

Statistics 215B: Lab 3

UC Berkeley, Spring 2011

Hoxie Ackerman

March 31, 2011

1 Split-Level Alfalfa Cultivation

Early designers of experiments often used a one-factor-at-a-time approach to exploring the variable space: after optimizing with respect to one variable, they would fix that variable at the determined optimal level and explore others variables in the same manner. This design approach has a number of shortcomings, however: treatment combinations are often missed, interactions can't be estimated, randomization is impossible, and experiments often take longer, allowing temporal effects to creep in [1]. To ameliorate these issues, Fisher proposed fully randomized and crossed experiments, suggesting that Nature "will best respond to a logical and carefully thought out questionnaire" [2]. Indeed, randomized and fully-crossed experiments are more efficient due to the 'hidden' replications that occur with every observation, allow us to evaluate interactions, and often let us address more questions simultaneously.

Though the benefits of randomized, fully-crossed experiments are impressive, it's not always possible to vary the factor levels in the ways required for a randomized, fully-crossed experiment. For example, if changing a machine part is too labor or time intensive, experimenters might run through all other factor combinations in a random order with the machine at one level, then change the machine part and run all other factor combinations in a new random order. Thus, this experiment was performed as randomly as possible, given the constraint of only one machine part change. Said another way, the randomizations of all other factors were nested within the machine randomization. As another example, agriculturalists wishing to evaluate irrigation systems and fertilizer types might be unable to irrigate any unit smaller than an entire field but able to fertilize subsections of a single field easily. Thus, fertilizer randomization would take place within irrigation randomization (though these factors are actually crossed). In both of these examples, the experimenters are executing a split-plot design, since the entire experimental space is first split into main plots at one resolution, then those plots are split into sub-plots at a higher resolution. These types of design are common in agricultural experiments, and a good example of an agricultural split-plot design follows.

1.1 Exploratory Data Analysis

The data under consideration involve the impact of two factors on alfalfa yields in the following year. The two factors are alfalfa Variety and the Date of the third cutting. Three varieties of alfalfa were planted: Cossack (C), Ranger (R), and Ladak (L). Within each variety, harvests were performed at four different dates: no third cut (N), September 1 (S1), September 20 (S20), and October 7 (O7). These twelve combinations were repeated across eight Blocks, which were physically partitioned sections of the field to account for heterogeneity in the growing environment, giving us seventy-two observations and accompanying yields [3].

Examining the distributions of yields across cut dates and varieties by block (Figure 1), a few visual trends become apparent. Dates S1 and S20 show lower yields than the other two dates do, and variability across cut dates is relatively constant. No variety really behaves differently than the other two in terms of average yields, though Ranger variability is slightly smaller. By the fairly consistent vertical stacking of colors in both plots of Figure 1, it seems that some blocks enjoyed higher yields than other blocks did.

Also worth considering are interaction effects. There are three possible two-way interactions between factors, and a good way to initially understand them is via interaction plot, as given in Figure 2. Average yields were calculated for each factor level on the x-axis; these levels were then **sorted** from smallest to largest, and their trends across a second variable are plotted. The mostly parallel lines in Figures 2b and 2c suggest that interactions between (Block, Date) and (Date, Variety) are marginally significant at worst in these data. The crossing lines in Figure 2a suggest that there might be an interaction between Block and Variety: perhaps soil or other growing conditions differ between blocks, and different varieties thrive in different settings? This possible interaction is explored quantitatively in the next section.

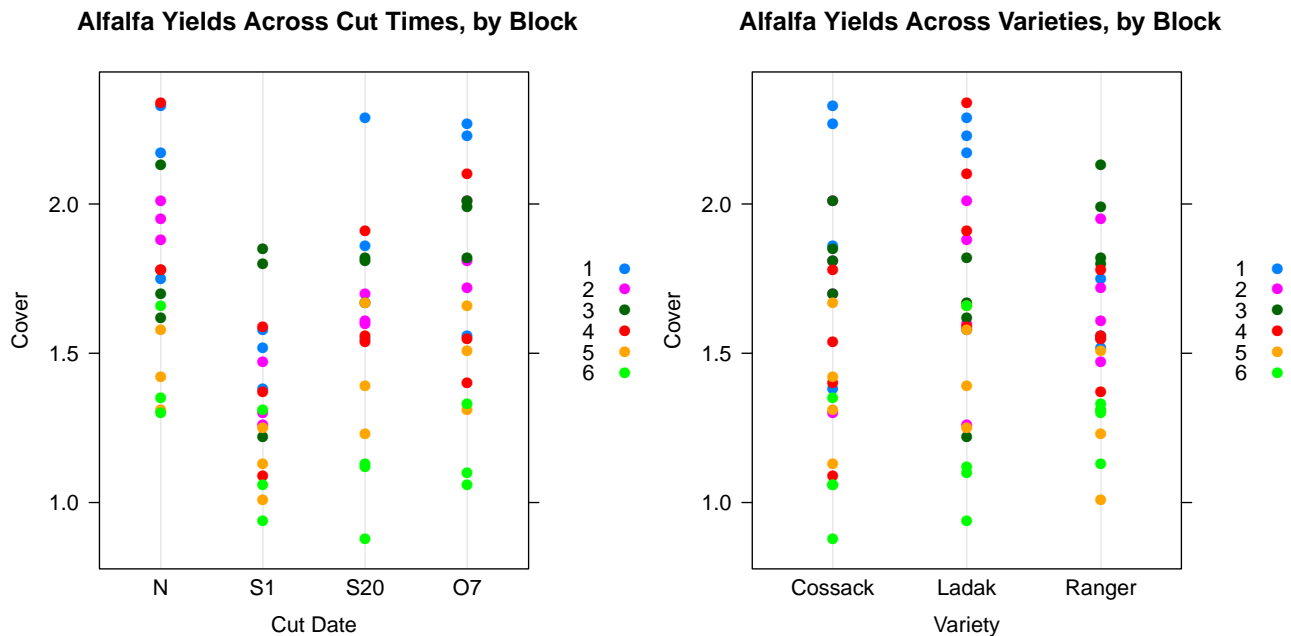


Figure 1: Distributions of yield across each factor, grouped by Block. (a) Dates S1 and S20 seem to have lower yields. Variability is relatively constant across times. (b) Though Ranger variability is slightly lower than the others, all three varieties have similar average yield. The fairly consistent stacks of colors in both plots suggest that blocks will explain a fair amount of variability in yield.

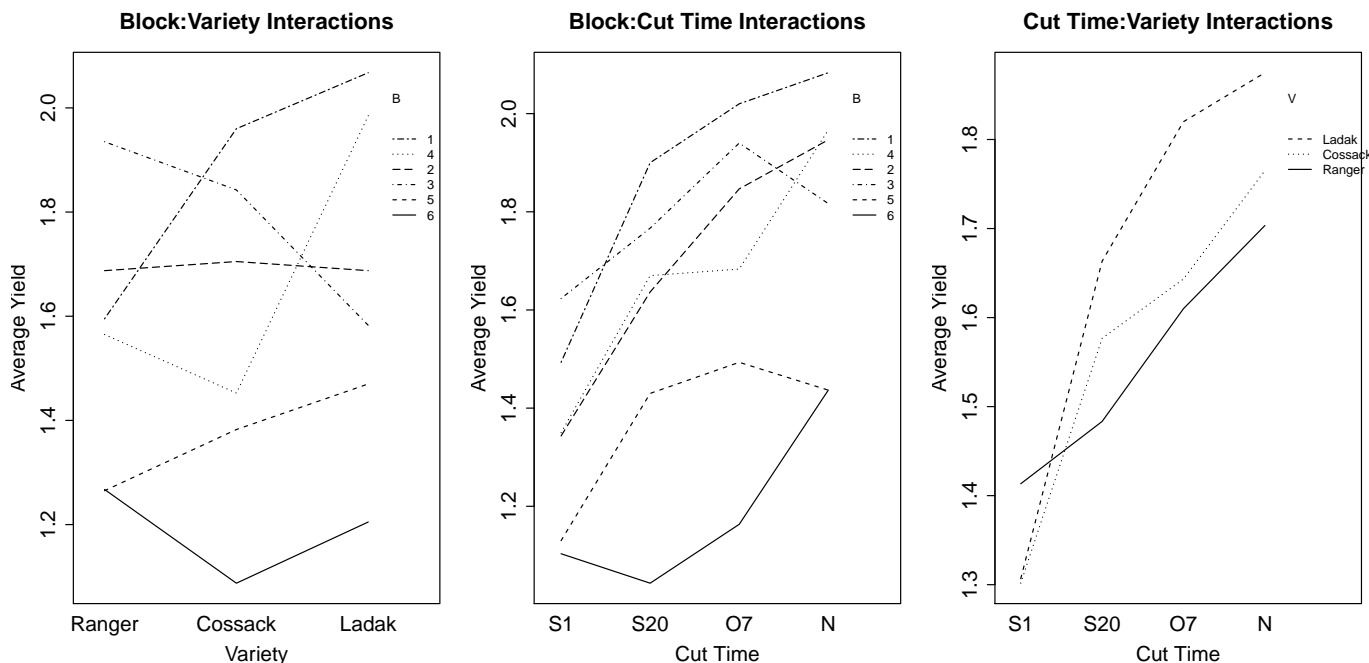


Figure 2: Interaction plots for the three pairwise interactions, with factor levels sorted from least to greatest. The mostly parallel lines in (b) and (c) suggest minimal interaction between (Block, Date) and (Date, Variety). The jagged, crossed lines in (a) suggest interactions between Block and Variety might be non-negligible.

1.2 Hasse Diagram

Hasse diagrams in statistics are figures that concisely summarize the relationships and properties of factors in an experiment. Though simple to learn how to construct and interpret, they can be quite powerful: images are always good ways to convey information, and Hasse diagrams specifically describe nesting of factors, their fixed/random status, and their degrees of freedom. They also provide guidelines for the calculation of sums of squares and concisely summarize potentially complicated experimental designs [4].

The Hasse diagram for the alfalfa experiment is given in Figure 3. Random effects have parentheses around them, and fixed effects do not. Superscripts give the number of levels for each factor, and subscripts give the degrees of freedom for each factor. The diagram is read from top to bottom, with the main effects nested within the mean, the interactions nested within the main effects, and the residuals nested within the interaction effects.

Figure 3: Hasse diagram for the alfalfa experiment. Parentheses indicate random effects. Superscripts represent number of levels, and subscripts represent degrees of freedom.

1.3 Analysis of Variance

ANOVA tables provide a framework for summarizing the breakdown of variability due to different factors in an analysis of variance setting. Using the Hasse diagram given previously, it's simple to break all factors into their various strata and understand the associated degrees of freedom. Doing so and calculating the sums of squares produces the full ANOVA table given in Table 1.

To better understand how the full ANOVA model decomposes total variability into the different randomness components, we can also make ANOVA tables containing only fixed and only random effects. The fixed effects are those under our control, with changing levels based on our assignments to different treatments. The random effects are those essentially out of our control but accounted for by inclusion in the design. Thus, total variability is a combination of the fixed effects, due to differences in treatments, and the random effects, due to differences between units. This is a balanced design, and we see that even when we break the full ANOVA model into its fixed and random components, the sums of squares quantities for each factor are preserved. Given that blocks explain more variability than the fixed effects do, it's fortunate that blocking was used in this experiment.

Stratum	Source	df	SS
Mean		1	183.58
B	Block	5	4.1498
BV	Variety	2	0.1780
	Resid	10	1.3623
	<i>(Total)</i>	12	1.5403
BD	Date	3	1.9625
	Resid	15	0.5501
	<i>(Total)</i>	18	2.5126
BDV (Error)	VD	6	0.2106
	Resid	30	0.7084
	<i>(Total)</i>	36	0.9190
Total		72	192.71

Table 1: Full alfalfa ANOVA table for Block (B), Variety (V), and Date (D), arranged by strata based on Figure 3.

Fixed				Random			
Stratum	Source	df	SS	Stratum	Source	df	SS
Mean		1	183.58	Mean		1	183.58
Error	Variety	2	0.1780	Block	Block	5	4.1498
	Date	3	1.9625				
	VD	6	0.2106				
	Resid	60	6.7707	Resid	Resid	66	4.9719
	<i>(Total)</i>	71	9.1218				
Total		72	192.71	Total		72	192.71

Table 2: ANOVA tables for fixed effects only and random effects only. The BD and BV interactions are not present because they involve the interaction of a fixed and random effect. However, we can still see how total variability is partitioned based on random variability and factor variability.

1.4 Choice of Design

This experiment was conducted to evaluate the effects of different alfalfa varieties and cut dates on total yield the following year. This was essentially a fully crossed design of Variety and Date, with sufficient replications via Block to obtain reasonable degrees of freedom. Though Table 1 suggests that we'll be better able to detect significant differences between Dates, due to the 15 degrees of freedom vs. just 10 degrees of freedom for Variety, this difference stems from the fact that there were more Date levels than there were Variety levels, not from something inherently special about the design favoring Date. In fact, this fully crossed design doesn't really favor the analysis of one factor over the other; all factors enter the Hasse diagram at the top level (Figure 3) and are analyzed accordingly.

Furthermore, faced with assigning a new treatment at either the plot or subplot level, I don't think that the distinction will make a statistical difference. Assigning at the *plot* level would mean that each block was divided into three plots, one for each variety, then these plots were divided into k subplots, where the new treatment has k different levels, and then each of these k subplots would be divided into 4 sub-subplots, one for each harvest date. Assigning at the *subplot* level would mean that each block was divided into three plots, one for each variety, then these plots were divided into 4 subplots, one for each harvest date, and then each of these 4 subplots would be divided into k sub-subplots, with k as above. In both cases, the new treatment is fully crossed with B, V, and D, so the Hasse diagrams for the two designs would be identical, resulting in the same residual comparisons and sums of squares values. From a practical perspective, it might be easier to physically create one configuration of fields over the other, but that's the only difference I see between the two setups in terms of testing effects statistically.

2 Optimizing Dog CT Scans

In some experimental design settings, the factors we work with are the exact factors that we're curious about. For example, we might want to understand how patients respond to specific dosages of a medicine, or how crop yields depend on which specific strain of alfalfa is planted and which specific date is used for a third harvest. These are all examples of fixed effects, where our concern lies with estimating and testing the significance of these specific factor levels.

However, it's often the case that some factors under consideration could be considered a sample from a larger population of all factor levels. For example, we might choose a subset of schools in which to try a new educational program, or we might select only a random sample of stores in a chain to try a new promotional campaign. In these examples, schools and stores, respectively, are random factors. Our goal is not to understand the effects of the specific schools or stores that we chose, but rather to understand variability in the population from which these factors were drawn. Blocks and test subjects are often sampled from a population and therefore considered random.

In this experiment, the random test subjects were dogs who were injected with a dye contrast used for computerized tomography (CT) scans. After injection, the dogs were scanned on a few occasions up to 21 days post-injection. The goal of this study was to determine the optimal amount of time to wait between injecting dogs and scanning them to see the strongest signal, as measured by image pixel intensity. Intensities in both left and right lymph nodes were recorded on all occasions. The data are a bit sporadic: not all dogs were observed after the same periods of time or even the same number of times. However, we should be able to estimate an effect curve by aggregating across all dog results, then locate the maximum of that curve to determine the best time delay between injection and scanning, as desired.

In terms of experimental design terminology, this is a repeated measures experiment with a longitudinal component to it: the same ten dogs were observed at multiple points in time post-injection. Time will obviously be an important factor in this experiment, and I believe it can be treated as a fixed effect: when we carry over this protocol to a new set of dogs, we'll still be measuring out the same intervals of time. The left/right lymph node component is a fixed effect, since these are the two levels that we'll always consider in future experiments. Since we'd like to generalize these results to all dogs and because the ten dogs in the experiment represent a sample from the population of all dogs, dogs will be a random factor.

2.1 Random Effects Model

To model pixel intensity based on the input variables we have, I created various visualizations to explore the data from different angles. Questions I was initially trying to answer visually included:

- What kind of overall (main effect) trends are there in the intensities?
- How do left and right intensities track and/or differ over time and between dogs?

Both of these questions are answered nicely by Figure 4. We see that the shapes of intensity distributions for both left and right lymph nodes are similar, starting low at Day 0, peaking somewhere between Days 7–12, and then declining over the rest of the observation period. Via the colors of the different lines, we see a strong alignment between Left/Right activity within the same dog. Across Left/Right, we see strong relationships within dog: dogs with larger values at earlier days tend to have larger values at later days in both nodes. I was careful not to use the word 'correlation' in the previous sentence, lest I imply a linear relationship: the fixed effect relationship between time and pixel intensity is at least second order, possibly higher.

Rather than starting with a large main effects model and paring it down as needed, I fit increasingly higher order (in Day) linear models to the data, both including and excluding an indicator variable for Side. In all models I fit, the coefficient for the Side indicator had a p-value around 0.3, implying that there is no significant difference in response levels between Left/Right in these data. Both the second-order and third-order models (without Side) fit the data well. Examining the regression coefficients (calculated but not shown), the cubed term in the third-order model had a p-value on the order of $1e-4$, while the fourth-order term in the fourth-order model was not significant, so I chose a third-order model without an Indicator variable for Side as my fixed effects model. If the goal of this experiment was to predict a future value as best

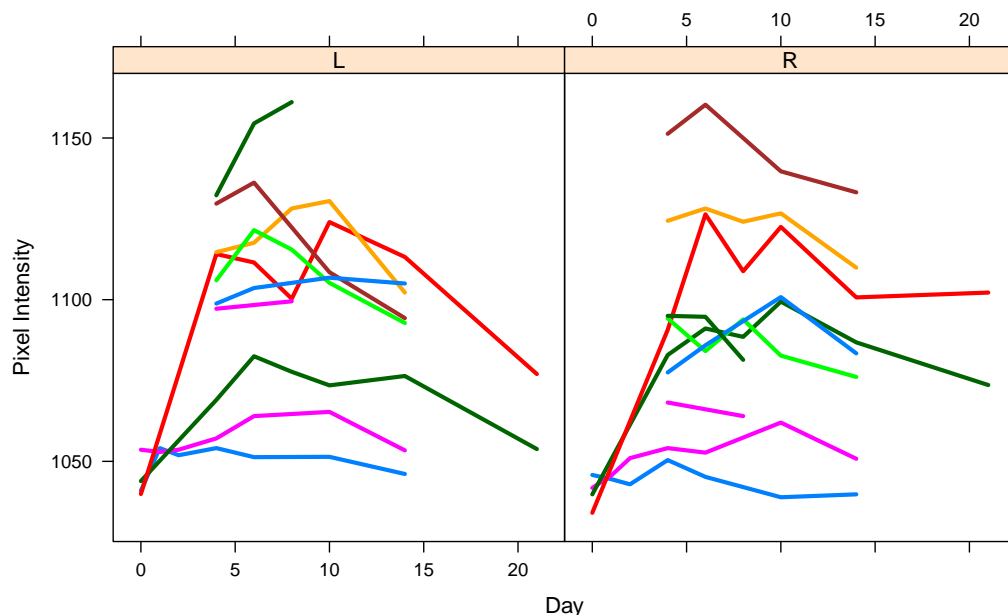


Figure 4: Pixel intensities over time in the left and right lymph nodes of 10 dogs. Features between Left and Right nodes don't appear to differ appreciably. Both sides show a low intensity within the first few days, a peak somewhere in the Days 7–12 range, and a decline across the remainder of the observation period.

as possible, it would probably be a good idea to include Side in the model; there *is* a small difference between sides (with Left values slightly higher than Right values for most of the experiment with a switch towards the end likely due to the small number of observations collected in later days) even if it's not statistically significant, and it only costs one degree of freedom to include a Side indicator. However, given that the goal of the experiment is to estimate the peak of the response curve, and that the response curves for both sides seem to peak at the same point in time, I dropped Side to simplify modeling.

To investigate needed random effects, I calculated and visualized the by-subject residuals resulting from the third-order model (Figure 5). The dotted black line in each panel is a loess smoothed summary line; though it doesn't follow the 0 line exactly, the subject residuals are centered mostly around zero. Note that the line tends to deviate most towards of the end of the experiment, when it's based on just two observations; when there are more data available, the line is closer to zero. The bump in both loess lines between Days 5–10 didn't disappear in fourth-order or fifth-order models. The fact that all subject residuals paths are not just noise around zero implies that random effects that provide control at the subject resolution will almost surely be helpful here. Furthermore, the residuals across time had clearly different intercept values, and there is definite linearity and even curvature in some of the residual paths. This suggested that a random intercept term, and perhaps random slopes and second-order terms too, might be appropriate.

After considering random effects models with a random intercept for each dog as well as first and second order random terms, the second-order random model seemed most appropriate based both on eyeballing the subject residuals in Figure 5 and based on AIC values from the various model fits (calculated but not given). To see how well this model fits our data, we can first look at the final residuals from the fit, as given in Figure 6. Though there do appear to be slightly more Left points above the loess smooth line in Figure 6a, we don't appear to be losing anything important by omitting Side from the model. Though there is a slight trend in the residuals at low fitted values, this trend quickly disappears. Some slight heteroskedasticity is present, though our impression of this feature is mostly fueled by the three extreme points on the plot around $x = 1020$, two above the line and one below. These three points are essentially the only outliers on the QQ plot given in Figure 6b; all in all, residual analysis suggests that this model fits the data well.

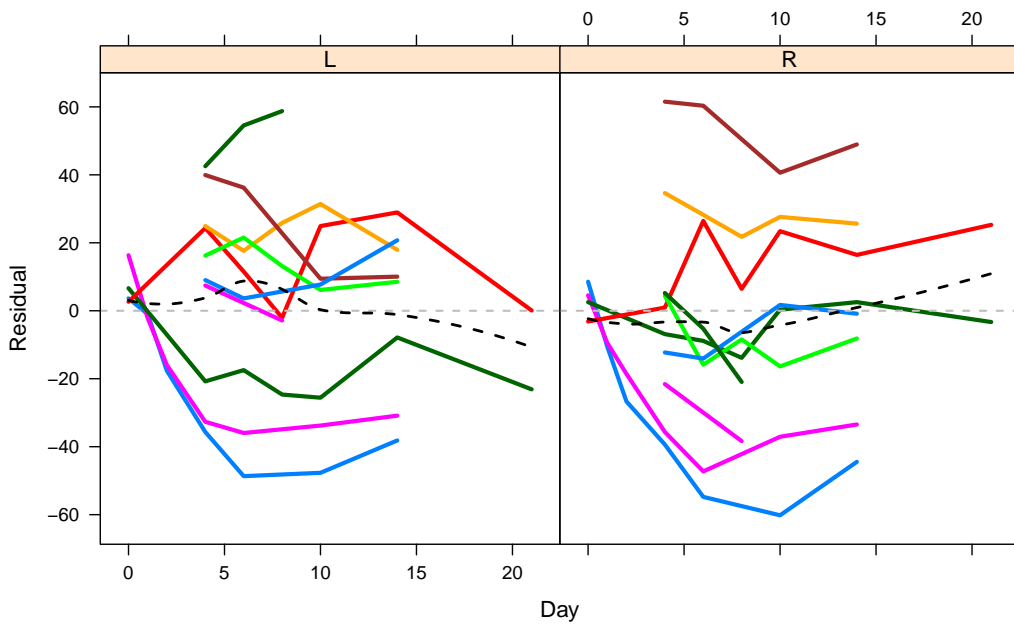


Figure 5: By-dog residuals from a third-order fixed effects model. The varying curve heights indicate a random intercept will be needed; random slopes for each dog might be useful too.

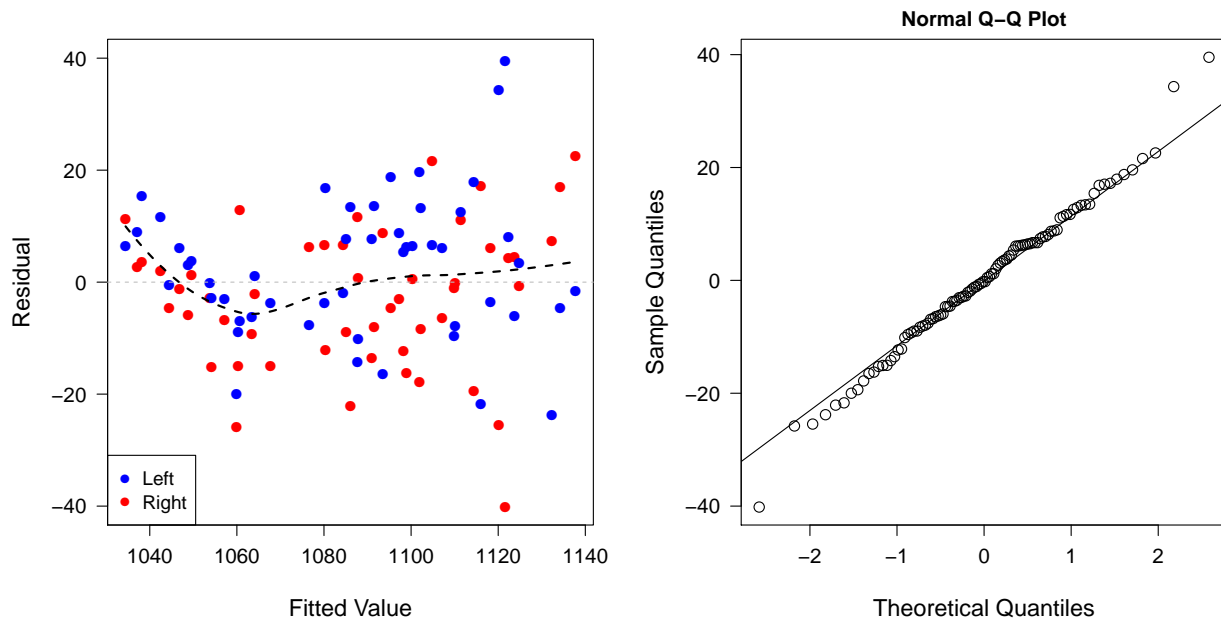


Figure 6: Residual diagnostics for a model with third-order Day fixed effects and second-order Dog random effects. (a) Over most of the range of fitted values, there is no trend in the residuals, and points from the Left and Right nodes are distributed approximately equally across the loess smoothing line. (b) Except for three outlying points, the residuals from this model are very close to Gaussian.

This model can be written formally as follows:

$$y_{ij} = \underbrace{\mu + \beta_1 \text{day}_i + \beta_2 \text{day}_i^2 + \beta_3 \text{day}_i^3}_{\text{fixed}} + \underbrace{\gamma_j^0 + \gamma_j^1 \text{day}_i + \gamma_j^2 \text{day}_i^2}_{\text{random}} + \epsilon_{ij} \quad (1)$$

where y_{ij} is the pixel intensity for dog j on day i , μ is a overall fixed effect common to all observations, β_1 , β_2 , and β_3 are the linear, quadratic, and cubic effects of time on intensity common to all dogs, γ_j^0 is the random intercept for dog j , γ_j^1 relates the effect of time on dog j individually, γ_j^2 is the second-order random effect of time on dog j , and ϵ_{ij} is a random error term specific to each observation. Aggregating up to the Dog level, we have:

$$y_j = X_j \beta + Z_j \gamma_j + \epsilon_j, \quad (2)$$

with $\beta = (\mu, \beta_1, \beta_2, \beta_3)^T$ a fixed, unknown vector common to all dogs, $\gamma_j = (\gamma_j^0, \gamma_j^1, \gamma_j^2)^T \sim \mathcal{N}(0, D_{3 \times 3})$ and $\epsilon_j \sim \mathcal{N}(0, R_{n_j \times n_j})$, where n_j is the number of observations we have for dog j . Thus, by the simple additive properties of the Normal distribution, we have:

$$y_j \sim \mathcal{N}(X_j \beta, Z_j D Z_j^T + \epsilon_j) \quad (3)$$

2.2 Estimating a Maximum

The goal of this experiment was to estimate the time, in days, at which the maximum pixel intensity is achieved. Just eyeballing the data, as given in Figure 4, the maximum seems to occur sometime in the Days 8–12 range. To pin this down further, we have a few options.

2.2.1 Option 1: Delta Method

We can take a parametric approach by fitting a quadratic model to the data and finding where the curve is maximized. The calculations for this procedure are given in Appendix A, but they boil down to basic calculus: take a partial derivative with respect to Day (so that in theory, whether or not random effects are included in the model is irrelevant), equate to zero, and solve. The resulting estimator for the maximum time point is a nonlinear function of our regression coefficients. However, since we know the asymptotic distribution of these coefficients and we've solved for the nonlinear function that we'll be using, we can use the Delta Method to find the variance of our estimator. Doing so (Appendix A) produces:

$$\widehat{\text{day}}^* = \frac{-\hat{\beta}_1}{2\hat{\beta}_2}, \quad \widehat{\text{Var}}\left(\frac{-\hat{\beta}_1}{2\hat{\beta}_2}\right) = \frac{1}{4\hat{\beta}_2^2} \left[\widehat{\text{Var}}(\hat{\beta}_1) - \frac{2\hat{\beta}_1}{\hat{\beta}_2} \widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) + \frac{\hat{\beta}_1^2}{\hat{\beta}_2^2} \widehat{\text{Var}}(\hat{\beta}_2) \right] \quad (4)$$

I fit the following quadratic model, which includes random terms, using the `lme4` package:

$$y_{ij} = \mu + \beta_1 \text{day}_i + \beta_2 \text{day}_i^2 + \gamma_j^0 + \gamma_j^1 \text{day}_i + \gamma_j^2 \text{day}_i^2 + \epsilon_{ij}, \quad (5)$$

Substituting in the calculated coefficient estimates, variances, and covariances into Equation (4) produced the following numerical values:

$$\widehat{\text{day}}^* = 8.86, \quad \widehat{\text{Var}}\left(\frac{-\hat{\beta}_1}{2\hat{\beta}_2}\right) = 0.904$$

The estimated coefficients, variances, and covariances will change with the fitting method. The `lmer` function uses restricted maximum likelihood (REML) methods by default, producing coefficients that should be less biased than MLE estimates. Working only with fixed effects, i.e. with the model given in Equation (5) without the γ terms, produces the following estimates:

$$\widehat{\text{day}}^* = 10.72, \quad \widehat{\text{Var}}\left(\frac{-\hat{\beta}_1}{2\hat{\beta}_2}\right) = 0.527$$

2.2.2 Option 2: Nonparametric Bootstrap

A key assumption of the delta method, as mentioned in Appendix A, is that $\hat{\beta}$ is asymptotically Normal with asymptotic mean β . Though it's often the case that estimators are normally distributed, we can relax these assumptions and consider an alternative method for estimating the variance of our estimate for the best day to perform the follow-up scan: the nonparametric bootstrap.

In the simplest nonparametric resampling scheme, sets of bootstrap observations are created by resampling the residuals and adding them to fixed effects estimates. This scheme won't take the correlation of the observations within the same dog into effect, though. I had initially planned to perform a block bootstrap by resampling residuals at the 'dog' level to avoid these correlation issues, since assuming that dogs are independent seems like a reasonable assumption to make, but the fact that different dogs were observed at different time points mean that different dogs will have different numbers of residuals corresponding to different time points.

Therefore, the only way I could see to reasonably bootstrap these data was to resample with replacement at the case level: choose bootstrap samples of ten dogs with replacement, implement a regression model on the resulting dataset, compute the maximizing ratio from the resulting fixed effect coefficients, and repeat over many resamplings. Like in Section 2.2.1, I considered two models: Equation (5) for the quadratic random effects model, and Equation (5) without the γ terms, i.e. just a quadratic fixed effects model.

I performed 10,000 separate resamplings for each model, and the distributions of best times can be seen in Figure 7. Both distributions are skewed left slightly, and it appears that the random effects model exhibits slightly larger variability than the fixed effects model does. As the median is a better estimate of center than the mean in skewed settings, I calculated median values of 8.49 and 10.57 for the random and fixed models, respectively. These values are close to the theoretical values of 8.86 and 10.72 calculated in Section 2.2.1, as we'd expect: these bootstrapped distributions should be good approximations to the true distributions of the statistics.

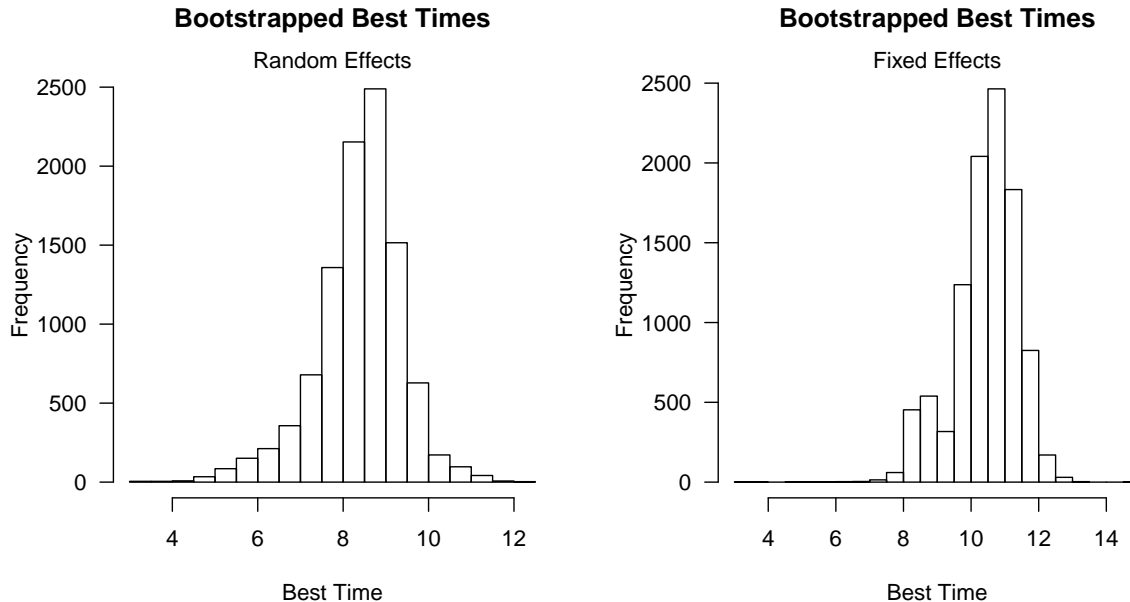


Figure 7: Results of 10,000 separate bootstrap estimates of the optimizing time in settings where Dog was treated as random (a) and fixed (b).

The variance values for the random effects and fixed effects models are given simply by the variance of the 10,000 bootstrapped optimal time estimates, and are 1.01 and 0.915, respectively. The random effects result was close to the asymptotic result calculated via the Delta Method (0.904), though the bootstrapped

fixed effects variance estimate was larger than the Delta Method would suggest (0.527). This can probably be explained by the fact that the variance calculated in the Delta Method section is an asymptotic result, and we only have 10 observations in this problem, not enough to expect asymptotics to provide a good approximation to our setting. I would be more inclined to believe the larger variance value, which probably better reflects the small sample size.

2.3 Considering A Robust Estimate

Producing an unbiased estimator of β in a random effects setting is relatively straightforward: as long as the mean component to y has been specified correctly, ignoring or misspecifying the variance structure of the problem won't bias the estimator itself (though it will affect the variance). As an example, given a random effects model $y \sim N(X\beta, ??)$, where the variance of y is left intentionally unspecified, we can consider two estimators of β : $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$ and $\hat{\beta}_{RE} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$. Even though one estimator takes the variance structure into account and the other ignores it completely, both estimators are still unbiased for β :

$$\begin{aligned}\mathbb{E}[\hat{\beta}_{OLS}] &= (X^T X)^{-1} X^T \mathbb{E}[y] \\ &= (X^T X)^{-1} (X^T X) \beta = \beta \\ \mathbb{E}[\hat{\beta}_{RE}] &= (X^T V^{-1} X)^{-1} X^T V^{-1} \mathbb{E}[y] \\ &= (X^T V^{-1} X)^{-1} (X^T V^{-1} X) \beta = \beta\end{aligned}$$

Of course, to perform inference about our coefficient estimates, we need to take the variance of our estimators into account. Regardless of the true variance structure of y , we can use standard variance formulas to obtain:

$$\begin{aligned}\text{Var}[\hat{\beta}_{OLS}] &= (X^T X)^{-1} X^T \text{Var}[y] X (X^T X)^{-1} \\ \text{Var}(\hat{\beta}_{RE}) &= (X^T V^{-1} X)^{-1} X^T V^{-1} \text{Var}[y] V^{-1} X (X^T V^{-1} X)^{-1}\end{aligned}\tag{6}$$

If we're willing to assume a variance structure for y , say $\text{Var}[y] = \sigma^2 I$ as in OLS or $\text{Var}[y] = Z D Z^T + R = V$ as in a typical random effects model, then these variance formulas reduce nicely to $\text{Var}[\hat{\beta}_{OLS}] = \sigma^2 (X^T X)^{-1}$ and $\text{Var}[\hat{\beta}_{RE}] = (X^T V^{-1} X)^{-1}$, respectively.

Of course, this nice reduction of the expressions in Equation (6) depends on the correct specification of the variance of y : all appropriate higher order terms have been included, no extraneous effects were included, etc. Rather than assume a possibly erroneous variance structure to calculate $\text{Var}(\hat{\beta}_{RE})$, which would in turn mean our significance tests about fixed effect regression coefficients would be incorrect, we can instead try to estimate $\text{Var}[y]$ directly, then substitute this value into Equation (6). An estimate of $\text{Var}[y]$ often used is based on the residuals from an OLS regression:

$$\widehat{\text{Var}}[y] = n^{-1} \sum_{i=1}^n (y_i - X_i \hat{\beta}_{OLS})(y_i - X_i \hat{\beta}_{OLS})^T$$

This is referred to as a robust estimator of $\text{Var}(\hat{\beta}_{RE})$, robust in the sense that it doesn't depend on a possibly incorrect specification of variance. The pros of this estimator are that it is a consistent estimator, as long as the mean has been correctly specified, and if we're only interested in testing our regression coefficients, we don't need to spend time and computer cycles estimating the variance structure. The cons of this estimator are that it's a less efficient estimator than what would be obtained if the variance were specified correctly, and more importantly, no estimate of the covariance structure is obtained using this method [5].

To determine whether or not the robust estimator will be useful in our setting, we need to keep the goal of the experiment in mind: predicting the best time for a dog CT scan. Though I initially dismissed the need for a correct variance calculation, thinking that a point estimate would suffice, I realized that it's quite unlikely that the dog will be in the operating room exactly 8.86 days after the radioactive injection, and therefore that having a good idea of how much variability there is in this estimate could be quite useful. Given the

doubts I had while specifying which model to use, it might be nice not to need to rely on a correct model assumption to obtain an estimate of variability, especially since we only have 10 independent observations in this experiment. I was unable to implement the robust estimate calculation due to confusion on my part about which estimate of V_i^{-1} to use, so this is a though experiment at best, but there are definitely obvious benefits to the robust estimator in a small-sample setting.

A Calculations Related to Finding a Maximum

The random effects model with second order time effect is:

$$y_{ij} = \mu + \beta_1 \text{day}_i + \beta_2 \text{day}_i^2 + \gamma_j^0 + \gamma_j^1 \text{day}_i + \gamma_j^2 \text{day}_i^2 + \epsilon_{ij}.$$

Calculus tells us that to find the maximum value of day can be found by taking the partial derivative with respect to day and equation to zero:

$$\begin{aligned} \beta_1 + 2\beta_2 \text{day}^* &= 0 \\ \Leftrightarrow \text{day}^* &= \frac{-\beta_1}{2\beta_2} \end{aligned}$$

This is a nonlinear function g of true value $\beta^T = (\beta_0, \beta_1, \beta_2)$. The best point estimate of this quantity will be

$$\widehat{\text{day}}^* = g(\hat{\beta}) = \frac{-\hat{\beta}_1}{2\hat{\beta}_2},$$

and we can use the delta method to estimate $\text{Var}(g(\hat{\beta}))$. Assume that $\hat{\beta} \sim N(\beta, \text{Var}(\hat{\beta}))$ asymptotically. This suggests that $\hat{\beta}$ will be close to β , so we can use a first-order Taylor expansion to obtain:

$$g(\hat{\beta}) \approx g(\beta) + \dot{g}(\beta)^T (\hat{\beta} - \beta), \quad \text{where } \dot{g}(\beta)^T = \left(0, \frac{-1}{2\beta_2}, \frac{\beta_1}{2\beta_2^2}\right).$$

Thus,

$$\begin{aligned} \text{Var}(g(\hat{\beta})) &\approx \text{Var} \left[g(\beta) + \dot{g}(\beta)^T (\hat{\beta} - \beta) \right] \\ &= \dot{g}(\beta)^T \text{Var}(\hat{\beta}) \dot{g}(\beta) \\ &= \left(0, \frac{-1}{2\beta_2}, \frac{\beta_1}{2\beta_2^2}\right) \text{Var}(\hat{\beta})_{3 \times 3} \left(0, \frac{-1}{2\beta_2}, \frac{\beta_1}{2\beta_2^2}\right)^T \\ &= \frac{1}{4\hat{\beta}_2^2} \left[\text{Var}(\hat{\beta}_1) - \frac{2\hat{\beta}_1}{\hat{\beta}_2} \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) + \frac{\hat{\beta}_1^2}{\hat{\beta}_2^2} \text{Var}(\hat{\beta}_2) \right]. \end{aligned}$$

$\widehat{\text{Var}}(g(\hat{\beta}))$, our numerical value for the variance of the day at which the maximum intensity value is obtained, can be obtained by substituting in $\widehat{\text{Var}}$ and $\widehat{\text{Cov}}$ values, as obtained by regression output. A similar derivation is given in [6].

References

- [1] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li. *Applied Linear Statistical Models*. McGraw-Hill Irwin, 5th edition, 2005.
- [2] R. A. Fisher and J. Wishart. The arrangement of field experiments and the statistical reduction of the results. *Imperial Bureau of Soil Science and Technology*, 10(23), 1930.
- [3] G. W. Snedecor and W. G. Cochran. *Statistical Methods*. The Iowa State University Press, 7th edition, 1980.
- [4] S. L. Lohr. Hasse Diagrams in Statistical Computing and Teaching. *The American Statistician*, 49(9):376–381, 1995.
- [5] G. Molenberghs and G. Verbeke. *Linear Mixed Models for Longitudinal Data*. Springer New York, 2000.
- [6] S. Weisberg. *Applied Linear Regression*. Wiley, 3rd edition, 2005.