Structural Equation Modeling

Julian Ibarguen

25/08/2022

Contents

0.1 Introduction 1
0.2 Evaluation of assumptions 1
0.2.1 Univariate distribution 1
0.2.2 Bivariate distribution 3
0.2.3 Correlation matrix 3
0.3 Path analysis 6

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 were assumed non-normally distributed

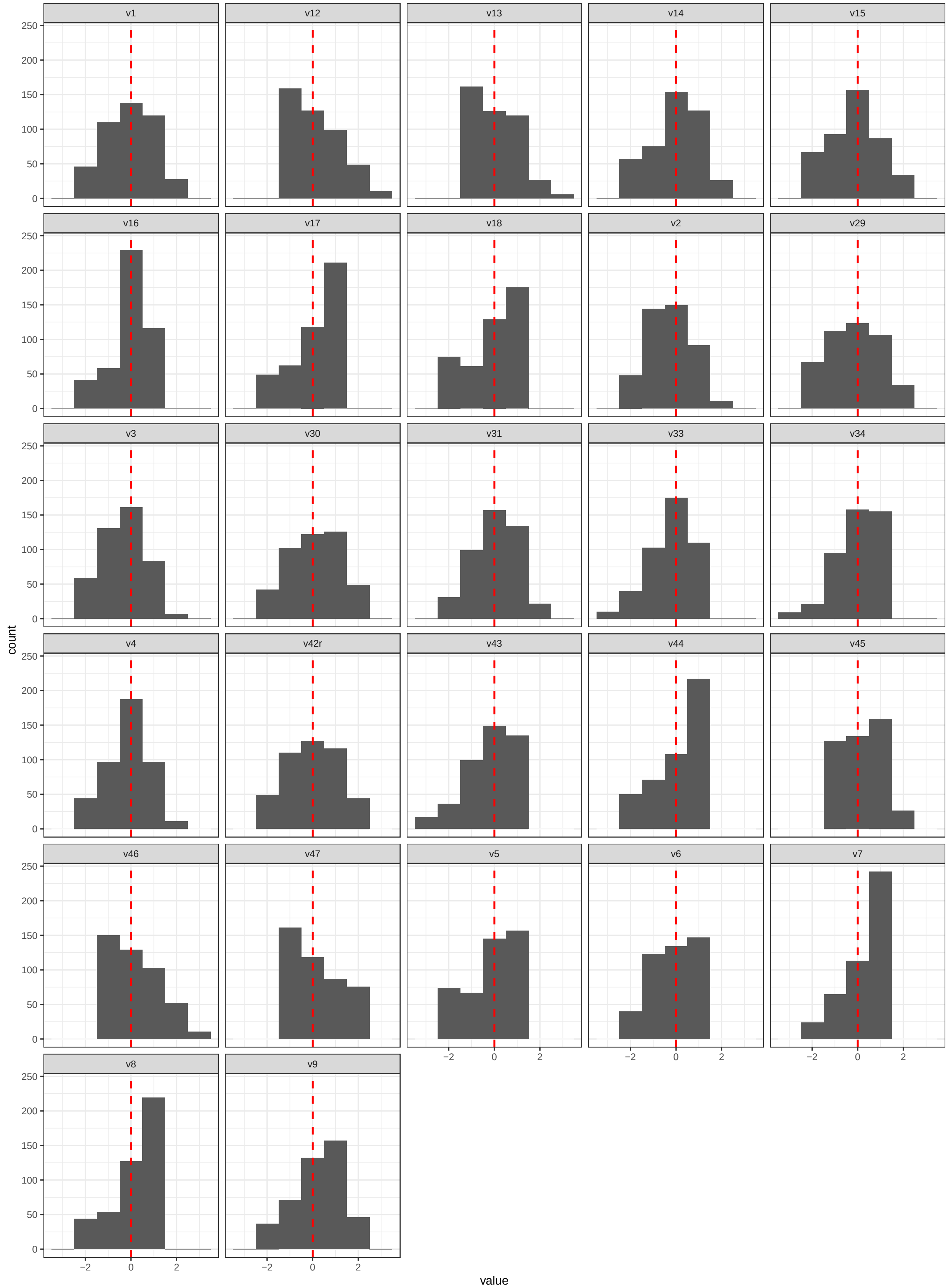


Figure 1: Histograms

Table 1: Descriptive statistics to assess approximation to normal distribution

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

0.2.2 Bivariate distribution

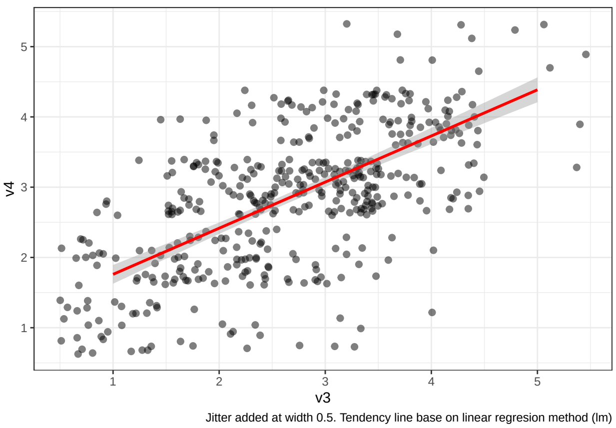
Based on the univariate analysis, we select three different variable combinations: 1) both are assumed normally distributed; 2) one is assumed normally distributed, but the other one no; 3) neither fo them is assumed normally distributed.

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

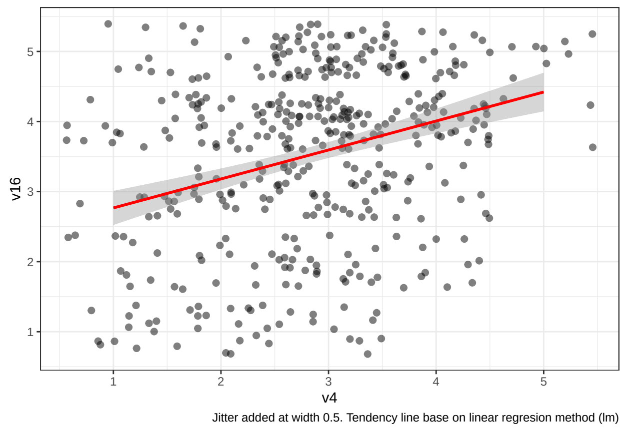
0.2.3 Correlation matrix

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.).

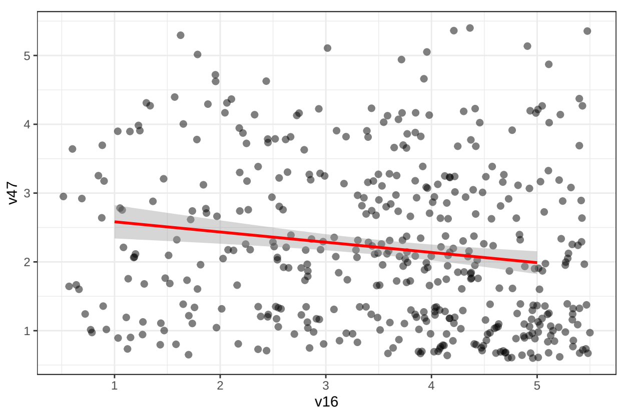
The highest correlation coefficient (r) in absolute terms was betwee v18, v17 with $r = $0.8344792 and R^2 = `rmax_r2. The lowest correlation coefficient ($r$) was between v45, v7 with $r = $-0.3372552 and $R^2$ =`r `min_r2`. Therefore, we cna conlcude that our variables are suitable for a SEM model



(a) Two variables assumed normally distributed



(b) One variable assumed normally distributed



(c) No variable assumed normally distributed

Figure 2: Bivariate distributions

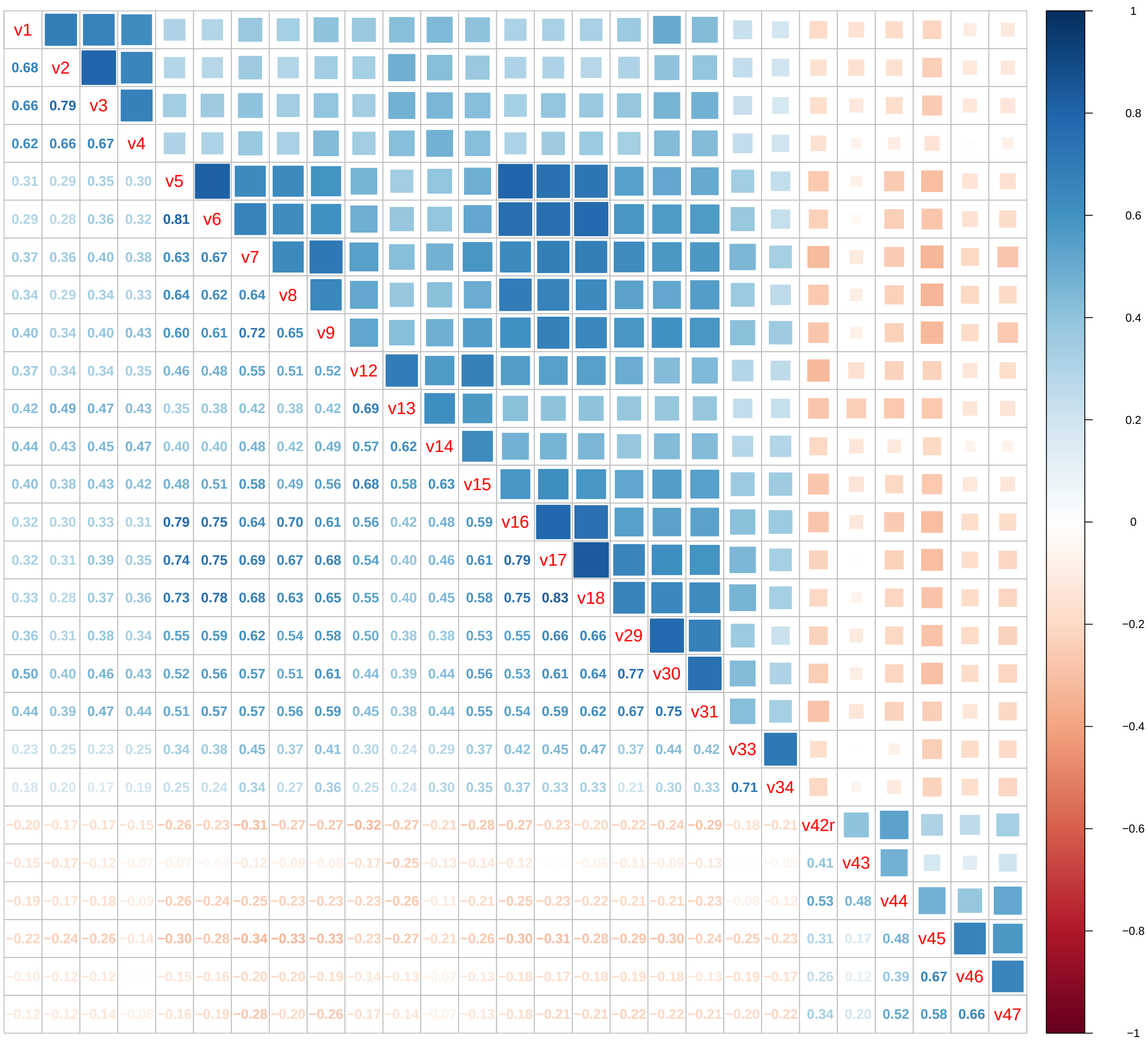


Figure 3: Correlation matrix

0.3 Path analysis

To perform the path analysis we make use of the 13 factors available in our data and we fit the following model:

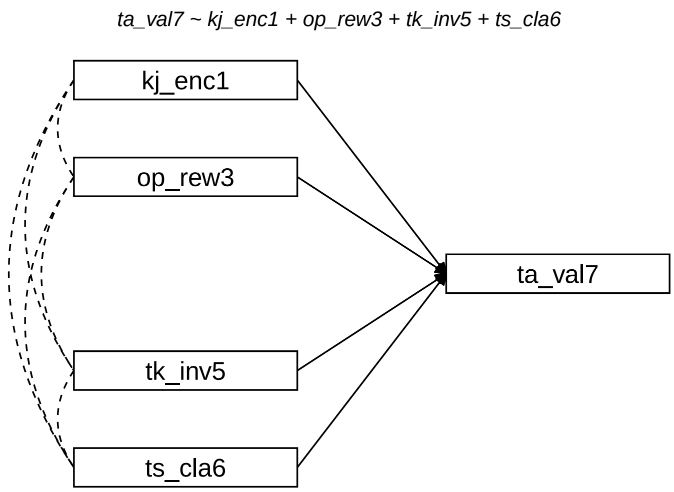


Figure 4: Model path

where:

Table 2: Labels for factor variables

variable_label	variable_name
1. Encouraging leadership	kj_enc1
2. Strategic leadership	sj_str2
3. Know-how rewarding	op_rew3
4. Know-how developing	ok_dev4
5. Incentive value of the work	tk_inv5
6. Clarity of the work	ts_cla6
7. Valuation of the work	ta_val7
8. Relationship-based learning	vk_rel8
9. Team spirit	rh_tes9
11. Psychical stress of the work	tr_psy11
12. Build-up of work requirements	tv_bui12
13. Commitment to work and organization	si_com13

Tghe summary of the model was the following:

```
## lavaan 0.6-12 ended normally after 16 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                         6
##
##                                                   Used       Total
##   Number of observations                           443         447
```

```
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   ta_val7 ~
##     kj_enc1           0.381    0.047    8.086    0.000    0.381    0.446
##     op_rew3           0.087    0.042    2.056    0.040    0.087    0.092
##     tk_inv5           0.228    0.040    5.697    0.000    0.228    0.221
##     ts_cla6           0.080    0.043    1.864    0.062    0.080    0.092
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ta_val7           0.959    0.140    6.842    0.000    0.959    1.093
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .ta_val7           0.365    0.025   14.883    0.000    0.365    0.473
##
## R-Square:
##                    Estimate
##     ta_val7           0.527
```

From the results of the model we obtain the following conclusions:

- How much dependent variable variance the four independent variables predict? $R^2 = 0.527$.
- Order the IVs in the following rows (best predictor comes first):

    1. The first (strongest) predictor for Valuation of the work is *1. Encouraging leadership* $r = 0.381$
    2. The second predictor for Valuation of the work is *5. Incentive value of the work* $r = 0.228$
    3. The third predictor for Valuation of the work is *3. Know-how rewarding* $r = 0.087$
    4. The fourth predictor for Valuation of the work is *6. Clarity of the work* $r = 0.08$

- Select Unstandardized estimates and complete the following sentences:

    - When *Encouraging leadership* goes up by 1, Valuation of the work goes up / down by ______.
    - When *Know-how rewarding* goes up by 1, Valuation of the work goes up / down by ______.