

MS COMPREHENSIVE EXAM

Stat 505-506 Take Home

August 21, 2012

Due: August 22, 2012 at ??

The Data:

<http://www.math.montana.edu/~jimrc/MScomp/milkProtein.csv>

I will also post this file on the web as

<http://www.math.montana.edu/~jimrc/MScomp/TakeHomeAug2012.html>

Here's what we know about these data:

Dairy cattle are said to “freshen” when they give birth to a calf and start milk production. The 78 cows in this experiment are all of one breed, and are quite similar. As they freshened (which occurs at random times for each cow), they were randomly placed in one of three experimental groups, meaning were assigned to a pasture which had barley, lupins, or both in it. The `Diet` variable tells us which pasture they were assigned. Milk protein content was recorded for each week, and the study ran 19 weeks. At this time, the pastures were grazed heavily enough that the experiment was ended.

The questions of interest is whether the treatment had an effect on mean protein and if the effect changed over time.

You will see that each cow's protein levels show a strong decrease in the first three weeks. The higher protein levels in these weeks are not related to diet, and these three weeks should be discarded to better address the Diet effect.

The `Time` variable in the data refers to time from when the cow entered the study, or “time since freshening”, therefore some cows were not observed for the full 19 weeks, since the cows do not all freshen at once. Because this view of time does not coincide with calendar time, I suggest that you reverse code time ($20 - \text{Time}$) so that all cows are on the same time, entering in different weeks and all stopping together.

Give a thorough analysis of the questions about effects of Diet (intercept and slope). Do these make sense using reverse time, or did I steer you wrong with that suggestion? Discuss any random effects you utilize (why did you choose them to be random? how large is the variation compared to underlying σ^2 ?) and any possible departures from the “usual assumptions”. Address, in particular, two aspects of the data:

1. The scientist has heard that missing data can cause problems with the analysis. What assumption is needed? Should we worry about missing data due to some cows starting later?
2. The pastures are not replicated, Only 3 pastures were available and they held 25, 26 or 27 cows each. How does that limit the inferences we wish to draw?

Your analysis should be summarized in a technical statistical report targeted toward a dairy scientist who is familiar with mixed effects and repeated measures models. Use Sweave or knitR and \LaTeX to produce the report and email a copy of the Sweave or knitR file to Jim Robison-Cox. Also turn in a hard copy to Josie by the appointed deadline. Include appropriate plots to explain your conclusions including diagnostic plots which consider the time series nature of these data.