Psychometrics - Assignment 02

Joshua Damm

2024-10-18

1 Introduction

Understanding and quanifiying life satisfaction has been a main objective of psychometric research, as it is thought to be a key indicator of well-being, happiness, and resistance towards phyiscal and psychiatric diseases. In this study, we used confirmatory factor analysis on data coming from a survey on the social integration of young adults (n = 1245) to investigate changes in life satisfaction over a span of 4 years from 1987 to 1991, using the same questionnaire to assess life satisfaction. The goal was to investigate how different indicators from the life satisfaction questionnaire relate to the latent construct of life satisfaction, and which statistical model is suitable to model the relationship between life satisfaction and the questionnaire indicators, taking into account statistical uncertainty and constraints. Furthermore, we wanted to investigate if life satisfaction changes in our cohort over the 4-year span, again taking into account statistical uncertainty.

2 Methods

1. Data set

The data comes from a survey on social integration in young adults. On two measurement points in 1987 and 1991 respecitvely, a questionnaire measuring the latent construct of life satisfaction was administered. The questionnaire constists of 3 items (indicators). The first item ("S341" in 1987 and "SS281" in 1991) states "In most respects my life is ideal", item 2 states ("S342" in 1987 and "SS282" in 1991) states "My life conditions are exellent", and item 3 ("S343" in 1987 and "SS283" in 1991) states "Overall I am satisfied with my life". Participants responded on a 7 point likert-scale ranging from "not at all applicable" to "completely" applicable. Therefore, a higher item score is applying a higher score on the latent trait life satisfaction. The items were identical over the two measurement points. The indicators are used to explain life satisfaction using confirmatory factor analysis. With the start of the data collection in 1987, 1775 participants ranging from 17 through 27 years of age participated in the study. Not all students filled out the questionnaires comprehensively. We used row-wise / list-wise deletion to remove participants when their reponses were not available for one of the item reponses across both measurement points. After the removal of missing data, n = 1245 participants remained for the final statistical analysis, including all item responses for both measurement points.

2. Statistical analysis

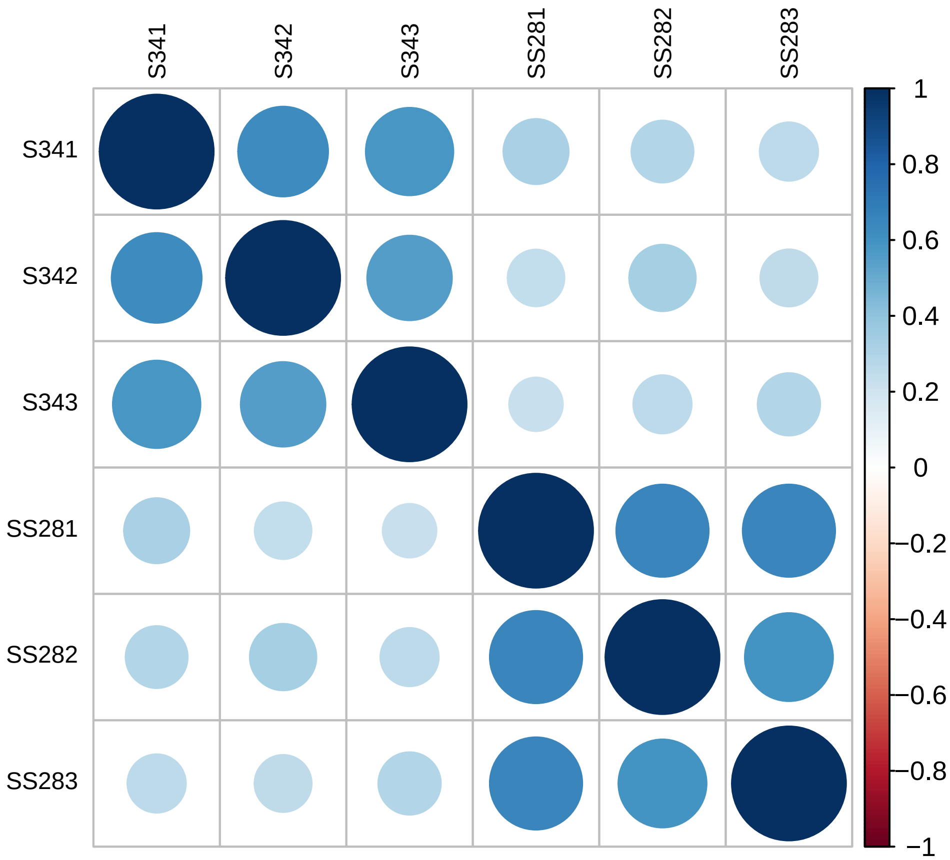
We used cofirmatory factor analysis which assumes that a latent factor life satisfaction loads on the respective items measuring it. For all the factor models we tested, we imposed a simple two-factor structure were all 3 items loaded on the respective latent life satisfaction factor for that measurement point (1987 life satisfaction loaded on 3 items, and 1991 life satisfaction loaded on the other 3 items). Since our two latent factors are not only correlated, but are though to reflect the same construct on the same participants after 4 years, we have a longitudinal design. Therefore, we tested for factorial invariance to see which assumptions and statements we can make about the underlying statistical properties of our factor model when comparing the two groups (measurement points) regarding their life satisfaction. The level of factorial invariance determines our ability to make statements about the comparisons of latent factors across both measurement points. We tested for configural invariance by comparing a model with configural invariance (same factorial structure across time points) with a model with weak factorial invariance (equal factor loadings across time points, and one freely estimated factor variance). We tested for strong factorial invariance (also called scalar invariance), by comparing a model with strong factorial invariance (equal intercepts across time points, and one freely estimated factor mean while the other is set to 0). We used χ^2-tests, RMSEA, CFI, AIC and BIC to test for model fit and to compare models.

Results

Looking at descriptive statistics, one can see that there is a tendency for higher life satisfaction scores in ascending item order. Descriptively, there does not seem to be a difference between the mean item scores over time (also see following table).

	n	mean	sd	median
S341	1245	5.072289	1.411794	5
S342	1245	5.502008	1.307671	6
S343	1245	5.874699	1.248296	6
SS281	1245	5.225703	1.256359	5
SS282	1245	5.606426	1.204610	6
SS283	1245	5.853815	1.185861	6

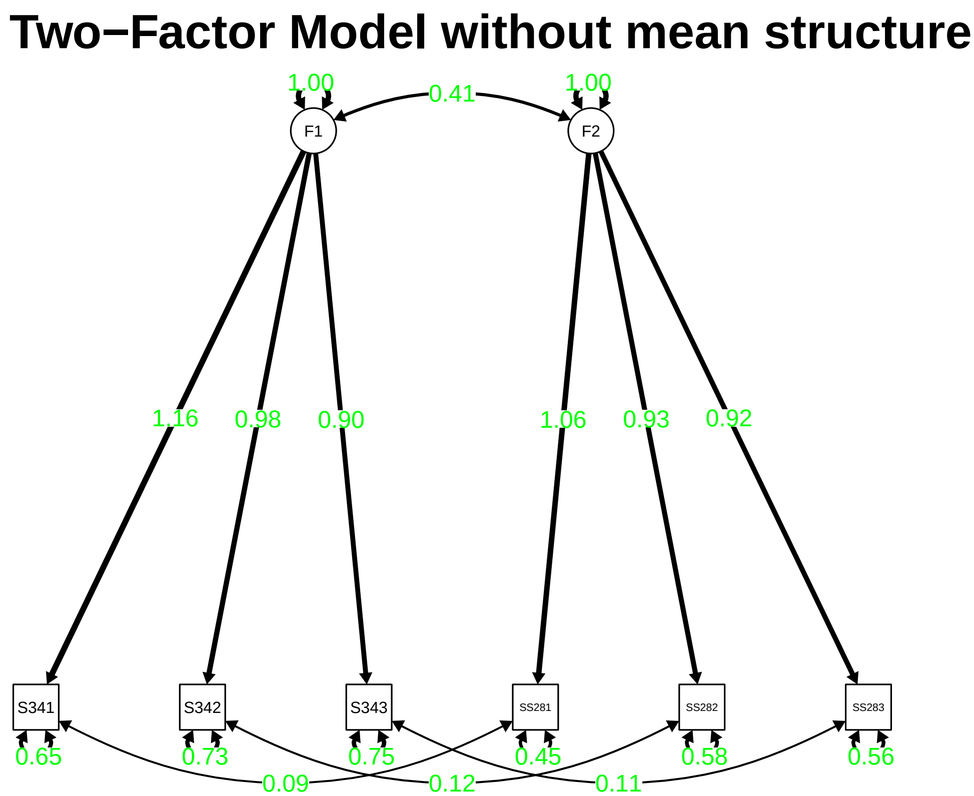
The following figure shows all pairwise item correlations.



We can see that the pairwise identical items measuring life satisfaction over both time points are correlated, which is to be expected. Interestingly, item correlations within groups are considerably higher than between groups. This means that participants at one time point responded more similar to questions of life satisfaction on different items compared to the similarity in responding to an identical item across the time points.

To get a priliminary overview of the factor model structure, we fit a two-factor model with fixed factor co(variances), and fixed covariances between the pairwise identical items. For model identification purposes, and because we do not know the scale of measurement for the latent factors, the latent factor variances are set to 1 (unit variance identification constraint). Therefore, we assume in this model that the two latent unobserved factors have the same spread in their underlying assumed distribution. Because the scales of the common factors are not defined, it is not possible to estimate both the variance and factor loading of a latent factor. All factor loadings (lambdas) can therefore be freely estimated. There is expected covariance between the two latent factors as they measure the same construct over time. Furthermore, the errors of the respective identical items over the two measurement points (factors) are correlated as well. This makes a lot of sense since the same questionnaire for life satisfaction was used both in 1987 and 1991, and therefore the same items and the errors in the two respective measurement points are expected to correlate.

The two latent factors shared significant covariance (phi21 = 0.41, SE = 0.03, z = 14.18, $p < 0.0001$). This means that the variations of the two factors are positively associated. When one factor increases, there is a tendency that the other factor increases as well, and vice versa. Both factors loaded significantly on the specified items (all $p < 0.0001$). The parameter coefficients for the factor loadings were standardized. This means that both the observed indicators and the latent factors have been transformed to the same scale. A unit increase in the indicator corresponds to a unit increase in the latent factor. F1 (life satisfaction in 1987) loaded significantly on items S341 (l11 = 1.163), S342 (l21 = 0.983), and S343 (l31 = 0.918). F2 (life satisfaction in 1991) loaded significantly on items SS281 (l42 = 1.163), SS282 (l52 = 0.983), and SS283 (l62 = 0.918). This shows that all items show a rather strong load on the factors. Since all loadings are close to one, you can on average assume that a unit increase on the indicators corresponds to a unit increase on the respective factor (when both the factors and the indicators were standardized). The following figure shows the two-factor model we esimtated with configural invariance.

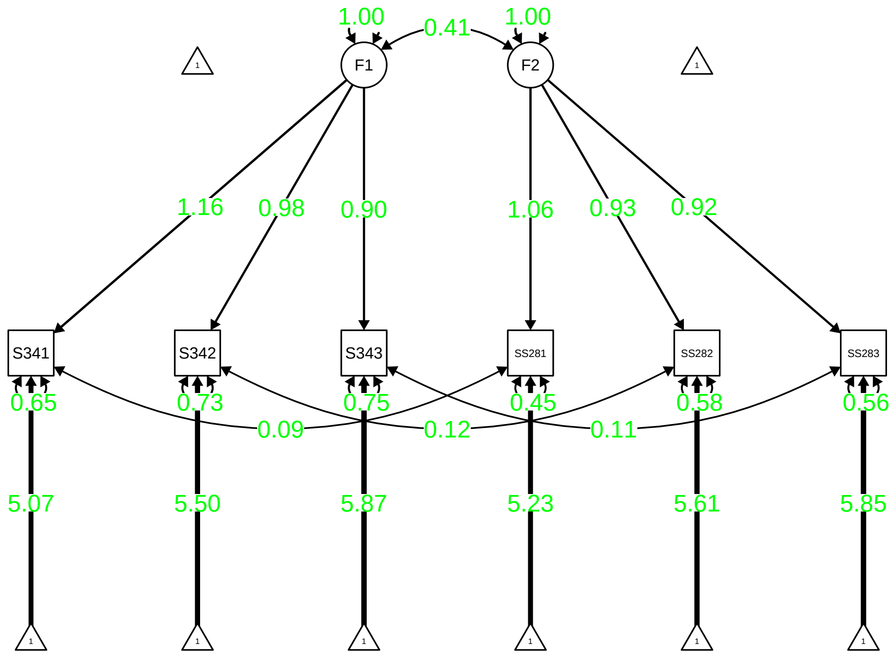


In the above specified two-factor model 16 model parameters had to be estimated with 21 observed parameters (6 variances, and 15 covariances) and 5 degrees of freedom. The model is identified, because the degrees are freedom are positive (df = 5), i.e. there are more observed parameteres than parameters that have to be estimated in the model.

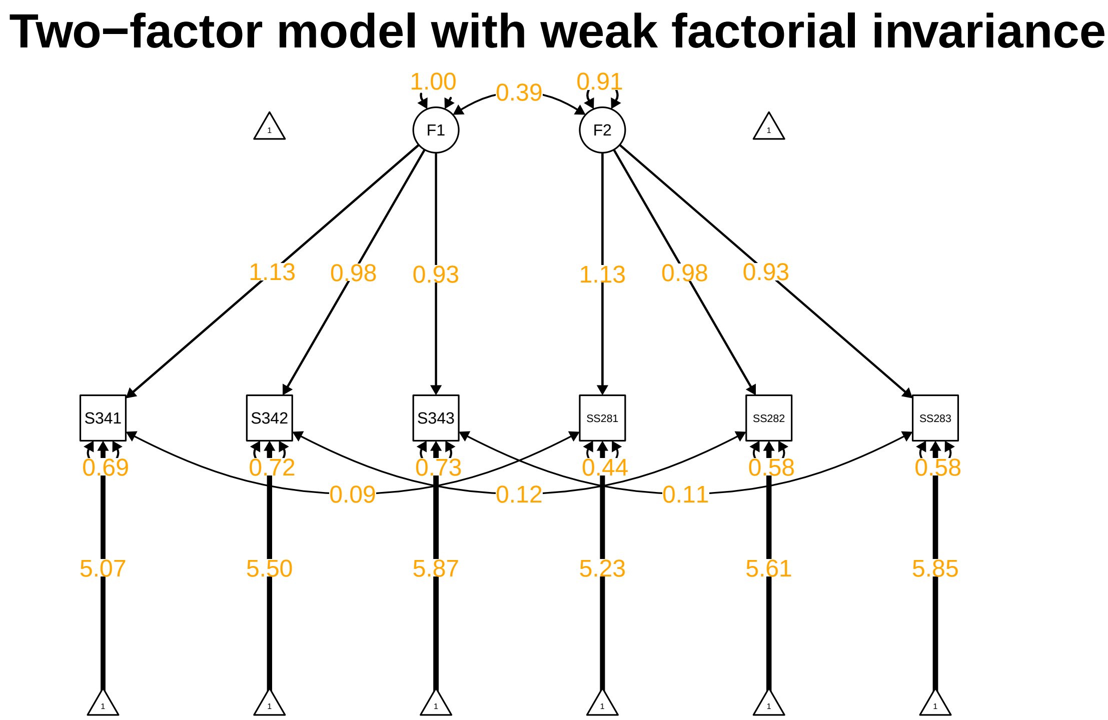
Next, we want to test for longitudinal invariance across the measurement points of 1987 and 1991. This is crucial, as the level of invariance determines which conclusions we can make from the modelling of our data. To compare the latent factor means of life satisfaction in 1987 and 1991, we need to have strong factorial invariance. A two-factor model with assumed configural invariance (and a mean structure by including intercepts on the indicators and fixed factor means of 0) is first compared with weak factorial invariance (equal factor loadings across groups and one freely estimated factor variance).

The configural invariance model fits the data well, i.e. there is no difference between the model-implied and observed variance-covariance matrix ($\chi^2(5) = 9.15$, p = 0.103).

Two−factor model with configural invariance



The weak factorial invariance model also fits the data well, i.e. there is no difference between the model-implied and observed variance-covariance matrix ($\chi^2(7) = 13.762$, p = 0.056). However, the χ^2 test for model fit shows a tendency for significance, which means that there is a tendency for an overfit of the model. The χ^2-test for model comparison suggests that there is no evidence that the weak invariance model has a worse fit than the configural invariance model ($\chi^2(2) = 4.6163$, p = 0.01), suggesting that weak invariance holds. The AIC and BIC of the configural invariance (AIC = 21755, BIC = 21868) and weak factorial invariance (AIC = 21756, BIC = 21858) suggest that the weak invariance model fits better do the data based on the BIC, although there is no substantial difference based on the AIC values. The difference in the CFI values between the configural and the weak invariance model of 0.001 suggests that the weak invariance model fits well and the weak factorial invariance holds. In terms of the RMSEA, the configural model (RMSEA = 0.026), and the weak invariance model (RMSEA = 0.028) show both a good fit (both RMSEA < 0.05). The RMSEA for the weak invariance model is slightly higher, however, the difference is negligible, suggesting that weak invariance holds. The following figure shows the final statistically valid two-factor model with weak factorial invariance.



The strong factorial invariance model does not fit the data well, i.e. there is a significant difference between the model-implied and observed variance-covariance matrix ($\chi^2(9) = 29.918$, $p < 0.001$). The χ^2-difference test for model comparison suggests that there is evidence that the strong factorial invariance model has a worse fit than the weak invariance model ($\chi^2(2) = 16.156$, $p < 0.001$), suggesting that strong factorial invariance does not hold. The AIC and BIC for the weak (AIC = 21756, BIC = 21858) and strong (AIC = 21768, BIC = 21860) factorial invariance models suggest that the weak invariance model shows a better fit to the data, especially based on the AIC value. The difference in the CFI values between the weak (CFI = 0.998) and the strong invariance model (CFI = 0.993) of 0.005 still suggests that there is a small difference in fit of the two models. In terms of the RMSEA, the weak invariance model (RMSEA = 0.028), and the strong invariance model (RMSEA = 0.043) show both a good fit (both RMSEA < 0.05). In total, the different fit indices are not suggesting the same conclusion. However, taking into account the significant χ^2-test of model fit and significant χ^2-difference test for model comparison, one can conclude that strong invariance does not hold. The best fitting final model with valid assumptions is therefore the model with weak factorial invariance. It is surprising that strong factorial invariance does not hold as the same questionnaire is given to the same participants after 4 years. Therefore, there must be further factors that cause varying intercepts in participants.

Finally, one objective of this study was to investigate whether life satisfaction improves or deteriorates from 1987 to 1991. We therefore want to compare the means of the latent factors across both time points. In the strong factorial invariance model, the first mean of the latent factor life satisfaction in 1987 was set to 0 for model identification. The second factor (life satisfaction in 1991) was freely estimated at a significantly higher value (F2 = 0.087 SE = 0.034, z = 2.537, p = 0.011), indicating that life satisfaction increased over the 4-year span. However, strong factorial invariance does not hold in our sample. This poses a challenge for directly comparing latent means because without assuming strong factorial invariance, one cannot assume that the item intercepts are the same across time points. Since the item intercepts might differ, any change in the latent means might be confounded by these differences in intercepts. Changes in item scores could be either due to changes in the latent factors, or changes in baseline scores on the items when the latent factor is 0 (intercepts). We are interested in changes in the latent factor over time, and not in the pure changes in item scores. Another approach could be testing for partial scalar invariance, which provides more statistical certainty in comparing latent factor means, because it assumes that a subset of item intercept pairs can be assumed equal across time. However, even when considering a model with partial scalar invariance, there are still item intercepts that will have to be freely estimated, because strong factorial invariance does not hold. Therefore, there will always remain a potentially confounding source, and we will never be able to certainly state that changes in the factor means over time are due to the changes in the item scores over time, or due to variations in baseline scores (intercepts) on the respective items for which the intercepts had to be freely estimated.

Discussion

This study aimed to investigate changes in life satisfaction from 1987 to 1991 using a two-factor model based on three life satisfaction items administered at both time points. Descriptive statistics indicated no differences between the mean item scores for life satisfaction in 1987 and 1991 albeit no inference statistics were run. A two-factor confirmatory factor analysis (CFA) model was tested, where life satisfaction as a latent factor at each time point loaded on the three corresponding items. The factors shared a significant positive covariance, indicating that higher life satisfaction in 1987 was associated with higher life satisfaction in 1991.

The model fit well with configural and weak factorial invariance, suggesting that the factor structure and loadings were consistent across time points. However, the strong factorial invariance model, which tests for equal item intercepts across time points, did not fit the data well, and showed a worse fit compared to the weak factorial invariance model. This indicates that strong scalar invariance does not hold in our sample, putting limitations regarding latent factor comparison over time.

Despite the lack of strong factorial invariance, the analysis indicated that life satisfaction in 1991 was slightly higher than in 1987. However, due to the absence of strong scalar invariance, this result should be interpreted cautiously, as the difference may be confounded by changes in item intercepts rather than purely reflecting changes in the latent life satisfaction construct. Further analysis, such as testing for partial scalar invariance, could provide more insight but would still leave some uncertainties regarding the true change in life satisfaction over time.

Despite the rather large sample size ($n = 1245$) we might have lost a significant amount of statistical power by removing 530 participants due to missing data. Possibly, imputation techniques could have increased power and therefore allowed for detecting strong factorial invariance. Generally however, the sample size was still considerably high, and the sample size between both groups (time points) was equal. Another possibility is that other factors lead to the fact that strong factorial invariance could not be reached, such as an unsuitable amount of items or specified latent factors. Given the simple factorial structure and the low amount of items in this longitudinal design, however, it is still surprising that strong factorial invariance could not be found. Items within groups were considerably higher correlated than between groups. A reason for this could be that there were external factors that influenced particpants baseline response behaviors independently of their values of the latent factor of life satisfaction. The participants' home country and the background of the data collection process is not stated specifically, and there were a lot of political upheavals from the late 1980s and early 1990s which could have influenced participants' response behavior.

All taken together, the research questions if life satisfaction in young adults changed from 1987 to 1991 could not be answered, because the statistical requirements for making such a statements were not met. However, there are hints that life satisfaction could have improved slightly over time. Furthermore, the statistical and conceptual relationship between the respective items of the life satisfaction questionnaire and their underlying latent factor of life satisfaction could be further pinpointed. Life satisfaction loaded equally on all identical questionnaire items over both measurement points, indicating that the relationship between life satisfaction and the constructs measured by the respective items remained the same over a 4-year span.