The construction and Estimation of Multidimensional Latent Factor Models without Parametric Assumptions

Landon Hurley[1]

July 3, 2023

[1] ljrhurley@gmail.com

- Today's talk: deals with estimation upon non-parametric latent manifolds introduced yesterday.

- Linear Factor Analysis (FA) was formalised by Eckart and Young (1936): a linear Whitney embedding was proven for the Frobenius norm ℓ_2 abelian sub-space space wrt columns, assuming a Gramian similarity matrix. – Technically treated as a contraction when $\Sigma^{p \times p} \mapsto \theta_{n \times q}^{\intercal} \times \theta^{n \times q}$.

- Thus, the covariance matrix must be abelian and invertible, possessing a linear spanning basis unique upon the sample set.

- There are alternative covariance matrices which may be employed. ℓ_2 presumes Gaussian data to ensure MLE equiv GM, but it is not necessary.

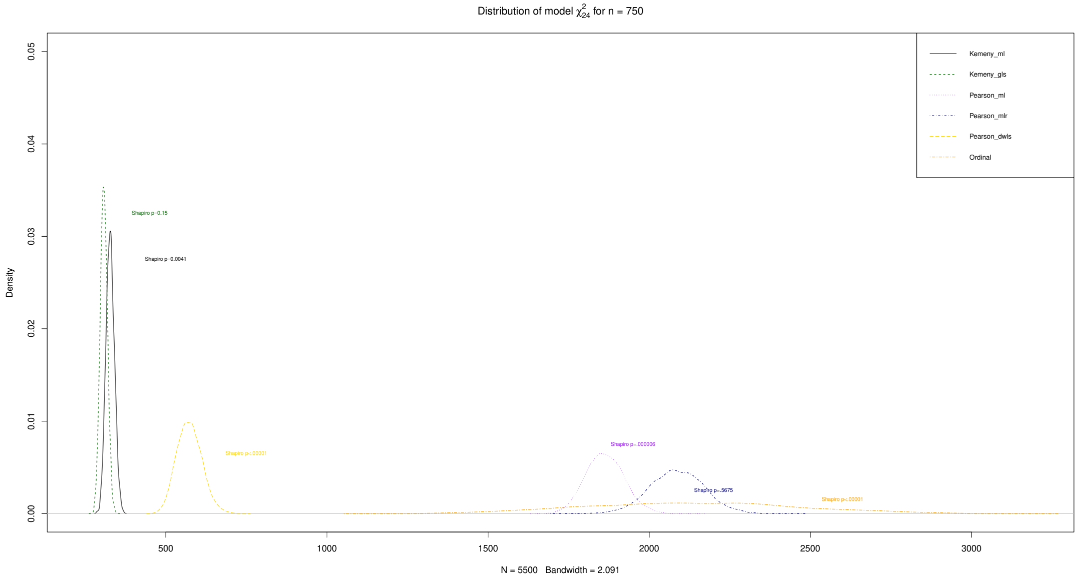


Figure: Visualisation of the distributions of the χ^2_{24} test statistics from Table 3.

Distribution of model χ^2_{24} for n = 7500

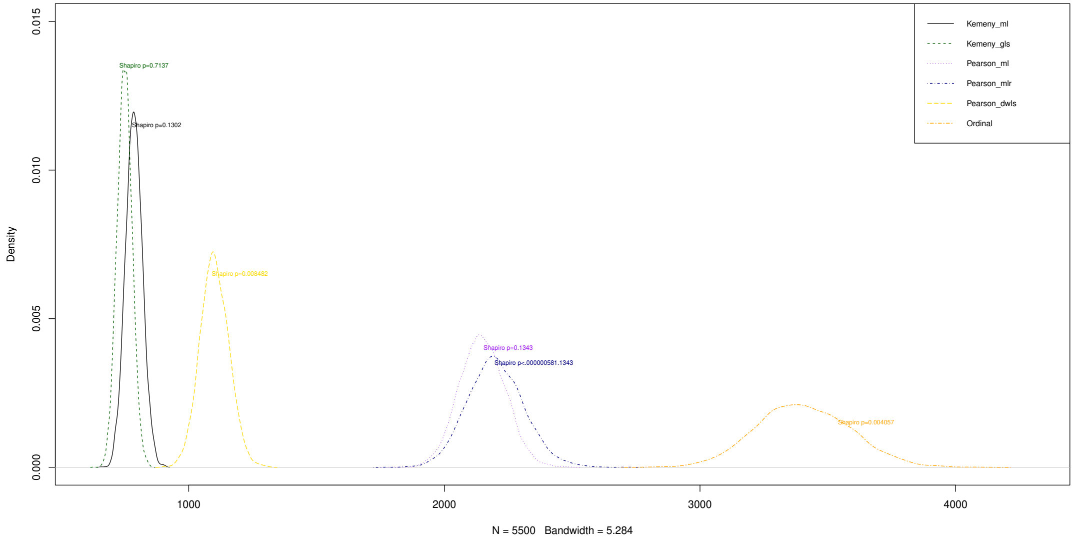


Table: Presentation of the observation of either a linear or non-linear manifold wrt the ranks and scores of upon a collection of random variables and their errors.

	Ranks	Ranks
Scores	Linear	Non-linear
Linear	(1)	(2)
Non-linear	(3)	(4)

- If the Neyman-Pearson lemma is true (or else reject all of Frequentist states...) then we expect an abelian function space to be expressible upon both the observed and latent domains and sufficient. This addresses points 1 & 2.

- Under the Eckart and Young (1936) theorem, nothing explicitly requires ℓ_2 to define the observed data linearly, only that the latent sub-space be a uniquely defined linear function of such data.

- Thus, a short-mapping is sufficient: both the latent and observed spaces being Gaussian is a stronger than necessary condition.

- For the sake of argument, consider what would be observed if we attempted to embed linear or non-linear observed data onto a non-linear latent variable.

- Unsurprisingly, attempting to approximate a non-linear surface with a linear greedy approximation would neither result in unbiased nor efficient estimates. Moreover, as the Gauss-Markov conditions are not met, the solution is not generalisable beyond the sample.

- Under such conditions, we would expect a non-linear latent manifold to produce inconsistent fit with the data: this would also explain the significant χ^2 test results, which replicate across multiple samples.

- Moreover, we would expect such tests to be excessively inconsistent with the proposed linear latent manifold: exceptionally strong rejection of the null hypothesis would be a necessary consequence.

- Such findings are, again unsurprisingly, entirely consistent with 80 years of Social Science research. They also can be explained as wishful thinking: after all, we have yet to actually propose an quantitative alternative.

Non-parametric linear short-mappings I

- We observe upon the real line $\mathbb{R}^{n \times p}$ the Borel-Cantelli lemma, reflecting the almost sure convergence of a solution. When said real data is Gaussian, we can leverage the ℓ_2 operator norm to define functions which are unbiased and minimum-variance.

- This allows us to relax both $n < \infty^+$, $p < \infty^+$, and still obtain Gauss-Markov estimators: this is necessary because it takes a very long time to acquire a study which contains the exhaustive population of people and the exhaustive population of features.

- Upon this population, by the CLT, we obtain factor scores: weighted linear combinations which asymptotically sum to a normal distribution.

- This is why we declared all latent variables to be Gaussian distributed.

- A natural, yet oddly unasked, question though is: Is the asymptotic latent Gaussian manifold stable?

Non-parametric linear short-mappings II

- Stability reflects the relaxation of the asymptotics to ensure the same distribution is obtained for finite samples.

- The Glivenko-Cantelli theorem ensures $n < \infty^{+}$ remains unbiased, as long as it is uniformly sampled.

- However, we rarely have exceptionally long test questions, such that $p \ll \infty^{+}$: are short tests guaranteed to reflect Gaussian latent variables?

- The answer, unsurprisingly, is no. The evidence is favour of this conclusion is exceptionally long-standing: it exists by the fields rejection of the Neyman-Pearson GOF test.

$$\mathbf{V}^{p\times p} = \Psi^{p\times p} + \Lambda^{p\times q}\Phi^{\intercal}_{q\times q}\Lambda^{\intercal}_{p\times q}. \tag{1}$$

- Generally, we tend to prefer such estimators. However, we rejected it because we almost always found Structural Equation Models to exhibit significant misfit.

Non-parametric linear short-mappings III

- The attempted justification was that the estimator is overly sensitive: however consider the assumptions:
    1. The observed manifold is uniquely defined in an unbiased manner, and can be summarised upon the ℓ_2-operator norm.
    2. The latent contraction ($\ell_2 \times \ell_2 \mapsto \ell_2$) is correctly defined and exhibits properties of a Gauss-Markov estimator of approximation.
    3. The latent manifold is Gaussian.

Gaussian latent variables I

- The Gaussian nature of errors in function approximation is a consequence of Eckart and Young (1936) being constructed as a Hadamard estimation problem.

- We construct p estimating equations for each of q latent dimensions, averaging over n sample individuals.

- By the central limit theorem, even if p variables are not normally distributed, their asymptotic weighted linear combination is still Gaussian.

- If they are linear functions, then we expect the data structure to be unbiased and possess minimum variance. Thus, all latent variables are asymptotically Gaussian.

- These conditions are, upon the test item population, identical to that of the Rasch model: we expect Test scores to uniquely determine the ranking of the sample proportionately to their true score $\theta_i^{n \times q}, i = 1, 2, \ldots, n$.

Gaussian latent variables II

- If the observed items do not share common discrimination though, there is no constant ΔT for test score. Once the 1PL is generalised, there is no longer bilinearity.
- As an estimation sub-space embedding of the Whitney theorem, the contraction upon $\mathbf{X}^{\mathsf{T}}_{n \times p} \times \mathbf{X}^{n \times p} \to \theta^{n \times q}$, we lose the ability to unique define a unique score.
- Wilson (1928) first revealed this, by identifying the loss of uniqueness upon quadratic polynomials and their lack of singular roots $\pm\lambda_j$, and Bochner expressed this as

$$\lim_{p \to \infty^+} \frac{p^2 - p}{2} + p \leq p^2,$$

which only achieves equality in the limit. Thus, factor score indeterminacy.

Stability of Gaussian latent variables I

- Next, consider relaxing to $p < \infty^+$. Under this relaxation, the linear embedding of the Gaussian errors upon $\mathbf{X}^{n \times p} \mapsto \theta^{n \times q} + \epsilon, \epsilon \sim \mathcal{N}(\mu^{q \times 1}, \Sigma^{q \times q}$, the observed values follows through upon the Gaussian latent variables.

- Thus, if the latent variables are normally distributed, we expect to obtain unbiased tight results upon finite samples wrt both $n \& p$. This supports the claim that the GOF χ^2-test should hold, uniformly via Glivenko-Cantelli being the most powerful misspecification detector.

Non-Gaussian latent manifolds I

- However, the distribution of the latent manifold is an empirical assumption, not a guarantee, unless performed upon the population of test items. This limit is therefore not guaranteed stable: asymptotically normal latent variables are not expected to be noramlly distributed upon finite samples when defined by only a few indicators.

- We offer a resolution to this problem by identifying a linear manifold and corresponding probability structure which is orthonormal to the ℓ_2 definition of the problem.

- Explicitly, we address the problem of Hadamard estimation problem as the ML or linear Gauss-Markov (GM) estimate of the latent scores which are ranked most accurately.

- This process is achieved by separating and defining a linear topology upon the ranks, and noting that, under certain general conditions, even non-linear scores are almost surely linearly ranked.

Non-Gaussian latent manifolds II

- The first problem consists of defining a sufficient domain for the problem: $S_n!$ is typical, but the lack of continuous random variables and their linear combination results in strong probability of ties occuring.

- We require a means of defining a metric space over permutations with ties: $\mathcal{M}_n = \{n^n - n\}_m$.

- Note that we explicitly exclude the now possible n permutation events which are degenerate from the population.

- Kemeny (1959) first identified a metric space for this, however his expression was too similar to Kendall (1938), and failed to recognise that the Kemeny space was a Hilbert space.

- Upon a Hilbert space, we achieve almost sure convergence via the Borel-Cantelli lemma, with a complete and unique probabilistic mapping for all n.

- As noted, a positive definite covariance matrix which is expressible upon the ℓ_2-norm function space is necessary to explore this idea.

- To do this, we introduce the Kemeny (1959) metric: it serves as a permutation parallel (formally, a projective geometric dual) to the ℓ_2 function space. It is orthonormal to those estimates, and can be estimated upon the same data.

- Note that any $\mathbf{X}^{n\times p} \mapsto \mathbf{A}^{p\times p} \times \mathbf{B}^{n\times n}$ is valid: any rectangular matrix can be uniquely and exactly (in the Gram-Schmidt sense) represented as the linear combination of two square matrices of different order.

$$\tau_\kappa(x,y) = -\tfrac{2}{n^2-n}\sum_{k=1}^{n}\sum_{l=1}^{n}\kappa(x)_{kl}\odot\kappa^\intercal(y)_{kl} = -\tfrac{2}{n^2-n}\left(\rho_\kappa(x,y)-\tfrac{n^2-n}{2}\right), \{x,y\}\in\overline{\mathbb{R}}^{n\times 1}. \tag{2a}$$

$$\rho_\kappa(x,y) = \frac{n^2-n}{2} + \sum_{k,l=1}^{n}\kappa_{kl}(x)\odot\kappa^\intercal_{kl}(y),\ \ k,l = 1,\ldots,n. \tag{2b}$$

$$\kappa_{kl}(x) = \begin{cases}\sqrt{.5} & \text{if } x_k > x_l\\ 0 & \text{if } x_k = x_l,\\ -\sqrt{.5} & \text{if } x_k < x_l\end{cases} \tag{2c}$$

$$\kappa_{kl}(y) = \begin{cases}\sqrt{.5} & \text{if } y_k > y_l\\ 0 & \text{if } y_k = y_l,\\ -\sqrt{.5} & \text{if } y_k < y_l\end{cases} \tag{2d}$$

- The Kemeny metric is a Hilbert space, and also a Gauss-Markov estimator. It is affine-linear over monotonically non-decreasing functions as well.

- It operates upon $\mathcal{M}_n$, the set of all permutations upon the extended real line, and converges for all n almost surely.

- $S_n \subset \mathcal{M}_n$ as well. It addresses the problem of the probability of observing ties uniquely.

- This allows us to express a distance matrix between all vectors $X^{n \times p}$ upon the sample, and an affine-linear transformation of the distances provides a correlation measure τ_κ as well.

- $X^{n=5\times 1} = [1,2,3,4,5]^{\mathsf{T}}$:

$$\kappa_{k,l}(X) = \left[\begin{array}{c|ccccc} & l=1 & l=2 & l=3 & l=4 & l=5 \\ \hline k=1 & 0 & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} \\ k=2 & -a_{2,1} & 0 & a_{2,3} & a_{2,4} & a_{2,5} \\ k=3 & -a_{3,1} & -a_{3,2} & 0 & a_{3,4} & a_{3,5} \\ k=4 & -a_{4,1} & -a_{4,2} & -a_{4,3} & 0 & a_{4,5} \\ k=5 & -a_{5,1} & -a_{5,2} & -a_{5,3} & -a_{5,4} & 0 \end{array}\right] \tag{3}$$

$$X_* = [-4a, -2a, 0a, 2a, 4a]^{\mathsf{T}} \tag{4}$$

$$\hat{s}_*^2(X) = \frac{1}{5-1}\sum_{i=1} X_*(i)^2 = \frac{40\cdot(\sqrt{\frac{1}{2}})^2}{4} = 5. \tag{5}$$

- Here, we express the skew-symmetric matrix of order $n \times n$, a natural representation of the permutation structure upon the population $\mathcal{M}_5$.

- These elements, $a = \sqrt{\frac{1}{2}}$ provide a linear spanning basis (confirmed via Gram-Schmidt), and allow for constructing an $n \times 1$ representation by the transpose of the marginalisation over all rows, acting as an affine-linear transformation of the pre-existing norm's basis.

- The linear combination of the former basis provides a Gauss-Markov estimator, defined as a generalisation of Kendall (1938) τ.

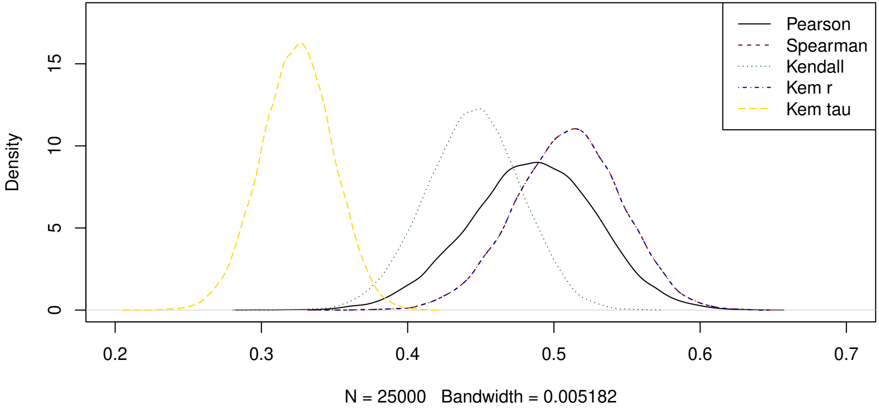
- The linear combination of the latter basis provides a Gauss-Markov estimator as a generalisation of Spearman (1904) ρ.

- As both operate upon the domain of the extended real line, $X^{n\times 1} \in \overline{\mathbb{R}}^{n\times 1}$, we observe that the Borel-Cantelli convergence guarantees exist upon both estimators, even when considering the Cauchy distribution. This is because the median is defined even when the mean is not.

Correlation performance

	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
r	0.485	0.044	0.486	0.485	0.044	0.297	0.642	0.344	-0.110	-0.006
ρ	0.509	0.036	0.510	0.510	0.036	0.345	0.635	0.290	-0.116	0.017
τ_b	0.445	0.033	0.445	0.445	0.033	0.292	0.563	0.271	-0.084	0.016
Kem ρ_κ	0.509	0.036	0.510	0.510	0.036	0.345	0.635	0.290	-0.116	0.017
Kem τ_κ	0.324	0.025	0.325	0.324	0.025	0.214	0.412	0.199	-0.073	0.018

Correlation distribution of A2 & A3



Another comparison under population independence

Table: Comparison of the average error in correlation for a bivariate pair, using various methods estimated over a number of sample sizes, each time for 15,000 iterations. Note that the biased variance of the Kendall estimator approaches our estimator as *n* increases.

	vars	Average Error	sd of estimator error	median	min	max
n=30	Kemeny ρ	0.000	0.187	0	-0.63	0.6
	Kemeny τ_κ	0.001	0.129	0	-0.45	0.46
	Pearson r	-0.001	0.187	0	-0.64	0.62
	Kendall τ_b	0.000	0.134	0	-0.47	0.48
n=1500	Kemeny ρ	0.001	0.082	0	-0.31	0.31
	Kemeny τ_κ	0.001	0.045	0	-0.21	0.10
	Pearson r	0.001	0.082	0	-0.31	0.31
	Kendall τ_b	0.000	0.055	0	-0.21	0.11
n=5000	Kemeny ρ	0.000	0.045	0	-0.17	0.14
	Kemeny τ_κ	0.000	0.030	0	-0.12	0.13
	Pearson r	0.000	0.045	0	-0.17	0.14
	Kendall τ_b	0.000	0.030	0	-0.12	0.13

- Mokken (1971) first proposed a non-parametric latent variable. He identified it via the same bilinearity condition that identifies the Rasch/1PL model: the linear ordering of the total scores is proportionate the the rankings upon the latent variable.

- However, the Mokken model is unfortunately more descriptive than quantitative: all we can really do with it is sort the scores.

- Moreover, the presence of ties precludes the identification of non-unidimensional latent variables: we cannot uniquely measure linear combinations of rankings with error, as sums of individual items produce surjective mappings. This occurs almost surely under the birthday paradox for small item response sets (e.g., 1-6).

- Note though that the latent variable scores are the relative rankings upon the test, questionnaire, or scale, denoting relative extremenes of responses.

- This is suspiciously similar to the Kemeny ρ_κ domain (in fact, it is the same): just a linear function of the relative rankings upon the linear or non-linear scores.

$$\mathbf{X}^{n\times p} = \Theta^{n\times q}\Phi^{q\times q}\Lambda^{\mathsf{T}}_{p\times q} + \Psi^{n\times p} + \mathbf{E}^{n\times p}. \tag{6}$$

$$\begin{aligned}
(n-1)\mathbf{V}^{p\times p} &= \mathbf{X}^{\mathsf{T}}_{n\times p}\mathbf{X}_{n\times p} \\
&= (\epsilon^{n\times p} + \Theta_{n\times q}\mathbf{w}^{\mathsf{T}}_{p\times q})^{\mathsf{T}}(\epsilon^{n\times p} + \Theta^{n\times q}\mathbf{w}^{\mathsf{T}}_{p\times q}) \\
&= \epsilon^{\mathsf{T}}_{n\times p}\epsilon_{n\times p} + \epsilon^{\mathsf{T}}_{n\times p}\Theta_{n\times q}\mathbf{w}_{p\times q} + \mathbf{w}^{\mathsf{T}}_{p\times q}\Theta^{\mathsf{T}}_{n\times q}\epsilon_{n\times p} + \mathbf{w}_{p\times q}\Theta^{\mathsf{T}}_{n\times q}\Theta_{n\times q}\mathbf{w}^{\mathsf{T}}_{q\times p} \\
&= (n-1)\Psi_{p\times p} + 0 + 0 + (n-1)\mathbf{w}\mathbf{I}\mathbf{w}^{\mathsf{T}} \\
&= (n-1)\Psi + (n-1)\mathbf{w}\mathbf{w}^{\mathsf{T}} \\
\mathbf{V}^{p\times p} &= \Psi^{p\times p} + \mathbf{w}_{p\times q}\mathbf{w}^{\mathsf{T}}_{p\times q} \qquad \text{(7a)}
\end{aligned}$$

$$\mathbf{V}^{p\times p} = \Psi^{p\times p} + \Lambda^{p\times q}\Phi^{\mathsf{T}}_{q\times q}\Lambda^{\mathsf{T}}_{p\times q}. \tag{7b}$$

Empirical Demonstration: Iris dataset I

- We performed CFA variable model for the Iris dataset, ultimately to conduct measurement invariance

- Note that: the standard errors upon continuous random variables are biased (too little variance on κ)

- We can correct this by multiplying s.e.'s by $(\frac{7}{11})^2$

- Not a problem for discrete (i.e., ordinal) data items. This numerical analytic correction confirmed by bootstrapping.

- Once corrected, everything can be interpreted and reported as expected.

- Note that there is now a direct inverse correlation between s.e. and χ^2, as expected/required by the Neyman-Pearson lemma.

Empirical Demonstration: Iris dataset

	$\chi^2_2 = 1.423, p \leq .491$		$\chi^2_2 = 1.397, p \leq .497$		$\chi^2_2 = 59.593, p \leq .000$		$\chi^2_2 = 33.890, p \leq .000$	
	Est	s.e.	Est	s.e.	Est	s.e.	Est	s.e.
Sepal.Length	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
Sepal.Width	-0.229	0.045	-0.229	0.046	-0.350	0.044	-0.342	0.058
Petal.Length	1.242	0.048	1.242	0.048	2.776	0.156	2.734	0.142
Petal.Width	1.079	0.043	1.076	0.043	1.027	0.069	1.067	0.062
Sepal.Length	0.423	0.024	0.417	0.024	0.235	0.026	0.121	0.016
Sepal.Width	0.894	0.042	0.883	0.042	0.134	0.014	0.082	0.012
Petal.Length	0.132	0.021	0.131	0.021	-0.339	0.056	-0.173	0.033
Petal.Width	0.296	0.021	0.299	0.021	0.107	0.013	0.071	0.009
F	0.537	0.042	0.542	0.043	0.446	0.073	0.431	0.073

Empirical Demonstration: Iris dataset, Measurement Invariance

	Df	χ^2	$\Delta\chi^2$	$\Delta(df)$	$\Pr(\Delta\chi^2_6))$
fit.config	6	8.4402	–	6	–
fit.loadings	12	17.3421	8.9018	6	0.1791737
fit.strong	18	44.7294	27.3873	6	0.0001225
fit.strict	26	53.1954	8.4660	8	0.3893233
fit.config2	6	26.98			
fit.loadings2	12	62.95	35.97	6	0.0000
fit.strong2	18	118.24	55.29	6	0.0000
fit.strict2	26	131.90	13.66	8	0.0912
fit.config2	6	30.91			
fit.loadings2	12	74.44	47.52	6	0.0000
fit.strong2	18	228.09	171.90	6	0.0000
fit.strict2	26	312.34	66.91	8	0.0000

Empirical Demonstration: Holzinger dataset, Measurement Invariance

Procedure		Df	χ^2	$\Delta\chi^2$	RMSEA	$\Delta(df)$	$\Pr(\Delta\chi^2_6)$
Kem_ML	fit.config	96	52.17				
	fit.loadings	114	60.84	8.67	0.00	18	0.9670
	fit.strong	132	119.97	59.13	0.17	18	0.0000
	fit.strict	159	124.67	4.70	0.00	27	1.0000
		Df	χ^2	$\Delta\chi^2$	RMSEA	$\Delta(df)$	$\Pr(\Delta\chi^2_6)$
Kem_GLS	fit.config2	96	47.03				
	fit.loadings2	114	52.39	5.36	0.00	18	0.9982
	fit.strong2	132	106.92	54.53	0.16	18	0.0000
	fit.strict2	159	114.30	7.38	0.00	27	0.9999
		Df	χ^2	$\Delta\chi^2$	RMSEA	$\Delta(df)$	$\Pr(\Delta\chi^2_6)$
Pear_ML	fit.config3	96	158.94				
	fit.loadings3	114	195.82	36.87	0.12	18	0.0054
	fit.strong3	132	259.21	63.39	0.18	18	0.0000
	fit.strict3	159	310.67	51.46	0.11	27	0.0031

Comparison of fitness statistics for Holzinger-Swineford

Table: Distributions of the likelihood-ratio tests for the Holzinger data set under different estimators, Ordinal defining the combination of polychoric and DWLS estimators, for various *n*.

n		mean	sd	median	mad	min	max	range	skew	kurtosis
n = 301	Kem_ml	135.78	8.54	135.41	8.50	107.94	172.14	64.20	0.19	0.06
	Kem_gls	125.25	7.10	125.11	7.07	100.22	156.27	56.05	0.02	0.05
	Pear_ml	762.03	39.11	759.76	38.54	642.64	952.66	310.03	0.36	0.28
	Pear_gls	838.45	52.84	837.61	53.23	654.54	1037.89	383.35	0.09	0.05
	Ken_ml	238.11	25.11	236.03	24.41	165.14	355.81	190.67	0.51	0.52
	Ken_gls	836.07	165.33	837.49	179.24	388.66	1371.87	983.21	-0.02	-0.56
n = 750	Kem_ml	327.49	13.01	327.15	13.05	284.63	371.73	87.10	0.11	-0.09
	Kem_gls	307.78	11.15	307.55	11.11	267.35	354.09	86.75	0.09	-0.01
	Pear_ml	1862.46	59.63	1861.04	60.92	1660.84	2146.83	485.99	0.16	0.08
	Pear_gls	2086.23	83.36	2086.02	83.58	1734.29	2443.53	709.23	-0.01	0.11
	Ken_ml	573.25	38.94	572.07	39.50	458.98	743.20	284.21	0.26	0.06
	Ken_gls	2128.23	326.40	2126.12	337.12	1209.75	3113.24	1903.49	0.06	-0.40
n = 7500	Kem_ml	3209.740	41.008	3209.908	40.780	3055.841	3345.246	289.405	-0.007	-0.056
	Kem_gls	3052.594	36.288	3052.480	36.659	2923.028	3175.430	252.402	0.013	-0.090
	Pear_ml	18402.979	186.499	18403.916	189.628	17784.339	19041.252	1256.913	0.064	-0.125
	Pear_mlr	20865.061	266.452	20865.209	274.161	19974.056	21818.144	1844.088	-0.021	-0.118
	Pear_dwls	5610.712	120.587	5607.139	122.131	5218.355	6090.508	872.153	0.089	0.006
	Ordinal	21664.151	1192.641	21666.358	1181.759	16652.545	27453.147	10800.602	0.012	0.483

Demonstration of differences in fitness for alternative latent spaces upon Big-5 Personality Index

Table: Model fitness for Gaussian and non-Gaussian latent variables upon the Big-5 model, and also when stratified by gender. For comparison, analysis upon MVN Pearson Covariance matrix misfits $\chi^2_{265} = 3843.296$. Under Rhemtuella correction, null homogeneous scaled $\chi^2_{265} = 5584.411$

Model	χ^2	DF	$\Pr(\chi^2_{df} = 0)$
Gaussian latent - Homogeneous	1439.265	265	.000
Non-Gaussian latent - Homogeneous	1045.181	265	.000
Gaussian latent - Gender - 1	572.530	265	.000
2	963.041	265	.000
Non-Gaussian latent - Gender - 1	399.291	265	.000
2	710.330	265	.000
Rhemtuella Gender - 1	3507.844	265	.000
2	2107.182	265	.000

Benfits of this approach I

- Better fitting, non-significant, models.
- Uniformly most powerful tests can be used to evaluate the question of whether latent variables truly are Gaussian.
- Tighter and more efficient (generalisable) empirical findings.
- Better performance upon small samples, since Ξ is pretty much always positive definite (most extreme case was 12×27.)
- Any bivariate distribution which share a common CDF is a linear function of the Kemeny metric.
- Allows us to solve ill-posed estimation problems uniquely (saddlepoint theorem), useful for missing data EM approaches wherein Gaussian errors are not possible

Benfits of this approach II

- Important implications for s.e. of missing data: if the expected Fisher Information Matrix is employed, it is biased downwards, inflating the Type 1 error rate dramatically.

- Changes very little in the overall analytical procedures: replace the covariance matrix, choose an estimator, adjust standard errors after fitting.

- Φ matrix remains intuitively interpretable as well: $\rho_\kappa \approx r$.

- Keeps linear models over otherwise non-linear scores (e.g., generalised partial credit model, ordinal regression, etc.)

- Note also that the polychoric correlation matrices presume latent normal distributions: if we show the latent variables to not be Gaussian, then the polychoric correlations are often biased.

Benfits of this approach III

- Resolves the biggest problem of Social Science Measurement: why do my causal models never fit the data well (without rejecting the Neyman-Pearson lemma).

Bibliography I

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*(3), 211–218. https://doi.org/10.1007/bf02288367

Kemeny, J. G. (1959). Generalized random variables. *Pacific Journal of Mathematics*, *9*(4), 1179–1189. https://doi.org/10.2140/pjm.1959.9.1179

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 81–93. https://doi.org/10.2307/2332226

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter Mouton. https://doi.org/10.1515/9783110813203

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72. https://doi.org/10.2307/1412159

Wilson, E. B. (1928). Review: The abilities of man, their nature and measurement. *Science*, *67*(1731), 244–248. https://doi.org/10.1126/science.67.1731.244