How is the 95% confidence interval of the variance component in a mixed model calculated?

Authored by
stats writer

July 1, 2024

The 95% confidence interval of the variance component in a mixed model is calculated by using the maximum likelihood estimation method. This involves calculating the standard error of the variance component, which is then used to determine the upper and lower bounds of the confidence interval. The confidence interval represents the range of values within which the true value of the variance component is likely to fall with 95% certainty. This calculation takes into account the variability within the data and the sample size, providing a reliable estimate of the variance component in the mixed model.

How is the 95% CI of the variance component in a mixed model
calculated? | Stata FAQ

NOTE: Code for this page was tested in Stata 12.

Below is a mixed model, where female is used to predict mathach,
the model includes a random intercept, where the level 2 units are defined by
the variable id.

xtmixed mathach female || id:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log restricted-likelihood = -23528.021
Iteration 1: log restricted-likelihood = -23528.021

Computing standard errors:

Mixed-effects REML regression Number of obs = 7185
Group variable: school Number of groups = 160

Obs per group: min = 14
avg = 44.9
max = 67

Wald chi2(1) = 62.83
Log restricted-likelihood = -23528.021 Prob > chi2 = 0.0000

--

mathach | Coef. Std. Err. z P>|z|

-------------+--

female | -1.358992 .1714418 -7.93 0.000 -1.695012 -1.022972
_cons | 13.34494 .2546749 52.40 0.000 12.84579 13.8441

--

--

Random-effects Parameters | Estimate Std. Err.

-----------------------------+--

id: Identity |

sd(_cons) | 2.858072 .1798756 2.526399 3.233288

-----------------------------+--

sd(Residual) | 6.232982 .0525962 6.130743 6.336926

--

LR test vs. linear regression: chibar2(01) = 938.95 Prob >= chibar2 = 0.0000

At the bottom of the output is the table that displays the estimates of the standard deviation of random effects (variances are shown if the var option is used). The standard deviation (SD) of the random intercept is displayed on the line beginning sd(_cons) and is estimated as 2.86 for this model. In addition to the estimate of the random intercept, the table includes the standard error of the estimate, and the 95% confidence interval (CI). You may notice that the lower bound for the CI is not equal to 2.858 - 1.96*.18 (= 2.506), the upper bound of the CI is also seemingly inconsistent with the output. Why isn't the CI calculated in the usual way (i.e. b +/- 1.96*se)? The CI

actually is calculated
in the usual way, it's just that the displayed
values just aren't the correct values to use to in the calculation.

To understand why the values displayed are not used calculate the 95% CI it is important to know that Stata doesn't actually
estimate the SD (or the variance) of the random effects, instead it estimates
the natural log of the SD (i.e., ln(sd)), which assures that the standard
deviation will always be positive. Stata then exponentiates the estimates so
that what you see is the SD (or the variance if the var option is used). Knowing this, we can see that the correct formula
for the confidence interval involves the natural logs of the coefficients and
standard errors displayed, specifically:

CI = exp(ln(sd) +/- 1.96*(ln(sesd))

where sesd is the standard error of the estimate of the

SD of the random
effect. Calculating the CI this was assures that the lower bound of the CI will never be below zero.
It also results in CIs that are not symmetric around the estimate of the SD or variance.

We can use the returned results that Stata stores after the
model is run to calculate the CI. The ln(sd) of the random intercept is stored
in the rather odd looking macro _b , and the standard error is stored in
_se. We can use those values, along with the display command, to calculate the lower and upper bounds of the CI. The two lines of
code below do just that:

```
display exp(_b - 1.96*_se)
2.5263929
display exp(_b + 1.96*_se)
3.233295
```

These values should be very close to the bounds of the CI shown in the

output, there may be some (very) small differences because Stata uses a more precise approximation to the correct z value for a 95% CI (1.959964...) than we used (1.96). Note that in order to get the confidence interval for the variance you will need to square the upper and lower bounds of the CI, the same way that you square the SD to get the variance.