Homework I

J Lacasa

2026-02-03

Read the exercises below, answer them, and knit the .Rmd file into an HTML of PDF file. Rename it to "YourLastName_hw1.Rmd" and submit it on CANVAS by Tue, February 10th.

1 In your own words, describe why you need statistical models in your research project.

2 Using mathematical notation, describe the simplest statistical model you could think of.

2.1 What are the assumptions behind this model?

2.2 How can said assumptions be relaxed?

3 In the context of mixed-effects statistical models:

3.1 Mention what you think are the most important characteristics/differences between fixed effects vs. random effects.

	Fixed effect	**Random effect**
Method of estimation	Maximum likelihood, least squares	Restricted maximum likelihood (shrinkage)

3.2 In your own words, explain to the applied user what you think are three key factors to decide whether to model an effect as fixed or random.

4 Write down the statistical models for the following data using mathematical notation.

4.1 Data generated by a designed experiment

A designed experiment studying the effects of plant density (low, medium, high) and genotype (5 genotypes) effects with a 3×5 factorial treatment structure and a split-plot design in an RCBD, where plant density (2 levels) is at the whole-plot level, and the genotypes are at the spli-plot level.

4.2 Opportunistic data

An opportunistic set of data that is the combination of multiple variety trials. The dataset contains information from 15 years, 60 locations (not always does a year include all 60 locations), and over 200 genotypes that are inconsistently present across locations (i.e., not always does a site-year include all genotypes). The objective thus far is to model yield as a function of years, location, and genotype, with a focus on studying the variability across years, locations, and genotypes.

4.3 Observational data

In this last example, the data correspond to observations of the native species in a given place. The data were collected in 40 points in a region in the Argentinean Patagonian Steppe (also known as Patagonian Desert) by a group of researchers. They collected 40 points in approximately random locations in that region for 25 years. At each site, the researchers randomly tossed a 1m^2 loop and counted the total number of exotic (i.e., non-native species) plants wherever it fell. The researchers registered the total number of plants and the total number of exotic plants. They are interested in the overall proportion of exotic plants, and whether it has been increasing in the past decades. One can assume that the overall number of total plants per m^2 is approximately contant in all the points.

5 Write down the statistical model for the following data, and fit said model using R.

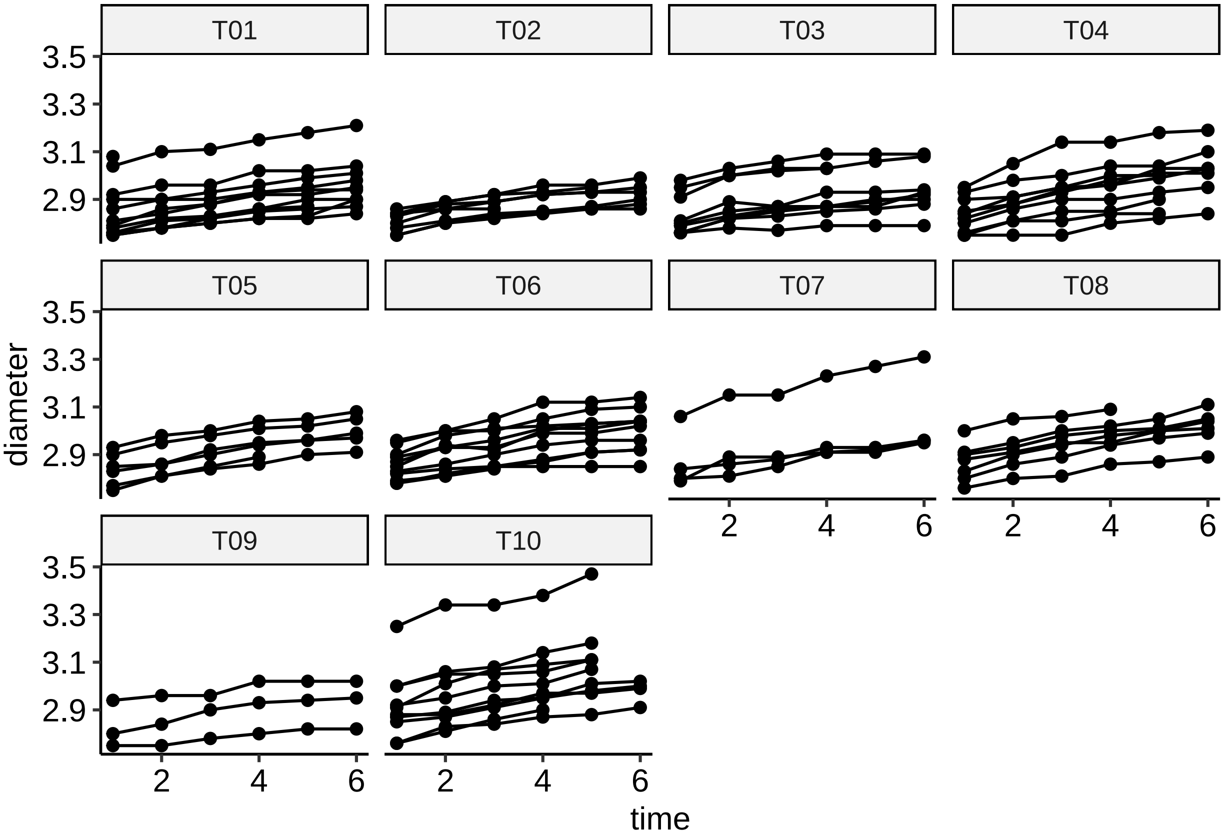
The data below were generated by an experiment that was run to study the growth of apples (in diameter). The experiment conducted at the Winchester Agricultural Experiment Station of Virginia Polytechnic Institute and State University. 25 apples were chosen from each of ten apple trees. The diameters of the apples were recorded every two weeks over a 12-week period.

5.1 What is the growth rate of the diameter for a single apple of unknown tree? Include a 95% confidence interval.

5.2 Compare the intra-tree and the inter-tree variability in growth rate.

```r
library(tidyverse)
library(ggpubr)

dat <- agridat::byers.apple
dat |>
  ggplot(aes(time, diameter))+
  geom_line(aes(group=appleid))+
  geom_point()+
  facet_wrap(~tree)+
  theme_pubr()
```


T01
T02
T03
T04
T05
T06
T07
T08
T09
T10
3.5
3.3
3.1
2.9
diameter
2
4
6
time