Structural Equation Modeling

Julian Ibarguen

25/08/2022

Contents

0.1 Introduction

We aim to fit a Structural Equation Model (SEM) following the exercise proposed by Petri Nokelainen from the Research Centre for Vocational Education, University of Tampere, Finland. Data and documentation for this exercise can be found here.

We used the proposed data set with 447 rows and 27 columns. The respondents in both samples are staff members of Finnish polytechnic institute for higher education. The original measurement instrument (Growth-oriented Atmosphere Questionnaire, GOAQ) has 13 factors and 92-items (Ruohotie, 1996; Ruohotie, Nokelainen & Tirri, 2002), but for the purposes of this exercise Nokelainen selected 13 factors and 27 items, all consisting of a five–point Likert scale from 1 (totally disagree) to 5 (totally agree) was applied (Nokelainen, n.d.)

0.2 Evaluation of assumptions

0.2.1 Univariate distribution

Although data is ordinal, thus non-normal in nature, we check for normality assumption. We will apply z re-scaling to see how they approximate to mean 0. Will also check kurtosis and skewnewss, and plot the histograms. We reject applying Saphiro test or other normality test, as being only 5 points scale ordinal is unlikely will provide any meaningful result.

Given a normally distribute variable would have a skewness of 0 and kurtosis of 3. For this purpose we calculated a synthetic measure for skewness and kurtosis as the average between the skewness and excess kurtosis ($kurtosis - 3$). Under this synthetic measure, a perfect normal distribution would have a value of 0. For our purpose, we select those variables whose deviance from the normal distribution is lower than 0.4, according to the synthetic measure create. The selected variables were further confirmed with the histogram.

Observing table 1 and figure 1, we assumed normality for the following variables: v2, v3, v4, v15, v31, v33. The remaining variables are assumed non-normally distributed

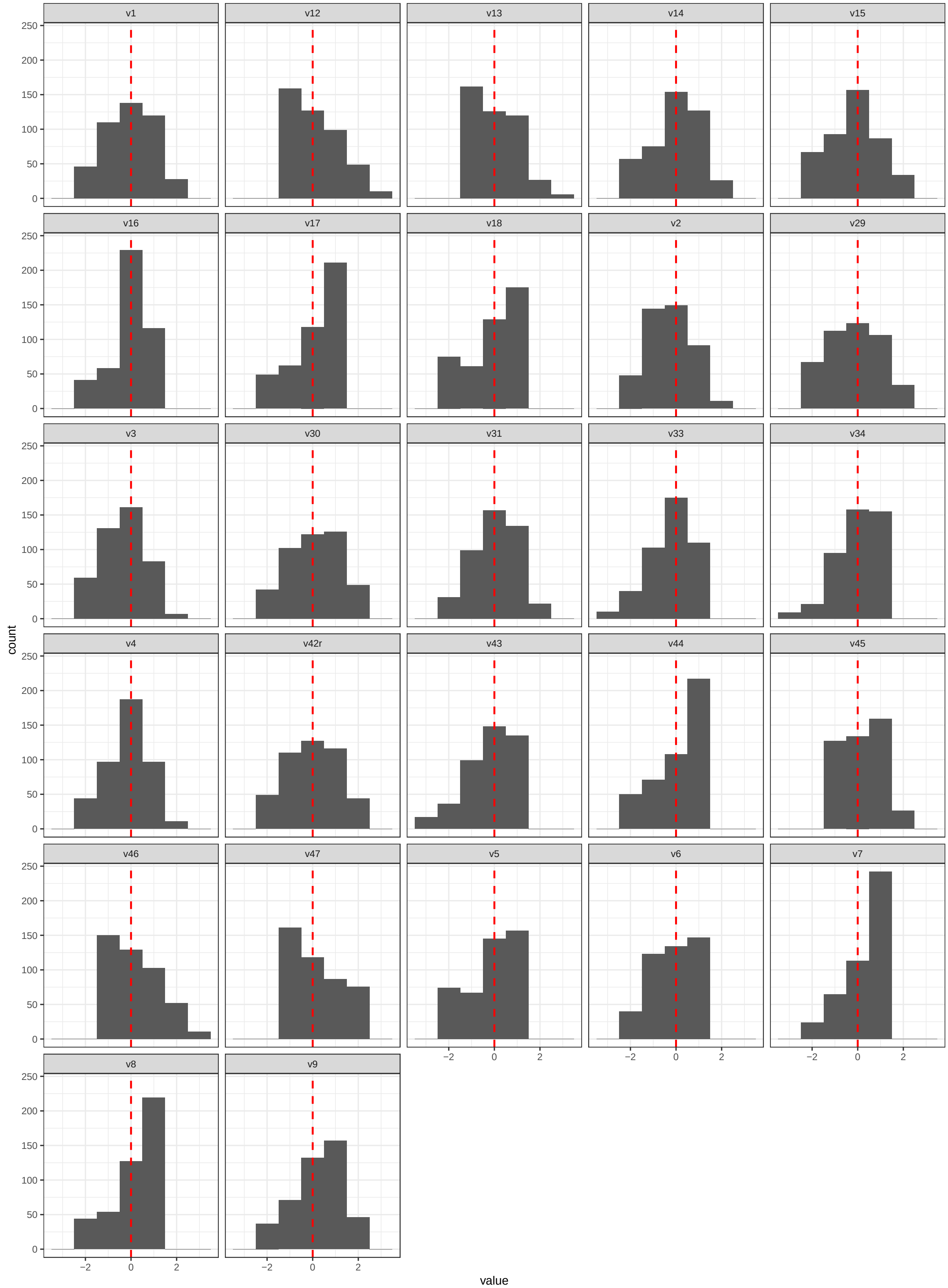


Figure 1: Histograms

Table 1: Descriptive statistics to assess approximation to normal distribution

variable	skewness	kurtosis	kurtosis_excess	avg_skew_kurt
v4	-0.203	2.592	-0.408	0.306
v3	-0.014	2.318	-0.682	0.348
v15	0.022	2.291	-0.709	0.366
v2	0.082	2.350	-0.650	0.366
v33	-0.610	2.876	-0.124	0.367
v31	-0.198	2.447	-0.553	0.375
v1	-0.072	2.239	-0.761	0.417
v43	-0.716	2.875	-0.125	0.420
v42r	-0.026	2.137	-0.863	0.444
v7	-0.496	2.600	-0.400	0.448
v9	-0.382	2.454	-0.546	0.464
v14	-0.271	2.333	-0.667	0.469
v29	0.042	2.084	-0.916	0.479
v30	-0.107	2.141	-0.859	0.483
v13	0.555	2.531	-0.469	0.512
v8	-0.419	2.368	-0.632	0.526
v34	-0.806	3.277	0.277	0.542
v5	-0.812	2.698	-0.302	0.557
v17	-0.372	2.243	-0.757	0.565
v46	0.558	2.384	-0.616	0.587
v12	0.607	2.431	-0.569	0.588
v18	-0.164	1.980	-1.020	0.592
v44	-0.341	2.133	-0.867	0.604
v6	-0.743	2.521	-0.479	0.611
v45	0.522	2.299	-0.701	0.612
v16	-0.578	2.295	-0.705	0.641
v47	0.596	2.271	-0.729	0.663

0.2.2 Bivariate distribution

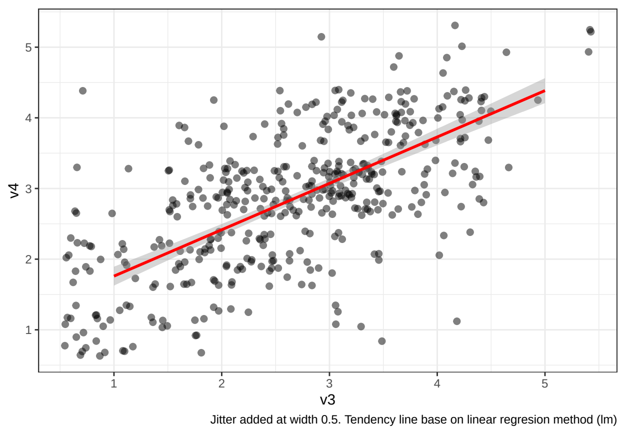
Based on the univariate analysis, we select three different variable combinations: 1) both are assumed normally distributed; 2) one is assumed normally distributed, but the other one no; 3) neither fo them is assumed normally distributed.

In figure 2, we can observe the linear relationship between the 3 different combination. When we assume normality in both variables, we observe linear relationship, which decreases for the second combination (one variable assumed normal and the other no), and the relationship disappear for the third combination

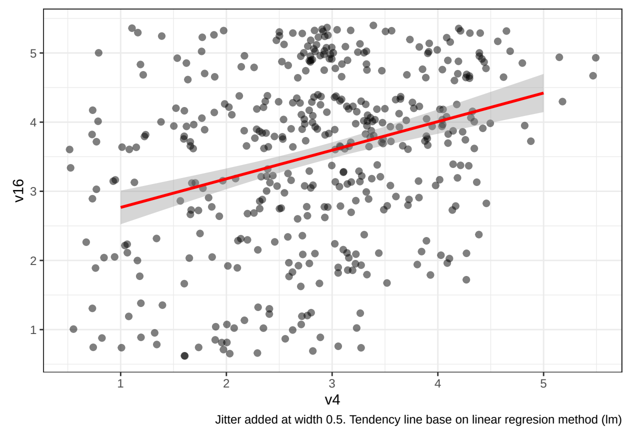
0.2.3 Correlation matrix

We observe now the correlation matrix between all 27 variables. For SEM and other factorial type of analysis, the correlation coefficient should be between ± 0.3 and ± 0.8. Too low correlations indicate weak inter-item dependency, too high correlations might indicate multicollinearity (Nokelainen, n.d.).

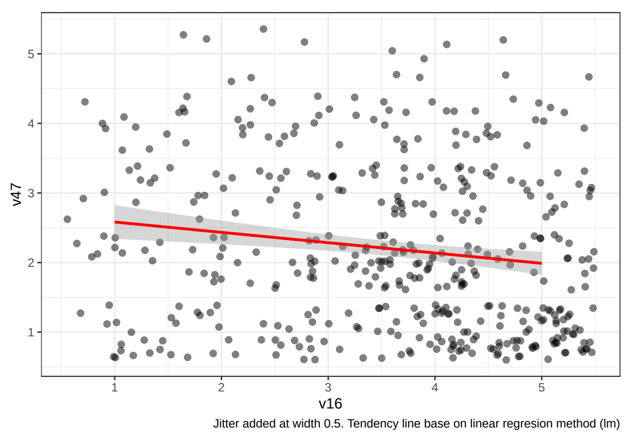
The highest correlation coefficient (r) in absolute terms was betwee v18, v17 with \$r = \$0.8344792 and R^2 = rmax_r2. The lowest correlation coefficient (\$r\$) was between v45, v7 with \$r = \$-0.3372552 and \$R^2\$ =r min_r2. Therefore, we cna conlcude that our variables are suitable for a SEM model



(a) Two variables assumed normally distributed



(b) One variable assumed normally distributed



(c) No variable assumed normally distributed

Figure 2: Bivariate distributions

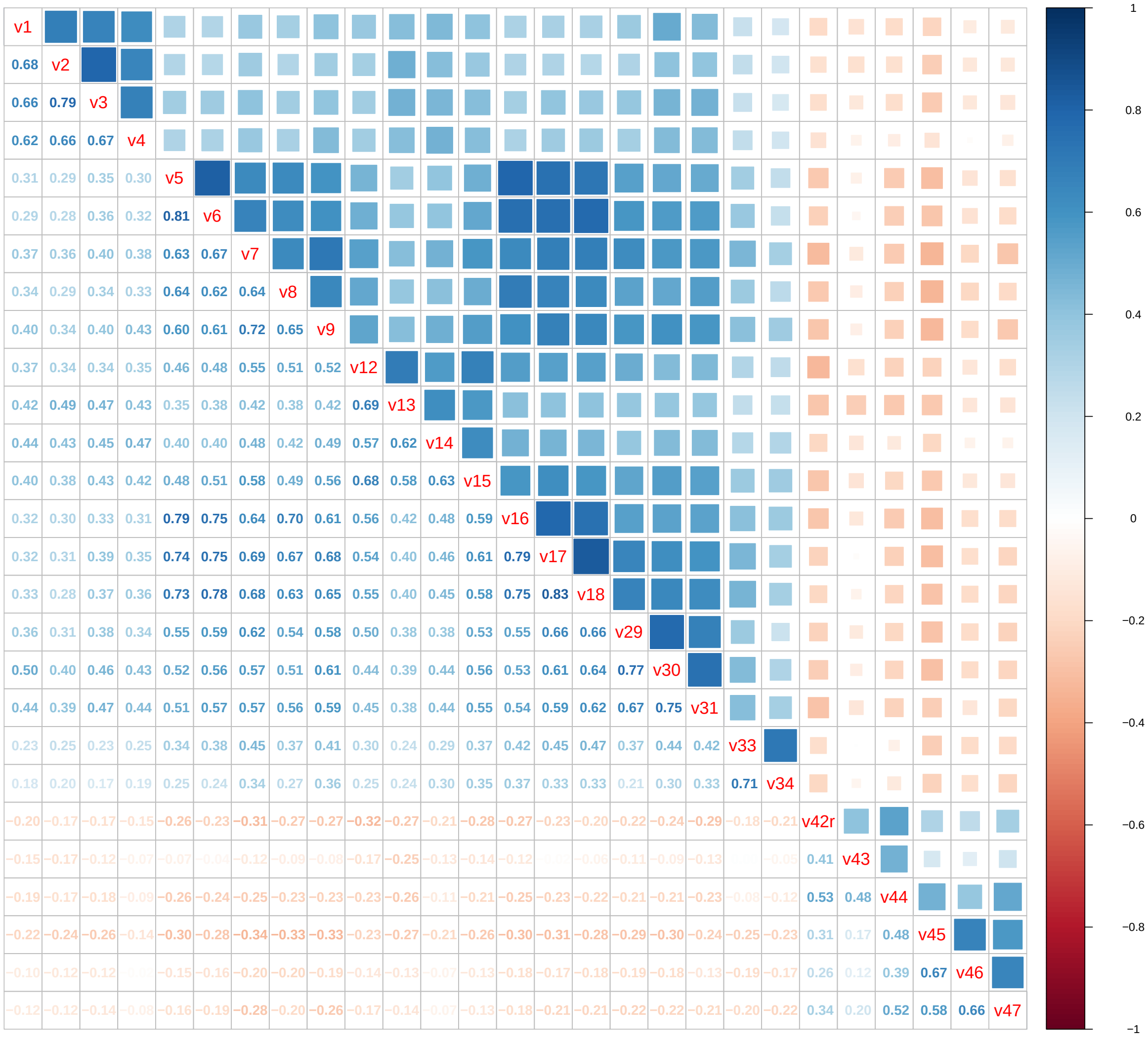


Figure 3: Correlation matrix

0.3 Path analysis

To perform the path analysis we make use of the 13 factors available in our data.