SAS `proc calis`: The basics[1]

STA431 Winter/Spring 2013

[1] See last slide for copyright information.

Overview

What it is and what it does

- SAS `proc calis` is model fitting software.
- It fits classical structural equation models to data, using numerical maximum likelihood (or optionally, other methods).
- Most of the output is about the details of the numerical search and how well the model fits.
- This is a narrow focus, compared to most other SAS procedures.
- Still, SAS tells you more than you need or want to know — as usual.

Three programs

- `proc calis` incorporates three programs that originated outside of SAS.
- They all use different, unrelated syntax for specifying the model.
- We will use the `lineqs`[2] syntax, which is the most convenient.

- First, read and label the data as usual in a SAS *data step*.

[2] Bentler and Weeks, *British Journal of Mathematical and Statistical Psychology, 1980.*

Specifying the model

Using `lineqs` syntax

Input includes:

- Names of the observable variables.
- Model equations, pretty much as you would write them by hand
  - Including the regression coefficients and the error terms – you name them.
  - No intercepts: The model is given in centered form and SAS bases everything on the sample covariance matrix.
  - Naming rules: Names of latent variables (including error terms) must begin with the letter F, D or E.
- Names must also be given to the variances and covariances of the explanatory variables and error terms. Anything unspecified is assumed zero.
- In the end, you give names to *all* the non-zero parameters in your model.

What happened to the intercepts?

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) \right\}$$

- Remember, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are both functions of $\boldsymbol{\theta}$.
- For regression without measurement error, expected values and intercepts are identifiable, but if there are latent variables that's rare.
- Re-parameterize, absorbing expected values and intercepts into $\boldsymbol{\mu}$.
- Estimate $\boldsymbol{\mu}$ with $\overline{\mathbf{x}}$ and it's gone.
- This is just a technical trick to allow the likelihood to have a unique maximum.
- But it does no harm, because *relationships* between variables are represented by the covariances.

Maximum likelihood

$$L(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2}(2\pi)^{-np/2} \exp -\frac{n}{2}\left\{tr\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)\right\}$$

$$L_2(\boldsymbol{\theta}) = |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-n/2}(2\pi)^{-np/2} \exp -\frac{n}{2}\left\{tr\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)\right\}$$

- Can maximize $L(\boldsymbol{\Sigma})$ over all $\boldsymbol{\Sigma} \in \mathcal{M}$, or maximize $L_2(\boldsymbol{\theta})$ over all $\boldsymbol{\theta} \in \Theta$.
- If the function connecting $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}$ is one-to-one and there is the same number of $\boldsymbol{\theta}$ and unique $\boldsymbol{\Sigma}$ values, call the parameter $\boldsymbol{\theta}$ *just identifiable*.
- In this case it's the same problem, and
- The invariance principle can be used to go back and forth between $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\theta}}$.
- Otherwise . . .

Maximize $L_2(\boldsymbol{\theta})$ over all $\boldsymbol{\theta} \in \Theta$

$$L_2(\boldsymbol{\theta}) = |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-n/2}(2\pi)^{-np/2} \exp -\frac{n}{2}\left\{tr\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\right)\right\}$$

- Actually, maximize the *log* likelihood.
- Well, actually, minimize the minus 2 log likelihood.
- Well, actually, minimize the minus 2 log likelihood plus a carefully chosen constant.

- The constant is based on the likelihood ratio test for goodness of model fit.

Likelihood ratio tests

In general

Setup

$$Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} P_\theta,\ \theta \in \Theta,$$
$$H_0 : \theta \in \Theta_0 \subset \Theta \text{ v.s. } H_1 : \theta \in \Theta_1 = \Theta \cap \Theta_0^c$$

Test Statistic:

$$G^2 = -2 \ln \left(\frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \right)$$

What to do

And how to think about it

$$G^2 = -2 \ln \left(\frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \right)$$

- Maximize the likelihood over the whole parameter space. You already did this to calculate the MLE. Evaluate the likelihood there. That's the denominator.
- Maximize the likelihood over just the parameter values where H_0 is true – that is, over Θ_0. This yields a restricted MLE. Evaluate the likelihood there. That's the numerator.
- The numerator cannot be larger, because $\Theta_0 \subset \Theta$.
- If the numerator is a *lot* less than the denominator, the null hypothesis is unbelievable, and
  - The ratio is close to zero
  - The log of the ratio is a big negative number
  - -2 times the log is a big positive number
  - Reject H_0 when G^2 is large enough.

Distribution of G^2 when H_0 is true

Given some technical conditions,

- G^2 has an approximate chi-squared distribution under H_0 for large n.
- Degrees of freedom equal number of (non-redundant) equalities specified by H_0.
- Reject H_0 when G^2 is larger than the chi-squared critical value.

Goodness of fit test for a covariance structure model
Multivariate normal data

Call it a "covariance structure" model because $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$.

- Compare fit of model to fit of the *best possible* model.
- The best possible model is the unrestricted multivariate normal:
  - Estimate $\boldsymbol{\mu}$ with $\overline{\mathbf{x}}$.
  - Estimate $\boldsymbol{\Sigma}$ with $\widehat{\boldsymbol{\Sigma}}$.
- Covariance structure model is re-parameterized to get rid of intercepts, so again, $\boldsymbol{\mu}$ is estimated with $\overline{\mathbf{x}}$.
- Compare

$$\ln L\left(\widehat{\boldsymbol{\Sigma}}\right) \text{ to } \ln L\left(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)$$

Likelihood ratio test

For goodness of model fit

Difference in fit (times two):

$$
\begin{aligned}
G^2 &= 2\left(\ln L\left(\widehat{\boldsymbol{\Sigma}}\right) - \ln L\left(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)\right) \\
&= -2\ln\left(\frac{L\left(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}{L\left(\widehat{\boldsymbol{\Sigma}}\right)}\right)
\end{aligned}
$$

It looks like a likelihood ratio test statistic.

More details

$$G^2 = -2 \ln \left(\frac{L\left(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}{L\left(\widehat{\boldsymbol{\Sigma}}\right)} \right)$$

If the covariance structure model is correct and

- The parameter vector is identifiable, and
- There are more unique variances and covariances in $\boldsymbol{\Sigma}$ than there are model parameters in $\boldsymbol{\theta}$, and
- Some other technical conditions hold

Then for large samples, G^2 has an approximate chi-squared distribution, with degrees of freedom the number of variances-covariances *minus* the number of model parameters.

Simplify $G^2 = 2\left(\ln L\left(\widehat{\boldsymbol{\Sigma}}\right) - \ln L\left(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)\right)$

Recalling $L(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2}(2\pi)^{-np/2} \exp -\frac{n}{2}\left\{tr\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)\right\}$,

$$\begin{aligned} G^2 &= -2\ln L(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})) - [-2\ln L(\widehat{\boldsymbol{\Sigma}})] \\ &= n\left(tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}) + \ln|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})| - \ln|\widehat{\boldsymbol{\Sigma}}| - p\right) \end{aligned}$$

A cute way to maximize the likelihood over $\boldsymbol{\theta} \in \Theta$

- Minimize $G^2(\boldsymbol{\theta})$: Just -2 log likelihood plus a constant.

$$
\begin{aligned}
G^2(\boldsymbol{\theta}) &= -2 \ln L(\boldsymbol{\Sigma}(\boldsymbol{\theta})) - [-2 \ln L(\widehat{\boldsymbol{\Sigma}})] \\
&= n \left(tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}) + \ln |\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln |\widehat{\boldsymbol{\Sigma}}| - p \right)
\end{aligned}
$$

- Actually, minimize the "Objective Function"

$$
tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}) + \ln |\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \ln |\widehat{\boldsymbol{\Sigma}}| - p
$$

- Multiply by n (or $n-1$) to get the G^2 statistic.
- This is what SAS `proc calis` does.

Saturated models

All the degrees of freedom in the data are "soaked up" by the model.

- If there are the same number of moment structure equations and unknown parameters and the parameter is identifiable, there is a one-to-one function between $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\theta}}$.
- In this case the parameter is called *just identifiable.*
- $L\left(\widehat{\boldsymbol{\Sigma}}\right) = L\left(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)$
- $G^2 = 0, df = 0$ and the standard test for goodness of fit does not apply.
- The model may still be testable some other way.

What does `proc calis` give us?

- An indication of whether the numerical search went okay.
- MLEs of all the parameters, standard errors and Z tests of $H_0 : \theta_j = 0$.
- The -2 log likelihood at the MLE, plus a constant.
- Likelihood ratio test for goodness of fit.

With the -2 log likelihood at the MLE (plus a constant) we can

- Fit a full and a reduced model.
- Test null hypothesis that the reduced model holds, using a LR test.
- G^2 is a difference between two -2 log likelihoods.
- The constant $(-2 \ln L(\widehat{\boldsymbol{\Sigma}}))$ cancels.

- This is all we really need.

Copyright Information