Latent variable modeling: A practical guide

Estimation and Model Fit

Waylon Howard

Webinar, March 05, 2025

ML estimation (typically the default)

- ML identifies the population parameters that are most likely given the observed data.
- A likelihood (or log-likelihood) function is used to quantify how well the proposed parameters explain the observed data.
- ML requires a population distribution (normal).

ML estimation 😏

A density function gives the shape of the normal curve

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5\frac{(Y_i-\mu)^2}{\sigma^2}}$$

L_i (the likelihood) gives the relative probability that Y_i came from a normal distribution with a particular mean and variance.

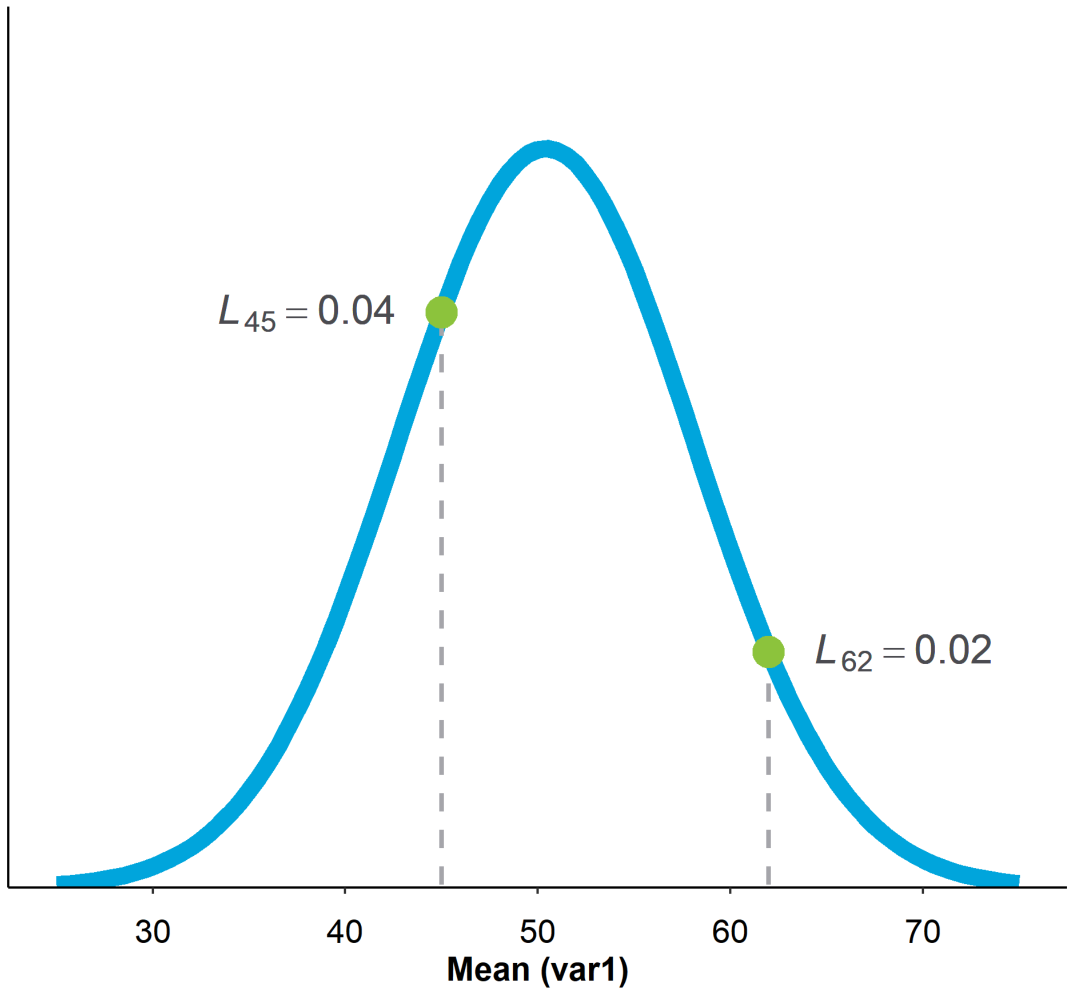
ML estimation

Applying the density function gives the relative probability (L_i) of each score from this normal distribution.

Population Values

var1 (μ = 50.42, σ = 7.65)

ID	var1	L_i
1	36.6	0.010201
2	41.8	0.027624
3	42.6	0.030908
4	43.1	0.032971
5	43.4	0.034205
6	44.2	0.037444
7	44.9	0.040166
8	46.3	0.045074
9	48.6	0.050658
10	49.0	0.051223
11	50.0	**0.052038**
12	51.6	0.051508
13	54.6	0.044915
14	54.8	0.044264
15	55.7	0.041102
16	57.2	0.035227
17	57.6	0.033589
18	60.3	0.022677
19	60.9	0.020433
20	65.3	0.007888



$L_{45} = 0.04$
$L_{62} = 0.02$
30
40
50
60
70
Mean (var1)

Maximum Likelihood

Multiple each (L_i) to get sample likelihood.

0.0000000000000000000000000000163666415977258

Fit of this data to μ = 50.42, σ = 7.65

To avoid small numbers, we take the log of the likelihood.

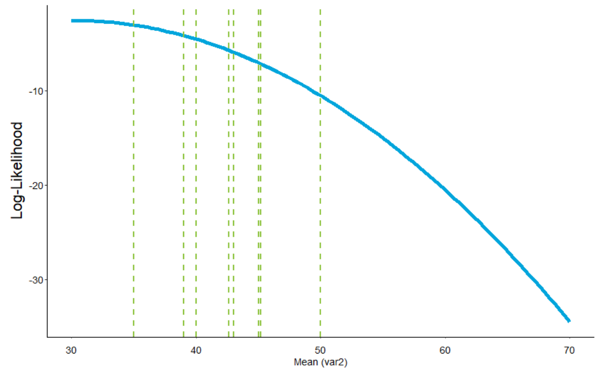
Add each $logL_i$ to get sample loglikelihood.

-68.58

ID	var1	L_i	$logL_i$
1	36.6	0.010201	-4.585258
2	41.8	0.027624	-3.589055
3	42.6	0.030908	-3.476754
4	43.1	0.032971	-3.412113
5	43.4	0.034205	-3.375376
6	44.2	0.037444	-3.284921
7	44.9	0.040166	-3.214733
8	46.3	0.045074	-3.099445
9	48.6	0.050658	-2.982664
10	49.0	0.051223	-2.97157
11	50.0	0.052038	**-2.955783**
12	51.6	0.051508	-2.966023
13	54.6	0.044915	-3.102986
14	54.8	0.044264	-3.117579
15	55.7	0.041102	-3.191692
16	57.2	0.035227	-3.345936
17	57.6	0.033589	-3.393553
18	60.3	0.022677	-3.786393
19	60.9	0.020433	-3.890587
20	65.3	0.007888	-4.842415

ID	var1	μ = 30	μ = 40	μ = 50	μ = 60	μ = 70
1	36.6	-3.326	-3.053	-4.487	-7.627	-12.474
2	41.8	-4.142	-2.982	-3.528	-5.781	-9.74
3	42.6	-4.309	-3.012	-3.422	-5.538	-9.361
4	43.1	-4.419	-3.036	-3.361	-5.391	-9.129
5	43.4	-4.487	-3.053	-3.326	-5.306	-8.992
6	44.2	-4.675	-3.105	-3.241	-5.085	-8.634
7	44.9	-4.849	-3.159	-3.176	-4.9	-8.33
8	46.3	-5.222	-3.293	-3.071	-4.556	-7.747
9	48.6	-5.906	-3.585	-2.971	-4.063	-6.862
10	49	-6.035	-3.645	-2.963	-3.987	-6.718
11	50	-6.368	-3.808	-2.954	-3.808	-6.368
12	51.6	-6.936	-4.103	-2.976	-3.556	-5.843
13	54.6	-8.118	-4.773	-3.135	-3.203	-4.978
14	54.8	-8.203	-4.823	-3.151	-3.185	-4.926
15	55.7	-8.591	-5.058	-3.231	-3.112	-4.699
16	57.2	-9.268	-5.479	-3.397	-3.021	-4.352
17	57.6	-9.455	-5.598	-3.447	-3.003	-4.266
18	60.3	-10.789	-6.471	-3.86	-2.955	-3.757
19	60.9	-11.102	-6.682	-3.968	-2.961	-3.661
20	65.3	-13.588	-8.416	-4.952	-3.194	-3.143
		-139.79	**-87.13**	**-68.62**	**-84.23**	**-133.98**

Possible population means for var2



Audition different parameters to quantify how well the proposed values explain the observed data.

Green dotted lines represent observed values for var2.

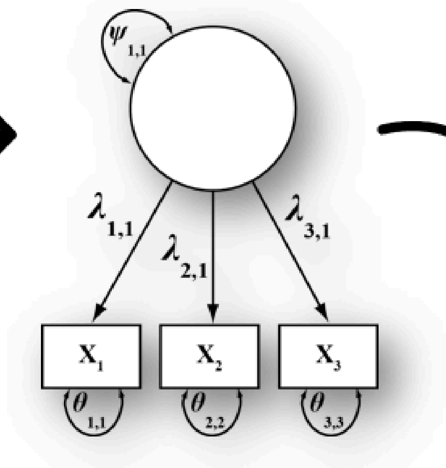
Handling Missing Data.

Listwise var2 M = 41.98.

ID	var2	μ = 30	μ = 40	μ = 50	μ = 60	μ = 70
1	40	-5.252	-2.341	-5.252	-13.984	-28.538
2	40	-5.252	-2.341	-5.252	-13.984	-28.538
3	35	-3.068	-3.068	-8.89	-20.533	-37.998
4	43	-7.26	-2.603	-3.767	-10.753	-23.561
5	42.6	-6.962	-2.538	-3.935	-11.154	-24.194
6	39	-4.698	-2.37	-5.863	-15.177	-30.314
7	45	-8.89	-3.068	-3.068	-8.89	-20.533
8	45.2	-9.066	-3.128	-3.011	-8.717	-20.243
9	50	-13.984	-5.252	-2.341	-5.252	-13.984
10	40	-5.252	-2.341	-5.252	-13.984	-28.538
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
		-69.68	-29.05	-46.63	-122.43	-256.44

Model Fit

Your Model



Your data

	X_1	X_2	X_3
X_1	0.64		
X_2	0.56	0.81	
X_3	0.48	0.53	0.77

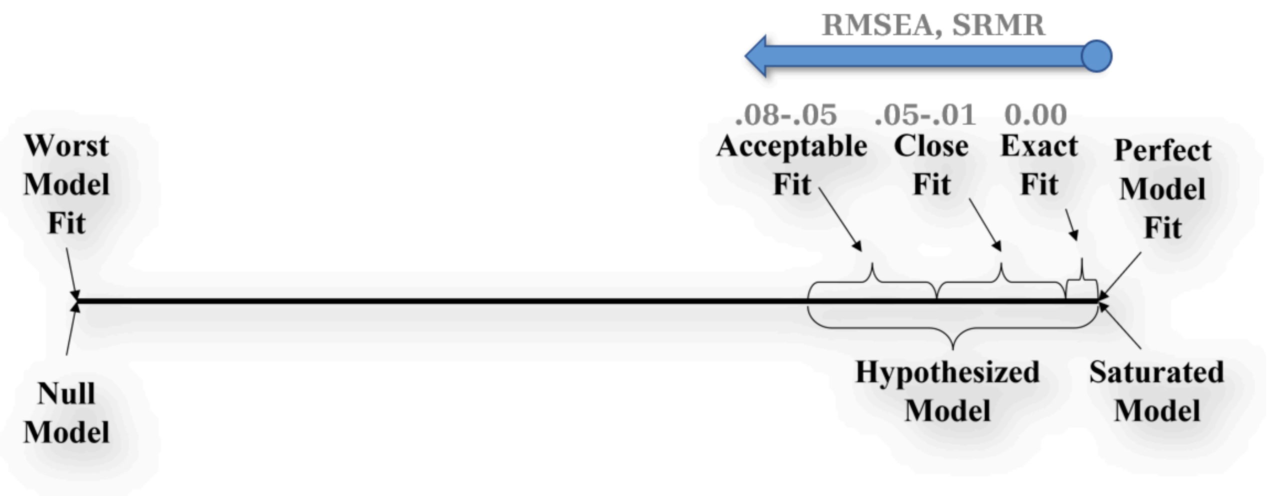
Model-implied

	X_1	X_2	X_3
X_1	0.60		
X_2	0.52	0.88	
X_3	0.53	0.58	0.73

Your data = Model-implied?

Chi-square (χ^2) = -2 * (Null Loglikelihood - Alternative Loglikelihood)

The model must be overidentified to assess fit.



Absolute Model Fit:

How far from perfect?

RMSEA, SRMR

> .10 poor fit

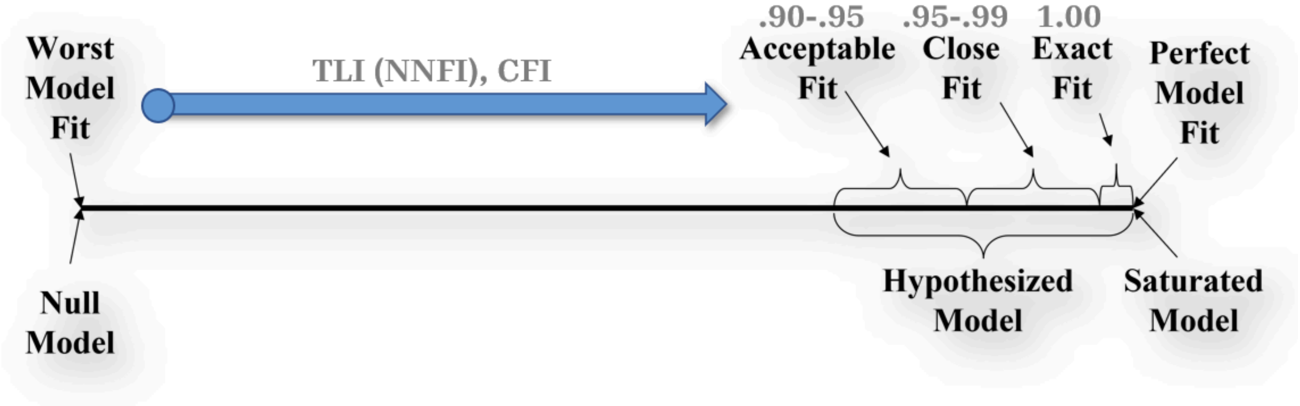
.08 - .10 mediocre fit

.05 - .08 acceptable fit

.01 - .05 close fit

.00 exact fit

Illustration adapted from Little, T. D. (2024).



Relative Model Fit:

How far from worst?

TLI, CFI...

< .85 poor fit

.85-.90 mediocre fit

.90-.95 acceptable fit

.95-.99 close fit

1.00 exact fit

Also: Modification indices, Fitted residual matrix, Parameter estimates...

Model Fit from Mplus

```
MODEL FIT INFORMATION

Number of Free Parameters                        19

Loglikelihood

          H0 Value                        -1365.848
          H1 Value                        -1351.359

Information Criteria

          Akaike (AIC)                     2769.696
          Bayesian (BIC)                   2844.509
          Sample-Size Adjusted BIC         2784.226
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              28.978
          Degrees of Freedom                      8
          P-Value                            0.0003

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.083
          90 Percent C.I.                     0.052  0.117
          Probability RMSEA <= .05            0.041

CFI/TLI

          CFI                                 0.989
          TLI                                 0.980

Chi-Square Test of Model Fit for the Baseline Model

          Value                            1939.234
          Degrees of Freedom                     15
          P-Value                            0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                               0.030
```

Chi-Square = -2[(*-1365.848*) - (*-1351.359*)] = 28.978

$$DF = \frac{v(v+1)}{2} - p = \frac{6(6+1)}{2} - 13 = 8$$

$$RMSEA = \sqrt{\frac{\frac{\chi^2_T - df_T}{N}}{df_T}} = \sqrt{\frac{\frac{28.978-8}{379}}{8}} = 0.083$$

$$CFI = \frac{(\chi^2_0 - df_0) - (\chi^2_T - df_T)}{(\chi^2_0 - df_0)}$$

$$= \frac{(1939.234-15)-(28.978-8)}{(1939.234-15)} = 0.989$$

$$TLI = \frac{(\frac{\chi^2_0}{df_0}) - (\frac{\chi^2_T}{df_T})}{(\frac{\chi^2_0}{df_0}) - 1} = \frac{(\frac{1939.234}{15}) - (\frac{28.978}{8})}{(\frac{1939.234}{15}) - 1} = 0.980$$

Reporting Model Fit

> Model estimation was performed using maximum likelihood estimation in Mplus 8. Model fit was evaluated using multiple fit indices, including the root mean square error of approximation (RMSEA) with its 90% confidence interval (Steiger, 1990), the Tucker-Lewis Index (TLI; Bentler & Bonett, 1980), and the Comparative Fit Index (CFI; Bentler, 1990). Model fit was deemed acceptable if RMSEA ≤ .08, TLI ≥ .90, and CFI ≥ .90 (Chen, 2007; Hu & Bentler, 1999). Overall, the goodness-of-fit indices indicated that the initial CFA model fit the data well, x2(24, n = 144) = 36.14; RMSEA = .059 (.000; .097) ; CFI = .980; TLI/NNFI = .970.

> Extended Reporting. Standardized residuals and modification indices indicated no areas of strain in the model solution. All factor loadings were statistically significant ($p < 0.001$)...

Any questions?