Principal Component Analysis Based on Self-Reported Humour Styles Questionnaire

Paul Long
School of Computing
National College of Ireland
Dublin, Ireland

Table of Contents

Word Count: 1404

Principal Component Analysis Based on Self-Reported Humour Styles Questionnaire

Introduction

This report is based on an undertaking of data analysis, performed on data collected from a 'Humor Styles Questionnaire', hosted on Openpsychometrics.org, as part of the study: "Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003)[1]. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. Journal of Research in Personality, 37, 48-75."

Data Description

The purpose of the original study was to study what people's humour styles were. These were categorised into two sets of factors – enhance or diminish relationships and whether the jokes would have a positive or negative impact on social setting – factorising into: affiliative, self-enhancing, aggressive and self- deprecating. This questionnaire contained 32 questions with self-reported answers ranging from 1 to 5, 1 being Totally Disagree, 2 being Somewhat Disagree, 3 being Neutral, 4 being Somewhat Agree and 5 being Totally Agree. Each question related to a specific factor – affiliative, self-enhancing, aggressive and self-deprecating, with each factor having 8 questions, whose mean scores rounded to 1 decimal place were calculated for each respondent. The groups were listed as questions 1, 5, 9, 13, 17, 21, 25 and 29 for affiliation, questions 2, 5, 10, 14, 18, 22, 26 and 30 for self enhancement, etc. There were 1,071 respondents.

Background

This report refers to a study done on the same data where factor analysis – specifically principal component analysis, was carried out on the data to investigate the fundamental groupings of the questionnaire. After understanding the principal components outputted by the analysis, these factors were then compared to the four original groupings to see if there was a significant overlap between them. This study was to ensure that there was significant separation between the four humour styles in the previous report. That is, outside of the assigned groupings, the scores from each question would cluster together naturally to recreate the prior assigned groupings. Following this, looking at whether the questions from each factor were strongly positively or negatively correlated had an impact on the calculations made by the study.

Initially, the analysis computed was exploratory factor analysis, so the components were calculated on the basis of eigenvalues – specifically eigenvalues greater than 1, to get a general understanding of what the components outputted would be. After this analysis was completed, there was some confirmatory factor analysis conducted, whereby the four factors accounting for the largest variance were compared with the four internal groupings of the study for comparison.

Data Alteration

The data didn't require much cleaning, as it contained only the numeric responses and the score for each of the four humour types. The only alteration that was undertaken was to reorder each row so that it was grouped with the other questions relating to that humour type. This aided in subdividing the correlation matrix into blocks of humour styles, thus making it easier to compare the correlation matrices to gain some further insight.

Assumptions

The data was tested to ensure that some preconditions were met for factor analysis. Bartlett's Test for Sphericity was run on SPSS to ensure that the correlations between variables were not zero. It was found that the correlations were not all zero $F(31, 1070) = 10915$, $p= 0.000$. It was proceeded on the basis that the data has some significant correlation in it.

Results

Principal Component Analysis was then undertaken using SPSS. The correlation matrix had it's strongest clustering within it's humour styles, with there being little to no clustering outside of this. This was then verified by computing a correlation heatmap using seaborn on Python. Within each cluster, there appeared to be a chequerboard structure, indicating that there were both strong positive and strong negative correlations within each grouping. Outside of the humour groupings, the highest correlations seemed to be between affiliation, and self-deprecating and self-enhancing, as well as between aggression, and self-deprecating and self-enhancing.
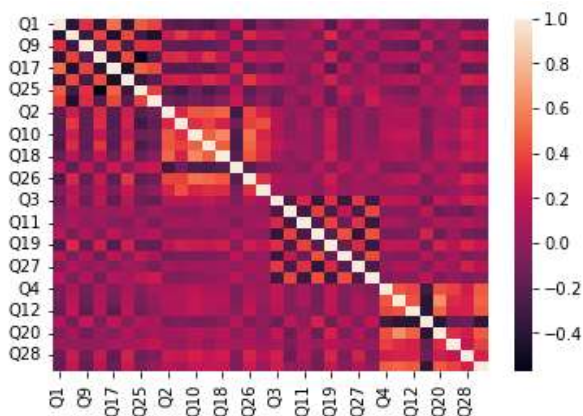

The factor threshold had an Eigenvalue level of 1. There were 7 factors that were produced as a consequence of this. However, one of these factors had an Eigenvalue of 1.003. This was significant as 7 factors did not correspond to the inflection point of the Scree Plot. The point of inflection of the Scree plot was at 6 factors. However, given that the 7th factor met the threshold required, it remained as a factor to be studied.
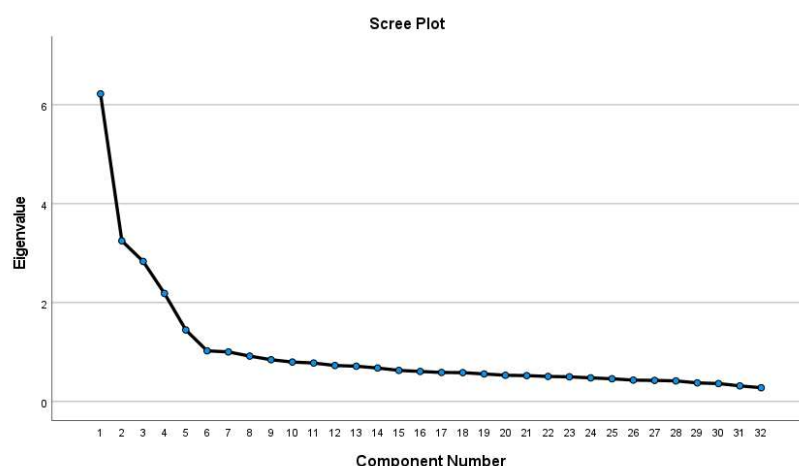
The seven factors produced accounted for roughly 56% of the total variance, with the first four factors accounting for roughly 42% of the variance.

The correlation matrix between each question and the seven factors were examined. It was clear that there was clustering, and that the factors had significant relationships with the original groups. Specifically, it appeared that each factor had a significant relationship to one of the humour styles groups.

The analysis of the factor correlations appeared to show that factor 1 was related to Affiliation, factor 2 was related to self-deprecating, factor 3 was related to aggression and factor 4 was related to self-enhancing. However, factor 1 happened to be more related to self-enhancing than factor 4 did, having a higher level of correlations than the rest of the factors. And factor 4 happened to have quite a high correlation with affiliation comparatively, having a lower level of correlations than the rest of the factors, despite both of their strongest correlations being to the groups listed above. The other three factors did not appear to have a particularly meaningful impact, with correlations being mostly low, occasionally jumping up to a high value for one question that had weaker correlations amongst the other factors.
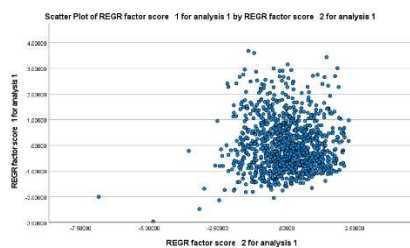
The rotated correlation matrix was then analysed. This was calculated using a varimax rotation to maximize for high and low loadings and minimize for medium loadings. This created a very clear relationship between the factors and the groups. In summary, each factor had a clear cluster associated to one of the groups – factor 1 was associated to affiliation, factor 2 was associated to self-deprecation, factor 3 was associated to self-enhancement and factor 4 was associated to aggression. It should be noted here that after the data was rotated, factors 3 and 4 switched.

This process was then conducted from a confirmatory factor analysis perspective, with the analysis set to extract four factors. Here, the factors accounted for roughly 45% of the total variance, which was greater than the summation for the largest four factors calculated previously. From this, the rotated correlation matrix showed that, again, there was a significant relationship between the factors. However here, factor 2 was associated to self-enhancement, factor 3 was associated with self-deprecation. So, factor 2 and factor 3 flipped. It was assumed that this change was due to a reordering in impact of the correlations in factors 2 and 3 when the additional 3 factors were not accounted for.

The individual questions that informed each grouping were then analysed. Particularly, the extent to which questions were positively and negatively skewed were looked at in the context of the factor-group correlation clusters. Whilst the sign for each factor was sometimes flipped across the whole list of questions, each cluster had a signs that matched the expected sign of the question in the correlation matrix. For example, affiliation had questions 1, 9, 17, 25 and 29 all having negative slants & questions 5, 13 and 21 having positive slants. Within Factor 1, the correlations of Q1, Q9, Q17, Q25 and Q29 were all positive and Q5, Q13 and Q21 were all negative. The difference here is that the signs were flipped, but the corresponding signs indicate that there is a significant relationship. Thus, each question within each factor has the same sign as it's counterpart has slant on that grouping.

Conclusion

In summary, exploratory principal component analysis was performed on the data from the humour styles self-reported survey to reduce the dimensions. It became clear that there were 4 factors that

were of relevance, but that of those factors, there were a primary four that were of major significance. Then, confirmatory factor analysis was performed to extract 4 factors. These factors clustered into groups that mirrored the groupings of the original study. Furthermore, within these groupings, each question had the same sign as it's counterpart's question's slant on the impact it had on that humour style. Thus, it was shown to a satisfactory extent that the original study's structure that was imposed upon it matched the inherent structure of the data.

Bibliography

https://www.kaggle.com/datasets/lucasgreenwell/humor-styles-questionnaire-responses?select=data.csv