Multilevel Models with the Von Guttenberg Tree Data

Andy Grogan-Kaylor

28 Jul 2020 14:02:30

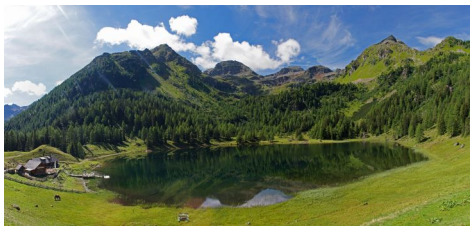

Figure 1: Norway Spruce and Larch Forest in Austrian Alps, https://ec.europa.eu/jrc/en/research-topic/forestry/qr-tree-project/norway-spruce

Data Source

The data used in this example are derived from the R package *Functions and Datasets for "Forest Analytics with R"*.

According to the documentation, the source of these data are: "von Guttenberg's Norway spruce (Picea abies [L.] Karst) tree measurement data."

Figure 2: Old Tjikko, a 9,550 Year Old Norway Spruce in Sweden

The documentation goes on to further note that:

> "The data are measures from 107 trees. The trees were selected as being of average size from healthy and well stocked stands in the Alps."

```
.     use gutten.dta, clear
```

Variables

`site` Growth *quality* class of the tree's habitat. 5 levels.

`location` Distinguishes tree *location*. 7 levels.

`tree` An identifier for the tree within location.

`age.base` The tree age taken at ground level.

It might be best to use a centered age variable, centered at the grand mean of tree age:

```
. egen ageMEAN = mean(age_base)

. generate ageCENTERED = age_base - ageMEAN
```

`height` Tree height, m.

`dbh.cm` Tree diameter, cm.

`volume` Tree volume.

`age.bh` Tree age taken at 1.3 m.

`tree.ID` A factor uniquely identifying the tree.

Your Quantitative Forestry Tasks (Should You Choose To Accept Them)

I clearly need to learn more about quantitative approaches to forestry, but I would like to model tree height as a function of tree age (measured at base), site, and location.

Spaghetti Plot

Use `spagplot` to examine these relationships. Remember that the `spagplot` syntax is `spagplot y x, id(id)`. What is the most appropriate id variable here?

```
. spagplot height ageCENTERED, id(tree_ID) scheme(michigan)

. graph export myspagplot.png, width(500) replace
(file myspagplot.png written in PNG format)
```

`spagplot height ageCENTERED, id(tree_ID) scheme(s1rcolor)` also works well here.

Multilevel Models

Unconditional Model

Run an unconditional model with `mixed`. Calculate the ICC. What does this number tell you?

Conditional Model

Now run a model with `mixed` where you examine the relationship of tree age, site, and location with tree height. Do any of your variables need to be treated as indicator (`i.`) variables? What do you find?

What if you examine age of tree^2 as part of your model?

```
generate ageMEAN2 = ageMEAN^2
```

Does the ICC have any meaning after your *conditional* model?

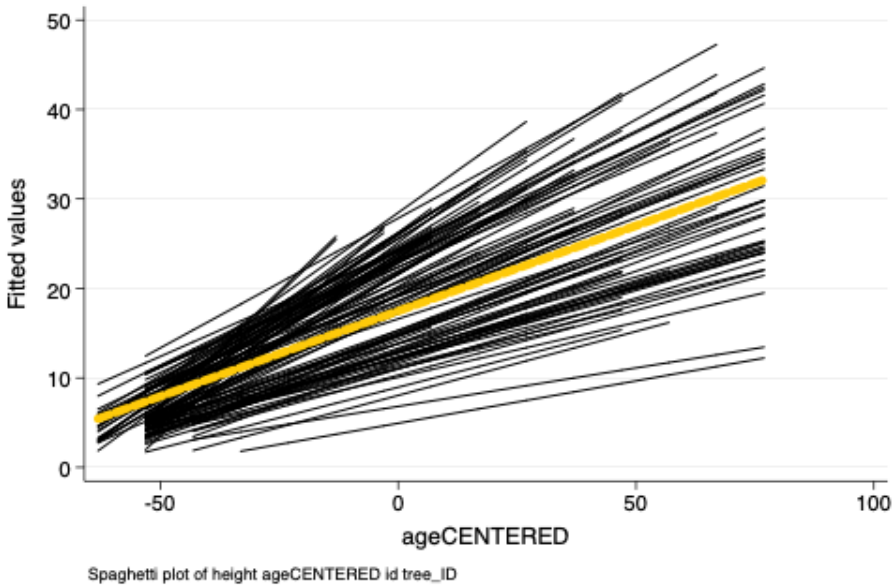

Figure 3: Spaghetti Plot