Untitled

Gerard Castell i Joan Valls

Friday, March 20, 2015

Multivariate model

Despite the fact of not being the optimal choice, a multivariate linear model was fitted to the data. It seems reasonable to think that observations are not independent, since they are comming from the same cow. However, this model can also lead to some interesting results.

From the bivariate linear regressions models fitted in the previous section, both time and dose seem to have some kind of effect in the PCV. Following this rationale, the first multivariant model used to explain differences in PCV among observations, was the one with the additive effects of the variables time and dose. Since both were significant, two F-tests were performed right after, to test if the models adding the new additive effect of number of births or the second order interaction between time and dose, were significantly different with respect to the first one. The variable with the lowest p-value under 0.05 was added to the model. This procedure was followed until no other covariates or interactions below 3rd order could be added, because they would not produce a significantly better model. Note that with this procedure the most parsimonious model was ensured to be found. Table 1 summarizes the results of all models considered. Finally, the best model explaining the data was,

$$PCV = 15.82 + 2.43X_{time} - 0.32X_{doseM} - 1.31X_{doseH} - 0.34X_{nbirth} + 0.48X_{doseM}X_{time} + 2.35X_{doseH}X_{time}$$

Formula	Variables	Estimate	Std. Error	t value	Pr(>|t|)
pcv~time + dose	(Intercept)	12.32	0.83	14.78	<0.00001
	time	3.55	0.37	9.63	<0.00001
	doseM	0.56	0.75	0.75	0.45
	doseH	3.2	0.72	4.47	0.00003
pcv~time * dose	(Intercept)	14.51	1.29	11.28	<0.00001
	time	2.28	0.68	3.34	0.001
	doseM	-0.47	1.81	-0.26	0.79
	doseH	-1.45	1.73	-0.84	0.41
	time:doseM	0.65	0.93	0.7	0.48
	time:doseH	2.5	0.87	2.87	0.005
pcv~time * dose + nbirth	(Intercept)	15.85	1.35	11.74	<0.00001
	time	2.43	0.66	3.67	0.0005
	doseM	-0.32	1.75	-0.18	0.86
	doseH	-1.31	1.67	-0.78	0.44
	nbirth	-0.34	0.13	-2.51	0.01
	time:doseM	0.48	0.9	0.54	0.59
	time:doseH	2.35	0.84	2.8	0.007

Table 1: All multivariant models considered for fitting the data, using a forward stepwise for selecting the model.

No more variables were added because the resulting model was not significantly different from the chosen one and were less parimonious. The analyses of the residuals of the model can be seen in Figure 1. Clearly, they present some issues: residuals are not independent from PCV, are not centered for high dose, no

homoscedasticity, etc. So even thought this is not the optimal model to fit the data (and this is also shown in the residuals plots), the model is pointing in a pretty obvious way that the high dose is the only one that an effect significantly different from low dose. One of the problems with the data is that there are no controls among the cows so, the effect of the doses cannot be compared with PCV without the treatment. We are suposing that any of the treatments won't go worse than not giving a treatment, and that the higher the dose the bigger the effect. The interaction between time and dose is also pointing that the effect of the high dose is different in each time, even when adjusting for the number of births.

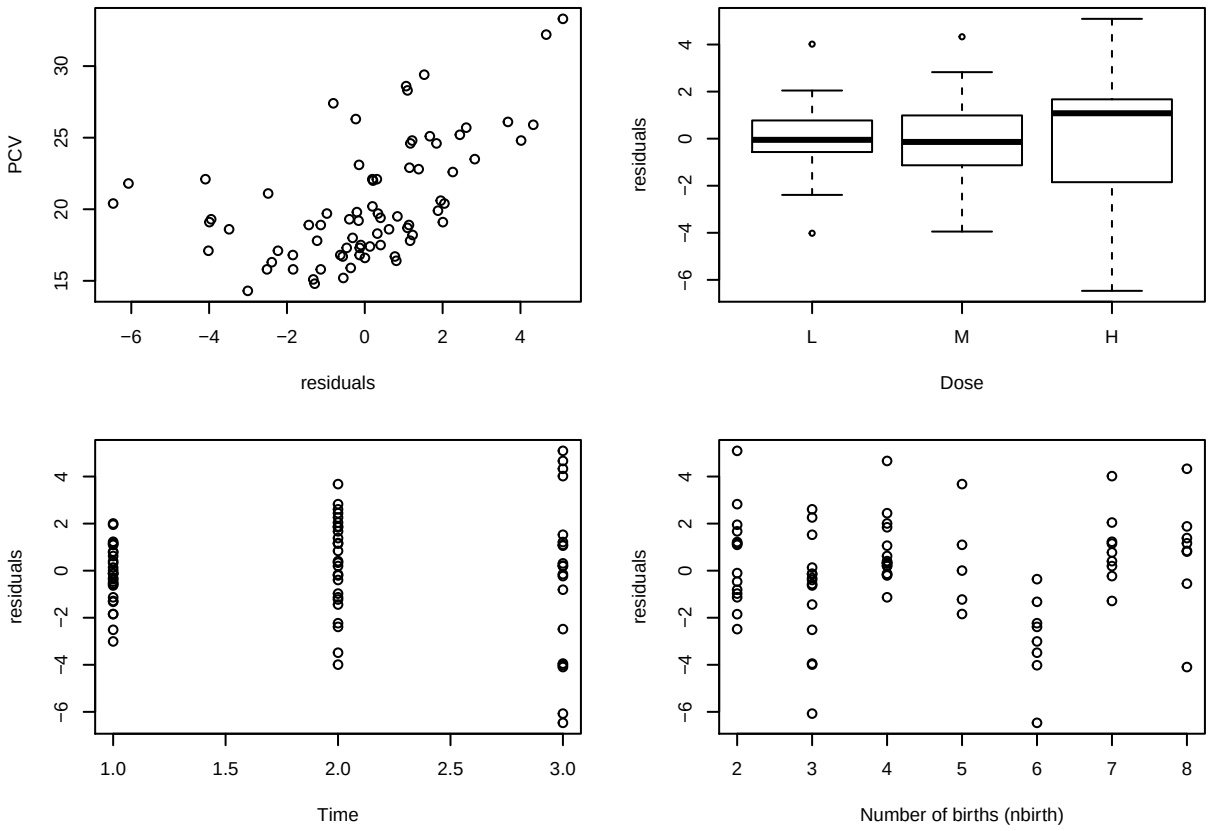

Figure 1: Analyses of the residuals of the model.