# How can I analyze a nested model using mixed in Stata?

Authored by

**stats writer**

July 1, 2024

Analyzing nested models using mixed in Stata involves utilizing the "mixed" command which allows for the estimation of multilevel or hierarchical models with nested data structures. This allows for the examination of both fixed and random effects within the model, as well as the correlation between these effects. The "mixed" command also allows for the inclusion of covariates and the specification of different covariance structures. Additionally, Stata provides several options for model diagnostics and hypothesis testing to assess the goodness of fit of the nested model. Overall, utilizing the "mixed" command in Stata provides a comprehensive and efficient approach to analyzing nested models.

## How can I analyze a nested model using mixed? | Stata FAQ

## Please note: The following example is for illustrative purposes only. The data presented is not meant to recommend or encourage the estimation of random effects on categorical variables with very few unique levels.

## Consider the following nested experiment: A study was conducted measuring the thickness of the oxide layer on silicon wafers. The wafers were produced on two different machines (`source`). Four lots of wafers were selected at random from each machine. From each lot three wafers were selected at random to be measured. Finally, on each `wafer` three positions were selected. So, we have `position` nested in `wafer`, `wafer` nested in `lot` which is nested in `source`. The primary concern of this experiment is to

## determine whether the two machines (`source`) differ in the thickness of their oxide layers.

## Let's load the data and look at our sample.

`use https://stats.idre.ucla.edu/stat/data/thickness, clearlist in 1/10`

## +------------------------------------------+

## | source lot wafer position thickn~s |

## |------------------------------------------|

## 1. | 1 1 1 1 2006 |

## 2. | 1 1 1 2 1999 |

## 3. | 1 1 1 3 2007 |

## 4. | 1 1 2 1 1980 |

## 5. | 1 1 2 2 1988 |

## |------------------------------------------|

## 6. | 1 1 2 3 1982 |

## 7. | 1 1 3 1 2000 |

## 8. | 1 1 3 2 1998 |

## 9. | 1 1 3 3 2007 |

## 10. | 1 2 1 1 1991 |

## +------------------------------------------+

`tabstat thickness, by(source) stat(n mean sd)`

## Summary for variables: thickness

**by categories of: source**

**source | N mean sd**

**---------+-----------------------------**

**1 | 36 1995.111 7.531943**

**2 | 36 2005.194 14.86668**

**---------+-----------------------------**

**Total | 72 2000.153 12.75518**

**---------------------------------------**

**Next, we will need to create a variable that indicates `lot` nested in `source`. We**
**will do this using the `egen` command with the`group` function.**

```
egen lotinsource = group(lot source), labeltab lotinsource
```

**group(lot |**
**source) | Freq. Percent Cum.**

**-----------+-----------------------------------**

**1 1 | 9 12.50 12.50**

**1 2 | 9 12.50 25.00**

**2 1 | 9 12.50 37.50**

**2 2 | 9 12.50 50.00**

**3 1 | 9 12.50 62.50**

**3 2 | 9 12.50 75.00**

4 1 | 9 12.50 87.50

4 2 | 9 12.50 100.00

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

Total | 72 100.00

From the table above it looks `lot` is crossed with`source`. This is not the case since a `lot` drawn from `source`1 is a different from a `lot` that is drawn from `source`2. Fortunately, `mixed` will be able to sort this out for us. Here is one way to parameterize this model.

`mixed thickness i.source || lotinsource: || wafer:, var`

Performing gradient-based optimization:

Iteration 0: log likelihood = -228.43197

Iteration 1: log likelihood = -228.43197

Computing standard errors:

Mixed-effects ML regression Number of obs = 72

-----------------------------------------------------------

| No. of Observations per Group

Group Variable | Groups Minimum Average Maximum

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

lotinsource | 8 9 9.0 9

wafer | 24 3 3.0 3

-------------------------------------------------------------

Wald chi2(1) = 2.03

Log likelihood = -228.43197 Prob > chi2 = 0.1537

------------------------------------------------------------------------------

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

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

2.source | 10.08333 7.068711 1.43 0.154 -3.771085 23.93775

_cons | 1995.111 4.998333 399.16 0.000 1985.315 2004.908

------------------------------------------------------------------------------

------------------------------------------------------------------------------

Random-effects Parameters | Estimate Std. Err.

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

lotinsource: Identity |

var(_cons) | 86.58149 50.1892 27.79739 269.6783

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

wafer: Identity |

var(_cons) | 35.86577 14.18759 16.51834 77.87428

**-----------------------------+---------------------------------------------**

**var(Residual) | 12.56944 2.565726 8.424908 18.75282**

**------------------------------------------------------------------------------**

**LR test vs. linear model: chi2(2) = 100.65 Prob > chi2 = 0.0000**

**Note: LR test is conservative and provided only for reference.**

**Note that the test for differences in `source` is not significant. Also, note that the variable `position` does not appear in the model. That's because variability due to `position` is accounted for by the residual variance. In the output above, lots nested in source (`lotinsource`) has a variance of 86.58, `wafer` has a variance of 35.87 and `position` (residual) has a variance of 12.57.**

**There is an alternative way to parameterize this model that is somewhat more efficient.**

```
mixed thickness i.source || lotinsource: || _all: R.wafer, var
```

**Performing EM optimization:**

**Performing gradient-based optimization:**

**Iteration 0: log likelihood = -228.43197**

**Iteration 1: log likelihood = -228.43197**

**Computing standard errors:**

**Mixed-effects ML regression Number of obs = 72**

**-----------------------------------------------------------**

**| No. of Observations per Group**

**Group Variable | Groups Minimum Average Maximum**

**----------------+-----------------------------------------**

**lotinsource | 8 9 9.0 9**

**_all | 8 9 9.0 9**

**-----------------------------------------------------------**

**Wald chi2(1) = 2.03**

**Log likelihood = -228.43197 Prob > chi2 = 0.1537**

**-----------------------------------------------------------------------------**

**thickness | Coef. Std. Err. z P>|z|**

**-------------+-----------------------------------------------------------------**

**2.source | 10.08333 7.068711 1.43 0.154 -3.771085 23.93775**

**_cons | 1995.111 4.998333 399.16 0.000 1985.315 2004.908**

------------------------------------------------------------------------------

------------------------------------------------------------------------------

Random-effects Parameters | Estimate Std. Err.

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

lotinsource: Identity |

var(_cons) | 86.58149 50.1892 27.79739 269.6783

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

_all: Identity |

var(R.wafer) | 35.86577 14.18759 16.51834 77.87427

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

var(Residual) | 12.56944 2.565726 8.424908 18.75282

------------------------------------------------------------------------------

LR test vs. linear model: chi2(2) = 100.65 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

All of the results as the same as in our first model, however some of the labels for the
variance components differ.

This design is completely balanced so the `mixed` results will be identical to those

## using the `anova` command.

```
anova thickness source / lot|source wafer|lot|source
```

**Number of obs = 72 R-squared = 0.9478**

**Root MSE = 3.54534 Adj R-squared = 0.9227**

**Source | Partial SS df MS F Prob > F**

**-----------------+--------------------------------------------------**

**Model | 10947.9861 23 475.999396 37.87 0.0000**

**|**

**source | 1830.125 1 1830.125 1.53 0.2629**

**lot|source | 7195.19444 6 1199.19907**

**-----------------+--------------------------------------------------**

**wafer|lot|source | 1922.66667 16 120.166667 9.56 0.0000**

**|**

**Residual | 603.333333 48 12.5694444**

**-----------------+--------------------------------------------------**

**Total | 11551.3194 71 162.69464**

**For more information on nested models see: "Multilevel and Longitudinal Modeling Using Stata" by Sophia Rabe-Hesketh and Anders Skrondal (2012)**