individuals who indicated that their results were accurate and suitable for research which constitutes 79% of all survey responses. The survey is 10 questions long; the questions are listed below:

Q1. My sexual appetite has gotten in the way of my relationships.
Q2. My sexual thoughts and behaviors are causing problems in my life.
Q3. My desires to have sex have disrupted my daily life.
Q4. I sometimes fail to meet my commitments and responsibilities because of my sexual behaviors.

---

[7] Kalichman & Rompa. *Sexual Compulsivity Scale*. 1995. (Data file and code book). Openpsychometrics.org. Accessed April 5, 2019.

Q5. I sometimes get so horny I could lose control.
Q6. I find myself thinking about sex while at work.
Q7. I feel that sexual thoughts and feelings are stronger than I am.
Q8. I have to struggle to control my sexual thoughts and behavior.
Q9. I think about sex more than I would like to.
Q10. It has been difficult for me to find sex partners who desire having sex as much as I want to.

Responses to the questions take values 1-4 (1=Not at all like me, 2=Slightly like me, 3=Mainly like me, 4=Very much like me), these variables are treated as numerical. There is also a score column which is a sum of responses to all questions, this is treated as a quantitative variable. There is a gender column which takes values 0-3(1=male, 2=female, 3=other; 0=not answered) and is treated as a categorical variable. The column age indicates age in years and is treated as a quantitative variable. Ages below 14 are removed.

Data Preparation
Our data, as previously described, contains 10 question variables, a total score of all the questions, the gender and the age of each examinee. Upon further investigation of our data, we saw that there were two outlier ages of 999 and 123. We decided to remove these observations. We also found 194 missing values which indicate non-responses to the survey questions. These missing values were found across 133 observations which were removed to ensure a complete data set. Apart from this removal of observations, we did not do any other data cleaning.

One limitation of our data is that there is no grade on how truthful each examinee was on the survey. Often times, surveys include a question that allows each person to grade how honest they were on the survey. This can help to remove dishonest observations. Our data does not contain this honesty metric. Another limitation is our data appears to be obtained in July of 2012 so our data is somewhat out-of-date. Sexual Compulsivity surveys could have changed since then which could result in misleading conclusions.

Data Visualization & Exploratory Data Analysis
Our 5-number summary (see Appendix A) conveniently gives us the minimum, Q1, median, Q3, and maximum of our data. Questions 4 and 6 have the lowest and highest means with 1.9 and 3.1 respectively. This tells us that while people typically think about sex in the workplace, they slightly let their sexual behavior distract them from completing their work, which, this information may be useful to us if we decide to look into investigating, say, the performance quality of a given workplace.

A bar chart visualizes the spread of our data. Our chart (see Appendix B) depicts the spread of the survey questions. The spreads of each question are relatively even with the exception of the spreads of Questions 4 and 6, which disproportionately favored the ratings of "Not at all like me" and "Very much like me" respectively.

To determine the specific number of factors in our exploratory factor analysis, we graphed a Scree Plot (see Appendix D). The Scree Plot maps the eigenvalues of our factors on the y-axis, with the number of components along the x-axis. The point where the slope of the curve is clearly leveling off (the "elbow) indicates the number of factors that should be generated by the factor analysis. In other words, we used the Scree Plot to assess the optimum number of factors to take into account in our analysis, which happens around the "elbow", the point at which the function plateaus. With all this in mind, we extract our components that are above the cutoff line (1), the components below this cutoff contribute little to our solution. We only have 1 factor that satisfies all of these conditions, so we will only consider 1 factor in our EFA.

A correlation number chart (see Appendix C) was developed to visualize the correlations of all of the variables in our study. We see that all variables had a correlation above 0.3, satisfying the assumption of EFA that required that all variables have at least 0.30 correlations amongst each other. Perhaps worth noting, Questions 7 and 8 had the highest correlation with a correlation of 0.71, which were closely related to self-control of one's sexual thoughts and behavior. The lowest correlation was among Questions 9 and 10, with a correlation of 0.32.

There is no form of cross validation for exploratory factor analysis.

Results and Discussion:

Assumptions

Prior to running the analysis, we checked for the suitability of our data by verifying it through a number of assumptions. First, we checked to see if our data was metrical. This was found to be true of our data as our values are based on a numeric scale. We then checked to see if our sample size was large enough to yield reliable estimates of the correlations between our research variables (which are the survey questions). The standard rule of thumb is to have a sample size greater than 200. With a sample size of 3,241, we considered our sample suitably large for EFA. We also need the correlations between our variables to be at least .3 to be able to identify coherent factors. We see in Appendix C that all of the correlation values between our research variables have a correlation of at least .3. We also need a homogenous sample. Homogeneity can be checked through reliability analysis. We can test this by looking at the Coefficient Alpha (Cronbach's Alpha) and the Average Split-Half Reliability of our data. This tests the internal consistency, or reliability, of our data. For our Coefficient Alpha and our Average Split-Half value, we need a value of at least .8. Our data produced both an alpha and an Average Split-Half value of .90 so our data is homogenous. Lastly, we must check for the presence of outliers, which, we found there to be none in our data.

Evaluation and Results

We first found the Eigenvalues of each factor which, as mentioned in the methodology, are significant if greater than 1. Based on the calculated Eigenvalues, we only found one factor with a value (5.2299) greater than one. This indicates that there is only one hidden factor that is statistically significant and that has high variation in our dataset (see Appendix F). This initial conclusion addresses our business objective since these results indicate that the survey *solely and directly* measures sexual compulsivity . With only one factor, we could argue that the one factor is in fact sexual compulsivity. However, practically speaking, our model may also have two (0.9115) or three (0.8182) hidden factors as their Eigenvalues are extremely close to 1. These results can also be seen in our Scree Plot (see Appendix D). Accordingly, the possibility of more hidden factors being involved would contradict our business objective.

After building an EFA model using only 1 latent factor, we looked at the factor loadings to gauge how each question (Q1-Q10) correlates with the factor (MR1). As we expected, all of the factor loadings (associations) were positive and range from fairly strong to very strong (see Appendix E). For instance, Q8 has a factor loading of 0.805 which means it is strongly positively correlated with MRI, which we hypothesised to be sexual compulsivity. On the other hand, Q10 had the smallest factor loading (0.529) which means it is only somewhat positively correlated to MR1. The aforementioned findings regarding the factor loadings indicate that each question is directly and strongly correlated with the underlying factor. For our company, this means that each survey question is actually related to sexual compulsivity which is precisely what we are looking for in an effective survey. [8]

We were also able to obtain factor scores from our EFA model which indicates each individual observations' embodiment of MR1 (assumed to be sexual compulsivity).[9] For instance, observation 2405 has a factor score of 1.97 which means that this examinee has a rather high sexual compulsivity. Based on our primary business objective, this individual, due to their high sexual compulsivity (high factor score) would be more likely to commit sexual harassment in the workplace. Therefore, we would recommend our clients to monitor this employee.

With our initial EFA model finalized, we proceeded to evaluate the absolute fit of our single-factor model by using three different absolute fit statistics. We found our model to have a significant result with a chi-square value of 425.28, a Tucker-Lewis Index of less than the cutoff of 0.9 (0.844) and a Root mean Square Error of Approximation greater than the cutoff 0.05 (0.125) which all indicate that our model does not have a good fit. Accordingly, we cannot conclude that our model fits adequately nor that the discrepancy between observed and expected data is small. This suggests that our one-factor model may not actually be representative of the data set and has weak predictive power. For commercial purposes, this is concerning as it means we cannot definitively say that the Sexual Compulsivity Scale only measures the one hidden factor of sexual compulsivity.[10]

Due to the lack of absolute fit, we looked at relative fit to see how our one-factor EFA model compared to models with a larger number of factors. We first tested two- and three-factor models since they had fairly high Eigenvalues, as mentioned earlier. Both of these models had a lower BIC (196.17 and 88.38) than our one-factor model (673.04) which suggests they have better relative fit. However, we believe that the BICs are artificially deflated simply due to the increase in factors. Accordingly, we conclude that a one-factor model is still appropriate — due to it

[8] Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r
[9] "What Is a Factor Score?" *Psychology.okstate.edu*, psychology.okstate.edu/faculty/jgrice/factorscores/fs_q.html.
[10] Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r

uniquely having a statistically significant Eigenvalue — but we still acknowledge that there is the possibility of up to 5 underlying factors being present. [11]

Relating back to our business question of "Does Kalichman and Rompa's survey in fact solely measure sexual compulsivity?", we can only conclude that, given the statistical evidence we have, the survey likely *only has one underlying factor* which can naturally be assumed, but not guaranteed, to be sexual compulsivity. Therefore, we can communicate to our clients that our analysis indicates that the Sexual Compulsivity Scale is reasonably effective and achieves its purpose. While there is certainly room for further analysis, the Scale should be able to gauge our clients' sexual compulsivity and thus predict which employees are more likely to engage in sexual harassment.

Recommendations

Based on the results from our exploratory factor analysis, we would recommend our clients to use Kalichman and Rompa's survey as an accurate and direct measure of their employees' sexual compulsivity. Using this survey should help companies identify which of their employees are more prone to engage in inappropriate sexual behavior and ensure for better security and managing practices. A safer workplace tends to be a more productive one, which should, in turn, lead businesses to achieve both their social and financial goals.

Specifically, we would recommend that our clients focus on the responses to Questions 2, 3, 5, 7, and 8 as these items are most strongly related to sexual compulsivity. Due to their high factor loadings, the answers to these questions are highly correlated with sexual compulsivity and therefore are more effective measures of that trait. Companies should place extra weight on those items, especially in unclear situations such as if an individual's survey responses are inconsistent.

However, due to the uncertainty of the validity of our final model, we do not recommend our clients to *only* use the Sexual Compulsivity Scale. Due to the ambiguity of EFA, we cannot be certain that our sexual compulsivity theory is correct. Accordingly, clients should consider other things outside of the survey (e.g. employee history, references, criminal record, etc.) before drawing any final conclusions about a particular person. In summary, while corporate HR departments should use the Sexual Compulsivity Scale for employee evaluation, they should not solely rely on just the survey.

Looking ahead, statistical techniques like Confirmatory Factor Analysis (CFA) — which we mentioned would be useful earlier — and Principal Component Analysis (PCA) can be applied to further our analysis and develop more precise conclusions.

First of all, CFA can be used to verify the factor structure of our observed variables that was found through EFA. As opposed to EFA, this type of factor analysis is confirmatory, in that it enables us the ability to test a hypothesis regarding the structure or number of dimensions underlying our set of predetermined variables. For instance, if we believe there to be, say, two

[11] Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r

dimensions underlying our data and want to verify this, then a CFA can be done to do this. If we were to run a CFA on the same data as we ran our EFA, then we will get a good fitting model. However, doing this would not utilize the full research benefits of a CFA, since, to reiterate, the point of a CFA is to show that a factor structure fits data it has never seen before. In a general application of the CFA, we can specify the number of factors required in our data and which measured variable is related to which underlying variable.[12][13]

If we desire to further our data feature extraction methods, then a comprehensive PCA may be another method to consider. Essentially, PCA combines our variables in a way where the "least important" variables are dropped while still retaining the most valuable ones. Through the PCA, these re-modified variables all become independent of one another, opening the door for additional methods like linear regression, etc… A PCA is approached with the mindset of answering how we take all of the variables we've collected to focus on only a few of them. By making efficient reductions, there will be fewer relationships between variables to consider and, as a result, we are less likely to overfit whatever linear model we decide to pursue doing in the future.

[12] Child, D. (1990). The essentials of factor analysis, second edition. London: Cassel Educational Limited.
[13] Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis, Psychometrika, 34, 183-202.

Works Cited

A. Brussow, Jennifer. "Factor Analysis in R." Datacamp, C. Ismay & B. Robins. Datacamp.https://www.datacamp.com/courses/factor-analysis-in-r
B. Child, D. (1990). The essentials of factor analysis, second edition. London: Cassel Educational Limite
C. Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis, Psychometrika, 34, 183-202.
D. Kalichman & Rompa. *Sexual Compulsivity Scale*. 1995. (Data file and code book). Openpsychometrics.org. Accessed April 5, 2019.
E. Salkind, Neil J. *Encyclopedia of Research Design.* Thousand Oaks, CA: SAGE Publications, Inc., 2010. *SAGE Research Methods*. Web. 19 Apr. 2019, doi: 10.4135/9781412961288.
F. Statistics Solutions. (2013). Confirmatory Factor Analysis [WWW Document]. Retrieved from http://www.statisticssolutions.com/academic-solutions/resources/directory-of-statistical-analyses/confirmatory-factor-analysis/
G. Statistics Solutions. (2013). Exploratory Factor Analysis. Retrieved from https://www.statisticssolutions.com/factor-analysis-sem-exploratory-factor-analysis/

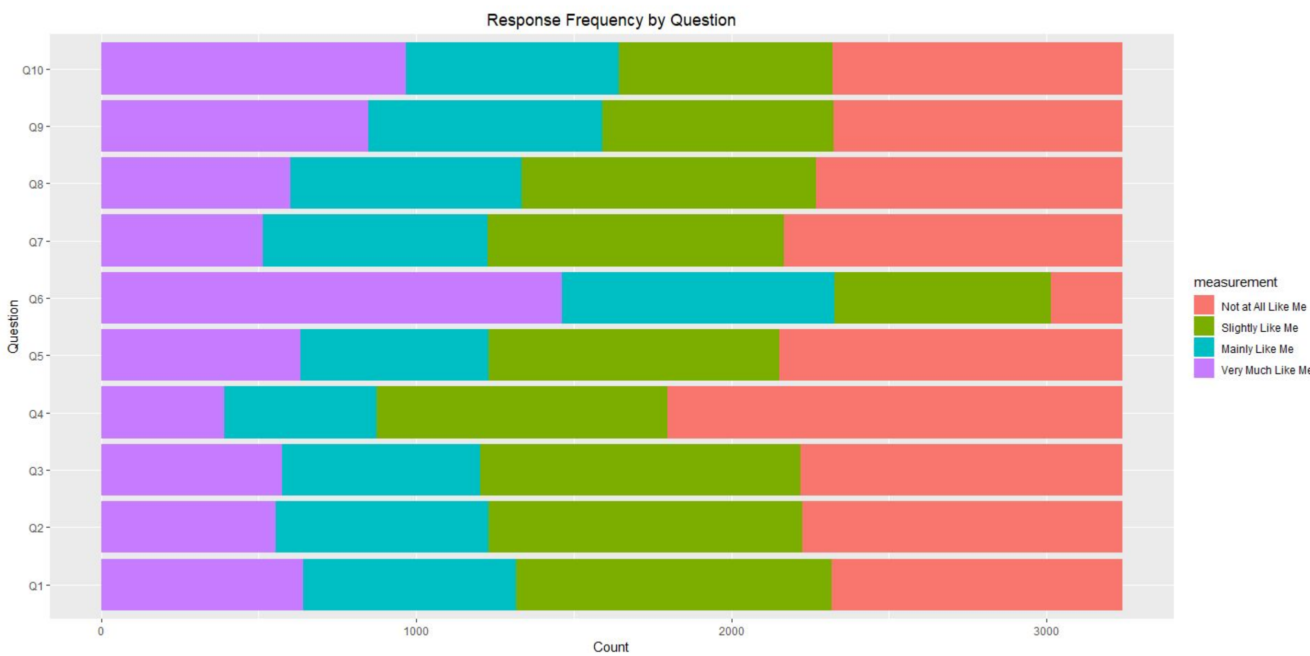
Appendix

Appendix A

	n	mean	sd	median	min	max	Q1	Q3	missing_obs
1	3241	2.319346	1.087879	2	1	4	1	3	FALSE
2	3241	2.236655	1.071731	2	1	4	1	3	FALSE
3	3241	2.233570	1.079007	2	1	4	1	3	FALSE
4	3241	1.944462	1.036376	2	1	4	1	3	FALSE
5	3241	2.238198	1.115562	2	1	4	1	3	FALSE
6	3241	3.099352	0.965454	3	1	4	2	4	FALSE
7	3241	2.204875	1.068916	2	1	4	1	3	FALSE
8	3241	2.297748	1.086061	2	1	4	1	3	FALSE
9	3241	2.469608	1.157271	2	1	4	1	4	FALSE
10	3241	2.521135	1.189304	3	1	4	1	4	FALSE

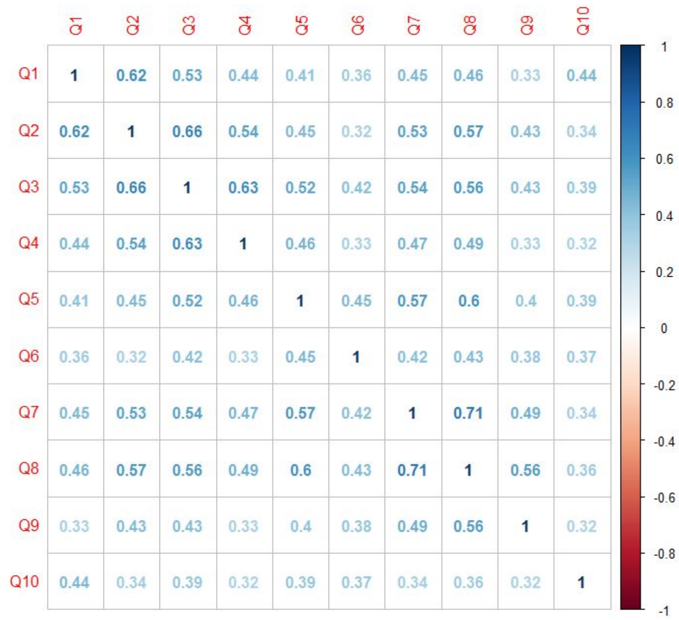
*5-number summary of the relevant variables of our dataset (the ten questions from the survey)

Appendix B



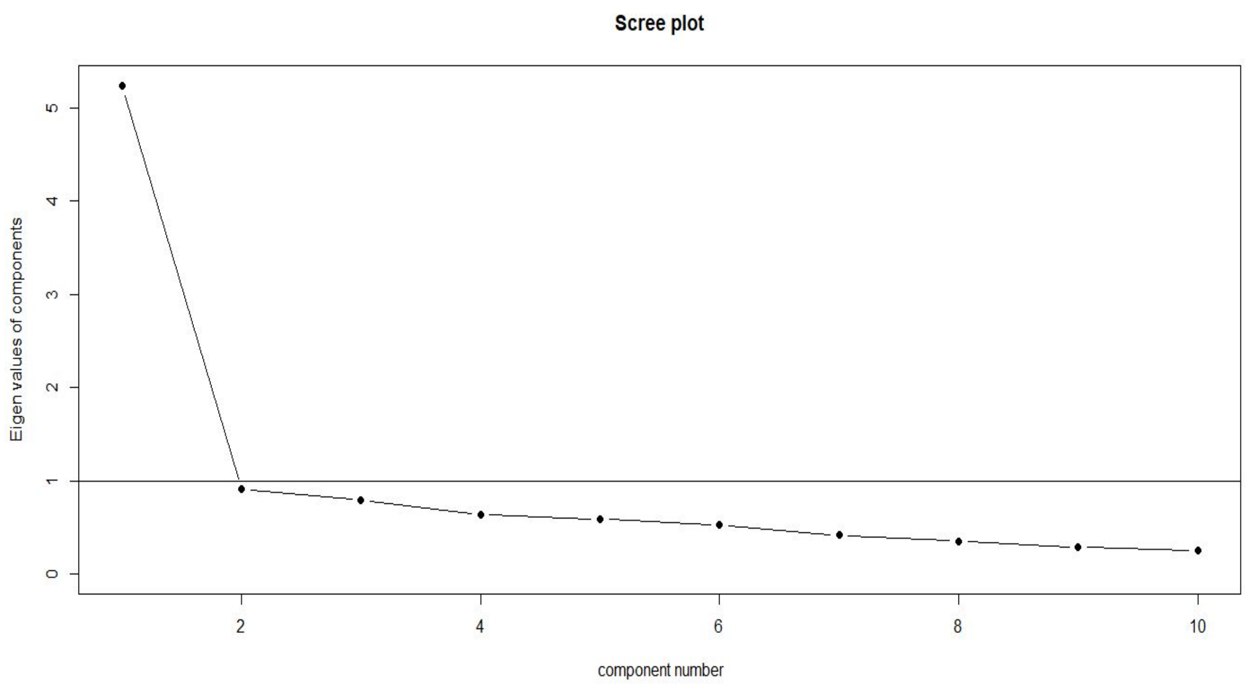
*Stacked bar chart displaying the response frequency of each question in the survey

Appendix C



*Correlation matrix used to test covariance of our variables to satisfy assumptions of Exploratory Factor Analysis

Appendix D



*Scree plot which indicators how many hidden factors in the data set are statistically significant; if the Eigenvalue is greater than 1, than that factor is significant

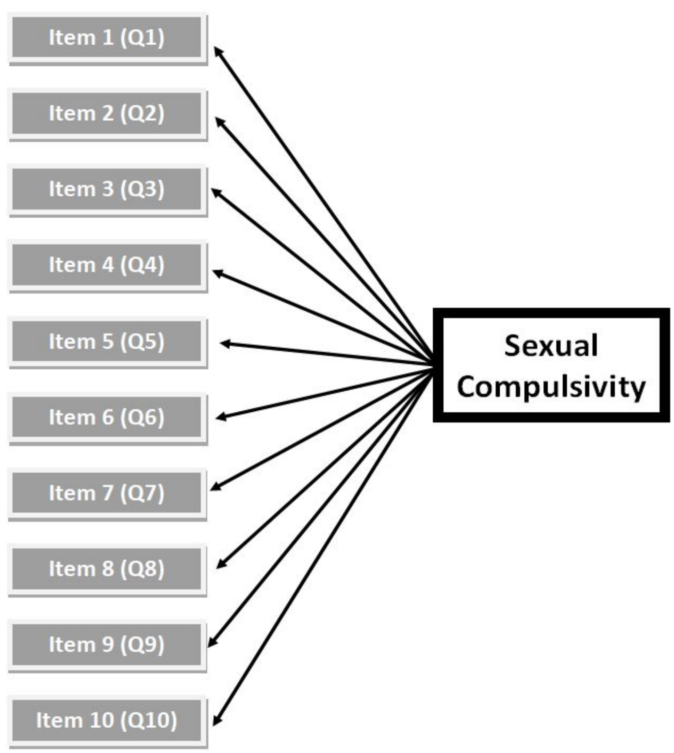
Appendix E

```
> EFA_model$loadings

Loadings:
    MR1
Q1  0.650
Q2  0.746
Q3  0.793
Q4  0.673
Q5  0.707
Q6  0.554
Q7  0.757
Q8  0.805
Q9  0.612
Q10 0.529
```

*R output of factor loadings which indicates how well each item (Q1-Q10) correlates with MR1 (the hidden factor); greater loading means stronger correlation

Appendix F



*Final factor structure that represents our data construct based on the results of EFA; only one hidden factor was found and all items have some positive relationship with it