Using Mplus (and Stata) to estimate Item Response Theory (IRT) models

Parameter estimates and their conversion to the two parameter logistic (2PL) metric

Rich Jones rich_jones@brown.edu

3 Mar 2018

Overview

The 2PL IRT model for a binary test item is often given as equation of a line in point-slope form:

$$P(u = 1|\theta, a, b) = P(\theta) = F^{-1}\left(Da(\theta - b)\right)$$

where θ is the person ability parameter, a is the item discrimination parameter, b is the item difficulty parameter, D is a scaling constant used to place estimates on a logit metric to a normal probability (probit) metric, and F^-1 denotes the inverse logistic distribution function (defined later).

We will show how to obtain estimates in this metric using MPlus under different estimator and parameterization options, including:

- WLSMV estimator, probit link, theta parameterization (Wth)
- WLSMV estimator, probit link, delta parameterization (Wde)
- MLR estimator, logit link (Ml)
- MLR estimator, probit link (Mp)
- Bayes estimator (Bay)

All models fix latent trait variance to 1 and estimate all factor loadings and thresholds.

Conversion of parameter estimates to 2PL metric (with $D = 1.814 = \sqrt{\frac{\pi^2}{3}}$) and conversion to standardized factor loading metric.

Simulation

Simulate dichotomous item respose data conforming to a two parameter logistic (2PL) IRT model using Rich Jones' Stata:simirt.ado.

```
. local obs "5000"

. *                          1    2   3   4   5
. local truecorrelations ".707 .5 .85 .707 .5 .3 .707 .5 .707"

. local truepvalues      ".05 .10 .15 .17 .3 .5 .5 .75 .75"

. simirt, nitems(9) r(`truecorrelations') n(`obs') pv(`truepvalues')
there are 9 items
                                       true 2PL parameters        sample statistics
           input parameters            ------------------------                  2PL
           --------------------------  slope    slope    loca-                   -------------------
   item    corr     threshold  Pvalue  (D=1.7)  (D=1.0)  tion      corr   Pvalue slope(D=1.7) location
   ------------------------------------------------------------------------------------------------
    1      0.707     1.645     0.050   1.000    1.699    2.327     0.705  0.046   0.993       2.394
    2      0.500     1.282     0.100   0.577    0.981    2.563     0.506  0.099   0.586       2.540
    3      0.850     1.036     0.150   1.614    2.743    1.219     0.852  0.160   1.626       1.167
    4      0.707     0.954     0.170   1.000    1.699    1.350     0.711  0.176   1.012       1.310
    5      0.500     0.524     0.300   0.577    0.981    1.049     0.482  0.299   0.550       1.093
    6      0.300     0.000     0.500   0.314    0.535    0.000     0.322  0.497   0.341       0.022
    7      0.707     0.000     0.500   1.000    1.699    0.000     0.702  0.503   0.986      -0.011
    8      0.500    -0.674     0.750   0.577    0.981   -1.349     0.483  0.744   0.552      -1.359
    9      0.707    -0.674     0.750   1.000    1.699   -0.954     0.699  0.751   0.977      -0.968
   ------------------------------------------------------------------------------------------------
  All items scored 0/1. The Pvalue is the proportion item=1. Corr is the correlation of the latent
  trait and the latent response variable underlying the item (i.e., the standardized factor
  loading). 2PL refers to two parameter logistic item response theory models, which can be
  parameterized with a scaling constant D that often assumed to be 1.0 or 1.7.
```

A csv (tester.csv) is available which contains the item response data on 5000 observations

Mplus models

Mplus:WLSMV/probit/theta Input

```
TITLE:
  Variable List -

  u1 :
  u2 :
  u3 :
  u4 :
  u5 :
  u6 :
  u7 :
  u8 :
  u9 :

DATA:
  FILE = __000001.dat ;
VARIABLE:
  NAMES =
    u1 u2 u3 u4 u5 u6 u7 u8 u9 ;
  MISSING ARE ALL (-9999) ;
  CATEGORICAL =
    all
    ;
ANALYSIS:
   ESTIMATOR = wlsmv ;
   PARAMETERIZATION = theta ;
OUTPUT:
MODEL:
f by u1-u9* ;
f@1 ;
```

Mplus:WLSMV/probit/delta Input

```
TITLE:
  Variable List -

  u1 :
  u2 :
  u3 :
  u4 :
  u5 :
  u6 :
  u7 :
  u8 :
  u9 :

DATA:
  FILE = __000001.dat ;
VARIABLE:
  NAMES =
    u1 u2 u3 u4 u5 u6 u7 u8 u9 ;
  MISSING ARE ALL (-9999) ;
  CATEGORICAL =
    all
    ;
ANALYSIS:
   ESTIMATOR = wlsmv ;
   PARAMETERIZATION = delta ;
OUTPUT:
MODEL:
f by u1-u9* ;
f@1 ;
```

Mplus:MLR/logit Input

```
TITLE:
  Variable List -

  u1 :
  u2 :
  u3 :
  u4 :
  u5 :
  u6 :
  u7 :
  u8 :
  u9 :

DATA:
  FILE = __000001.dat ;
VARIABLE:
  NAMES =
    u1 u2 u3 u4 u5 u6 u7 u8 u9 ;
  MISSING ARE ALL (-9999) ;
  CATEGORICAL =
    all
    ;
ANALYSIS:
   ESTIMATOR = MLR; LINK = logit ;
OUTPUT:
MODEL:
f by u1-u9* ;
f@1 ;
```

Mplus:MLR/probit Input

```
TITLE:
  Variable List -

  u1 :
  u2 :
  u3 :
  u4 :
  u5 :
  u6 :
  u7 :
  u8 :
  u9 :

DATA:
  FILE = __000001.dat ;
VARIABLE:
  NAMES =
    u1 u2 u3 u4 u5 u6 u7 u8 u9 ;
  MISSING ARE ALL (-9999) ;
  CATEGORICAL =
    all
    ;
ANALYSIS:
   ESTIMATOR = MLR; LINK = probit ;
OUTPUT:
MODEL:
f by u1-u9* ;
f@1 ;
```

Mplus:Bayes Input

```
TITLE:
  Variable List -

  u1 :
  u2 :
  u3 :
  u4 :
  u5 :
  u6 :
  u7 :
  u8 :
  u9 :

DATA:
  FILE = __000001.dat ;
VARIABLE:
  NAMES =
    u1 u2 u3 u4 u5 u6 u7 u8 u9 ;
  MISSING ARE ALL (-9999) ;
  CATEGORICAL =
    all
    ;
ANALYSIS:
   ESTIMATOR = bayes ;
OUTPUT:
MODEL:
f by u1-u9* ;
f@1 ;
```

Math

The following hold when there is a single common latent trait with unit variance:

Where

$$P(u_{ij}) = P(u_{ij} = 1|\theta_i, a_{2PL_j}, b_{2PL_j}) = \text{invlogit}\left[Da_{2PL_j}(\theta_i - b_{2PL_j})\right]$$

$$a_{2PL} = \lambda_{Wth} = \frac{\lambda_{Wde}}{\sqrt{1-\lambda_{Wde}^2}} = \frac{\lambda_{Ml}}{D} = \lambda_{Mp} = \lambda_{Bay}$$

$$b_{2PL} = \frac{\tau_{Wth}}{\lambda_{Wth}} = \frac{\tau_{Wde}}{\lambda_{Wde}} = \frac{\tau_{Ml}}{\lambda_{Ml}} = \frac{\tau_{Mp}}{\lambda_{Mp}} = \frac{\tau_{Bay}}{\lambda_{Bay}}$$

$$\text{factor loading} = \frac{\lambda_{Wth}}{\sqrt{1+\lambda_{Wth}^2}} = \lambda_{Wde} = \frac{\lambda_{Ml}D^{-1}}{\sqrt{1+(\lambda_{Ml}D^{-1})^2}} = \frac{\lambda_{Mp}}{\sqrt{1+\lambda_{Mp}^2}} = \frac{\lambda_{Bay}}{\sqrt{1+\lambda_{Bay}^2}}$$

$$D = 1.814 = \sqrt{\frac{\pi^2}{3}}$$

Wth = WLSMV estimator, theta parameterization

Wde = WLSMV estimator, delta parameterization

Ml = MLR estimator (logit link)

Mp = MLR estimator (probit link)

Bay = Bayes estimator

Math (table)

Model	b_{2PL}	a_{2PL}	$r_{\theta u^*}$
WLSMV (theta)	$\frac{\tau_{Wth}}{\lambda_{Wth}}$	λ_{Wth}	$\frac{\lambda_{Wth}}{\sqrt{1+\lambda_{Wth}^2}}$
WLSMV (delta)	$\frac{\tau_{Wde}}{\lambda_{Wde}}$	$\frac{\lambda_{Wde}}{\sqrt{1-\lambda_{Wde}^2}}$	λ_{Wde}
MLR (logit)	$\frac{\tau_{Ml}}{\lambda_{Ml}}$	$\frac{\lambda_{Ml}}{D}$	$\frac{\lambda_{Ml}D^{-1}}{\sqrt{1+(\lambda_{Ml}D^{-1})^2}}$
MLR (probit)	$\frac{\tau_{Mp}}{\lambda_{Mp}}$	λ_{Mp}	$\frac{\lambda_{Mp}}{\sqrt{1+\lambda_{Mp}^2}}$
Bayes	$\frac{\tau_{Bay}}{\lambda_{Bay}}$	λ_{Bay}	$\frac{\lambda_{Bay}}{\sqrt{1+\lambda_{Bay}^2}}$

Where

$$P(u_{ij}) = P(u_{ij} = 1|\theta_i, a_{2PL_j}, b_{2PL_j}) = \text{invlogit}\left[Da_{2PL_j}(\theta_i - b_{2PL_j})\right]$$

and $D = 1.814 = \sqrt{\frac{\pi^2}{3}}$ or some other acceptable constant.

Results

slope estimate

These are Mplus slope estimates, not standardized, under various estimator and parameterization options

Parameter	WLSMV (theta)	WLSMV (delta)	Ml	Mp	Bayes	Population
item 1	0.985	0.702	1.967	0.982	0.962	1.000
item 2	0.613	0.523	1.158	0.611	0.606	0.577
item 3	1.596	0.847	2.938	1.617	1.684	1.614
item 4	1.056	0.726	1.870	1.059	1.060	1.000
item 5	0.542	0.477	0.911	0.543	0.544	0.577
item 6	0.348	0.328	0.561	0.348	0.347	0.314
item 7	1.001	0.707	1.702	1.000	0.992	1.000
item 8	0.548	0.481	0.944	0.547	0.547	0.577
item 9	1.028	0.717	1.841	1.021	1.007	1.000

2PL slope estimate

These are Mplus slope estimates converted to 2PL a-parameters using the conversions on the *math* page, above

Parameter	WLSMV (theta)	WLSMV (delta)	MI	Mp	Bayes	Population
item 1	0.985	0.986	1.084	0.982	0.962	1.000
item 2	0.613	0.614	0.638	0.611	0.606	0.577
item 3	1.596	1.593	1.620	1.617	1.684	1.614
item 4	1.056	1.056	1.031	1.059	1.060	1.000
item 5	0.542	0.543	0.502	0.543	0.544	0.577
item 6	0.348	0.347	0.309	0.348	0.347	0.314
item 7	1.001	1.000	0.938	1.000	0.992	1.000
item 8	0.548	0.549	0.520	0.547	0.547	0.577
item 9	1.028	1.029	1.015	1.021	1.007	1.000

Factor loading estimate

These are Mplus slope estimates converted to factor loadings using the conversions on the *math* page, above

Parameter	WLSMV (theta)	WLSMV (delta)	MI	Mp	Bayes	Population
item 1	0.702	0.702	0.735	0.701	0.693	.707
item 2	0.523	0.523	0.538	0.521	0.518	.5
item 3	0.847	0.847	0.851	0.851	0.860	.85
item 4	0.726	0.726	0.718	0.727	0.727	.707
item 5	0.477	0.477	0.449	0.477	0.478	.5
item 6	0.329	0.328	0.295	0.329	0.328	.3
item 7	0.707	0.707	0.684	0.707	0.704	.707
item 8	0.481	0.481	0.462	0.480	0.480	.5
item 9	0.717	0.717	0.712	0.714	0.710	.707

threshold estimate

These are Mplus threshold estimates, not standardized, under various estimator and parameterization options

Parameter	WLSMV (theta)	WLSMV (delta)	Ml	Mp	Bayes	Population
item 1	2.368	1.687	4.494	2.364	2.328	2.327
item 2	1.508	1.285	2.684	1.506	1.497	2.563
item 3	1.873	0.994	3.415	1.890	1.938	1.219
item 4	1.355	0.931	2.378	1.357	1.354	1.350
item 5	0.599	0.527	0.997	0.600	0.599	1.049
item 6	0.007	0.007	0.012	0.008	0.007	0.000
item 7	-0.011	-0.008	-0.017	-0.010	-0.010	0.000
item 8	-0.749	-0.656	-1.258	-0.748	-0.748	-1.349
item 9	-0.970	-0.676	-1.706	-0.966	-0.963	-0.954

2PL location estimate

These are Mplus threshold estimates converted to 2PL b-parameters using the conversions on the *math* page, above

Parameter	WLSMV (theta)	WLSMV (delta)	Ml	Mp	Bayes	Population
item 1	2.404	2.403	2.285	2.407	2.420	2.327
item 2	2.460	2.457	2.318	2.465	2.470	2.563
item 3	1.174	1.174	1.162	1.169	1.151	1.219
item 4	1.283	1.282	1.272	1.281	1.277	1.350
item 5	1.105	1.105	1.094	1.105	1.101	1.049
item 6	0.020	0.021	0.021	0.023	0.020	0.000
item 7	-0.011	-0.011	-0.010	-0.010	-0.010	0.000
item 8	-1.367	-1.364	-1.333	-1.367	-1.367	-1.349
item 9	-0.944	-0.943	-0.927	-0.946	-0.956	-0.954

The deal with D

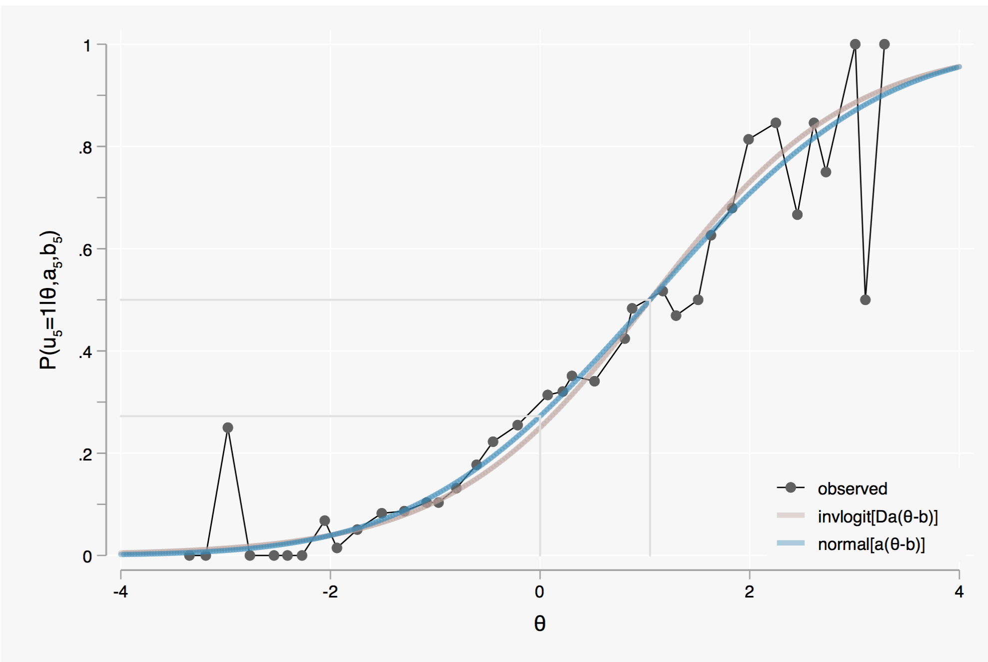
Lord (Lord F. Psychometric Monographs. 1952) introduced the item response theory model using a normal probability (i.e., probit) regression framework. Birnbaum (1968, Some latent trait models (chapter 17) in Statistical Theories of Mental Test Scores [F. Lord and M. Novick, eds]) expressed the model in the logistic regression framework. Logistic regression is more robust and much easier computationally, and in an era without powerful desktop computers quickly became the prefered parameterization. Both Lord's normal probability model and the Birnbaum logistic model made use of an assumed normally distributed underlying latent variable, but because logits and probits are on a different metric, the slopes and thresholds are on a different scale.

The scaling constant D was introduced to place logistic regression parameters on the scale of the normal probability or probit regression parameters (so that Birnbaum's results are backward-compatible to Lord's results). But probits and logits are non-linearly related, and a constant transformation is only a crude approximation (see Camilli G. Journal of Educational and Behavioral Statistics. 1994:19(3);293). What to use as a constant for D varies by author, software, situation, etc. Some use $D = 1.6$, others $D = 1.814 = \sqrt{\frac{\pi^2}{3}}$, but $D = 1.7$ is probably the most common.

Because D only applies when logistic regression is used, the constant D only appears when the estimator is MLR and the link is logit. When regressions are conducted in the normal probability (probit) metric, in *most* conditions the IRT a parameter is directly estimated by the measurement slope. The exception is when the delta parameterization is requested, because in this situation the response variable is a *latent* response variable with unit variance (when $\Delta = \text{I}$), and the slope estimate is on the scale of a standardized regression coefficient (i.e., a factor loading).

Predicted probabilities with probit or logit

Once we have item discrimination (a) and item difficulty (b) parameters in the 2PL (two parameter logistic) metric using a scaling constant (D) – that we now know are the same thing as parameter estimates in the 2PN (two parameter normal probability) metric – we can use the inverse logit function $\left(1 + exp^{-x}\right)^{-1}$ or the cumulative standard normal distribution function ($\Phi(x)$) to plot an item response function. Below we show an example for item u_5:



Reference lines are included that show the expected proportion making an error when $\theta = 0$ and the item difficulty (point on θ where expected probability of a randomly selected person with ability level θ makes an error on the item is 50%).

Epected proportion making an error when $\theta = 0$

Let's start with the expected proportion making an error on item 5 at the average ability level ($\theta = 0$). We'll use F^{-1} to denote the inverse logit function and D=1.814:

$$P(u_5 = 1|\theta) = F^{-1}\left[Da_5(\theta - b_5)\right]$$

$$P(u_5 = 1|\theta = 0) = F^{-1}\left[Da_5(0 - b_5)\right] = F^{-1}\left[-Da_5 b_5)\right]$$

$$P(u_5 = 1|\theta) = F^{-1}\left[-1.814 \times 0.577 \times 1.049\right]$$

$$P(u_5 = 1|\theta) = F^{-1}\left[-1.098\right]$$

$$P(u_5 = 1|\theta) = 0.250$$

We could also use the cumulative normal (yes, with the 2PL parameter estimates):

$$P(u_5 = 1|\theta) = \Phi\left[-a_5 b_5\right] = 0.272$$

Note that this is not the same as the pvalue specified in the simulation condition, which was .3. That is because the pvalue is the total proportion of people expected to make an error on u5 across all levels of θ.

Expected proportion making an error

If you wanted to know the expected proportion of people making an error on an item, that information is directly estimated by the item threshold Mplus WLS delta estimator is used.

$$P(u = 1) = \Phi\left[-\tau_{Wde}\right]$$

Where Φ indicates the cumulative normal distribution function. If all you had were 2PL (really, 2PN) parameter estimates, then:

$$P(u = 1) = \Phi\left[-br_{\theta u^*}\right] = \Phi\left[\frac{-ba}{\sqrt{1+a^2}}\right]$$

where $r_{\theta u^*}$ is the correlation of the latent trait and the latent response variable underlying the observed dichotomous item, i.e. the factor loading for the item. For example, with item 5:

$$P(u_5 = 1) = \Phi\left[-1.049\frac{0.577}{\sqrt{1+0.577^2}}\right]$$

$$P(u_5 = 1) = \Phi\left[-0.524\right]$$

$$P(u_5 = 1) = 0.300$$

We could also use the inverse logitistic function (F^{-1}):

$$P(u_5 = 1) = F^{-1}\left[-Db\frac{a}{\sqrt{1+a^2}}\right]$$

$$P(u_5 = 1) = F^{-1}\left[-D \times 1.049\frac{0.577}{\sqrt{1+0.577^2}}\right]$$

using $D = 1.7$,

$$P(u_5 = 1) = F^{-1}\left[-0.891\right]$$

$$P(u_5 = 1) = 0.291$$

If you have the factor loading and the proportion in error, what are the 2PL (2PN) item parameters?

And on the previous page we decided that

$$P(u = 1) = \Phi\left[-b r_{\theta u^*}\right]$$

so

$$\Phi^{-1}(P(u = 1)) = -b r_{\theta u^*}$$

$$b = -1 \times \frac{\Phi^{-1}[P(u=1)]}{r_{\theta u^*}}$$

And, we know that

$$a = \frac{r}{1-r^2}$$

And this explains how the data were generated using correlations and item error probabilities.

Stata SEM for kicks

Stata:sem, ologit item parameters are on the scale of Mplus/MLR estimates

```
. gsem F -> u1-u9 , ologit var(F@1)

Fitting fixed-effects model:

Iteration 0:   log likelihood = -22705.467
Iteration 1:   log likelihood = -22705.467

Refining starting values:

Grid node 0:   log likelihood = -21890.599

Fitting full model:

Iteration 0:   log likelihood = -21890.599  (not concave)
Iteration 1:   log likelihood = -21079.642
Iteration 2:   log likelihood = -20793.487
Iteration 3:   log likelihood = -20773.607
Iteration 4:   log likelihood = -20772.511
Iteration 5:   log likelihood = -20772.503
Iteration 6:   log likelihood = -20772.503

Generalized structural equation model           Number of obs     =      5,000

Response       : u1
Family         : ordinal
Link           : logit

Response       : u2
Family         : ordinal
Link           : logit

Response       : u3
Family         : ordinal
Link           : logit

Response       : u4
Family         : ordinal
Link           : logit

Response       : u5
Family         : ordinal
Link           : logit

Response       : u6
Family         : ordinal
Link           : logit

Response       : u7
Family         : ordinal
Link           : logit

Response       : u8
Family         : ordinal
Link           : logit

Response       : u9
Family         : ordinal
Link           : logit

Log likelihood = -20772.503

 ( 1)  [/]var(F) = 1
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
u1           |
```

```
            F |   1.966313   .1407637    13.97   0.000     1.690422    2.242205
--------------+----------------------------------------------------------------
u2            |
            F |   1.158222   .0732518    15.81   0.000     1.014651    1.301793
--------------+----------------------------------------------------------------
u3            |
            F |   2.941467   .1959325    15.01   0.000     2.557446    3.325487
--------------+----------------------------------------------------------------
u4            |
            F |   1.870557   .0960063    19.48   0.000     1.682388    2.058726
--------------+----------------------------------------------------------------
u5            |
            F |   .9109884   .0490784    18.56   0.000     .8147965     1.00718
--------------+----------------------------------------------------------------
u6            |
            F |   .5605937   .0391296    14.33   0.000     .4839011    .6372863
--------------+----------------------------------------------------------------
u7            |
            F |   1.702704   .0805913    21.13   0.000     1.544748     1.86066
--------------+----------------------------------------------------------------
u8            |
            F |   .9438373   .0541741    17.42   0.000     .8376579    1.050017
--------------+----------------------------------------------------------------
u9            |
            F |   1.840695   .1024814    17.96   0.000     1.639835    2.041555
--------------+----------------------------------------------------------------
/u1           |
         cut1 |   4.493492     .18976                       4.12157    4.865415
--------------+----------------------------------------------------------------
/u2           |
         cut1 |   2.683872   .0754859                      2.535922    2.831821
--------------+----------------------------------------------------------------
/u3           |
         cut1 |   3.417891   .1805091                        3.0641    3.771683
--------------+----------------------------------------------------------------
/u4           |
         cut1 |    2.37844   .0847616                      2.212311     2.54457
--------------+----------------------------------------------------------------
/u5           |
         cut1 |   .9972447   .0382557                      .9222648    1.072225
--------------+----------------------------------------------------------------
/u6           |
         cut1 |   .0121417   .0303604                     -.0473637     .071647
--------------+----------------------------------------------------------------
/u7           |
         cut1 |   -.016534   .0426188                     -.1000653    .0669974
--------------+----------------------------------------------------------------
/u8           |
         cut1 |  -1.258383   .0420198                      -1.34074   -1.176025
--------------+----------------------------------------------------------------
/u9           |
         cut1 |  -1.705365   .0708646                     -1.844257   -1.566473
--------------+----------------------------------------------------------------
       var(F)|          1  (constrained)
-------------------------------------------------------------------------------
```