Calf Weight Gain Data

Arden Miller

Executive Summary

The average daily weight gains (ADG) of calves during 3 to 9 months of age for three different breeds were compared. It was found that the ADG of calves is affected both by the ADG the calf's sire and by the ADG of the calf's dam. The impact of the dam's ADG was similar across breeds but the impact of the sire's ADG was clearly different for different breeds. For each increase of 1 unit in the dam's ADG, the predicted ADG of the calf increases by 0.512 units. For each increase of 1 unit in the sire's ADG, the predicted ADG of the calf increases by 0.643 units for breed 1, by 0.151 units for breed 2 and by 0.207 units for breed 3.

If we compare calves across breeds where the sire and dam of each calf have the average values of ADG (for that breed), then the predicted ADG is 2.99 for breed 1, 2.51 for breed 2 and 4.17 for breed 3.

Calf Data

This report considers data that was collected to explore the growth rate of calves. The average daily weight gain (ADG) during 3 to 9 months of age for 3 breed of calves was studied. It was thought that ADG is partly an inherited trait. Thus the ADG of each calf's sire and ADG of each calf's dam (when they were growing during 3 to 9 months of age) were considered as possible covariates. The data consists of the following measurements:

breed:	breed of calf, a factor with levels 1, 2 or 3.
adg:	average daily weight gain of calf.
sadg:	average daily weight gain of calf's sire.
dadg:	average daily weight gain of calf's dam.

The aim was to identify a suitable regression model for adg using the remaining variables as regressors. This model is then to be used to compare weight gains for different breeds. Figure 1 contains the pairwise scatter plots of the data. From these plots it is evident that adg tends to be highest for breed 3 and lowest for breed 2. There is also clear evidence that adg is related (in a positive manner) to both sadg and dadg.

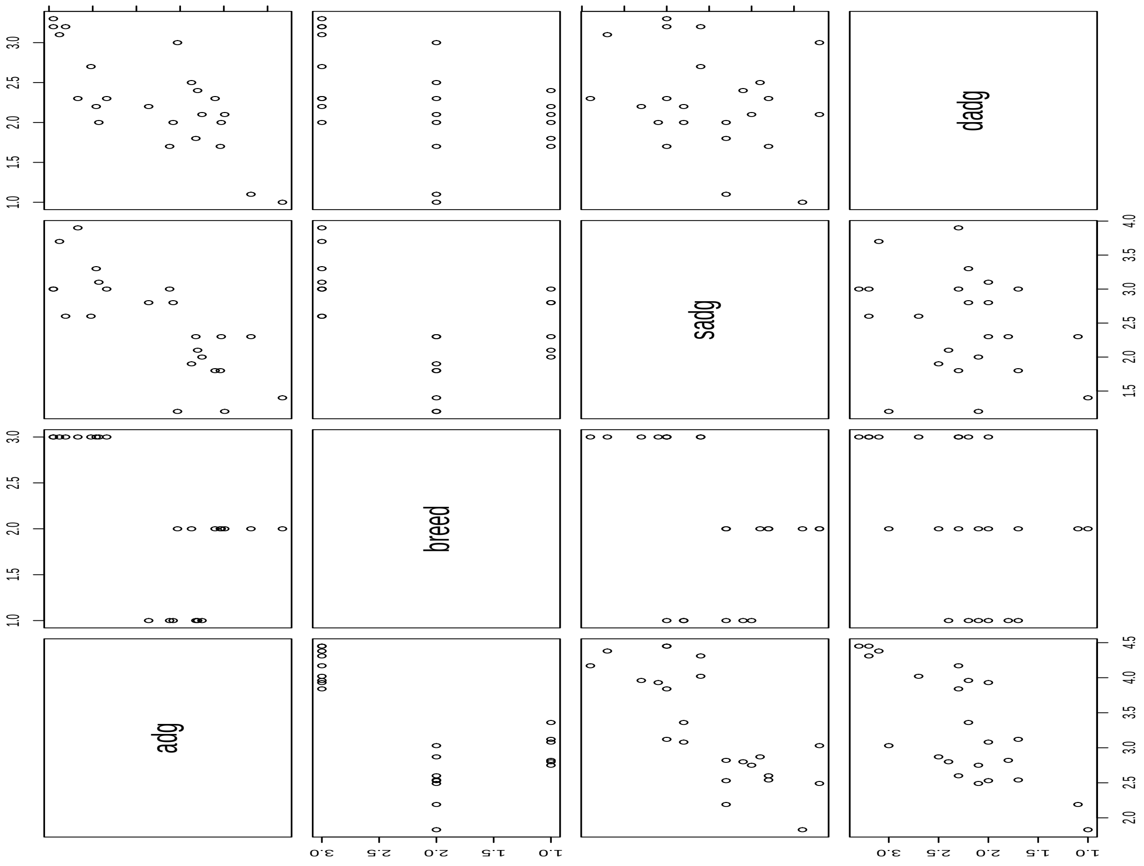


Figure 1: Pairwise Scatterplots of Calf Data.

1 Modelling "Average Daily Gain" of Calves

My analysis of the calf data indicated that both `sadg` and `dadg` are important regressors for predicting `adg`. Further, it was found that `sadg` interacted with `breed` whereas `dadg` did not. What this means is that the way `sadg` affects `adg` is different for different breeds but the impact of `dadg` on `adg` is (approximately) the same across breeds.

The fitted regression model produces the following expressions that can be used to predict the expected `adg` for calves of each breed based on `sadg` and `dadg`:

$$\begin{aligned} \text{breed 1:}\quad \mathrm{E}(\texttt{adg}) &= 0.339 + .643 \times \texttt{sadg} + .512 \times \texttt{dadg} \\ \text{breed 2:}\quad \mathrm{E}(\texttt{adg}) &= 1.242 + .151 \times \texttt{sadg} + .512 \times \texttt{dadg} \\ \text{breed 3:}\quad \mathrm{E}(\texttt{adg}) &= 2.137 + .207 \times \texttt{sadg} + .512 \times \texttt{dadg} \end{aligned}$$

For each breed the coefficient for `dadg` is 0.512. This means that if `sadg` is fixed then the estimated value of `adg` increases by 0.512 for each unit increase in `dadg` for each breed. In a similar manner the coefficients for `sadg` can be interpreted as the increase in `adg` for each unit increase in `sadg` if `dadg` is held constant. Note that the coefficients for `sadg` are different for each breed. Thus `sadg` has a the biggest impact on the predicted value of `adg` for Breed 1 (.643) and the smallest impact for Breed 2 (.151).

Comparing weight gains between breeds is a bit tricky. We need to take into account the levels of `sadg` and `dadg`. However, fixing the levels of these two variables at common levels for all three breeds is not reasonable since the ranges of values for `sadg` and `dadg` are quite different for the different breeds. For example, the range of `sadg` is from 1.2 to 2.3 for breed 2 and from 2.6 to 3.9 for breed 3. As an alternative, predictions were made using three scenarios for each breed: (i) a worst case scenario where `dadg` and `sadg` are set to their minimum values for that breed, (ii) an average scenario where `dadg` and `sadg` are set to their average values for that breed, and a best case scenario where `dadg` and `sadg` are set to their maximum values for that breed. Table 1 contains the predicted values for the average `adg` for each of these situations as along with a 95% confidence interval (an interval that we can be 95% confident will contain the true average `adg`). From this table it is clear that weight gains are highest for breed 3 and lowest for breed 2.

Breed		`sadg`	`dadg`	$\widehat{\texttt{adg}}$	95% CI
	min	2.00	1.70	2.50	(2.35, 2.64)
1	ave	2.50	2.03	2.99	(2.90, 3.07)
	max	3.00	2.40	3.50	(3.35, 3.64)
	min	1.20	1.00	1.94	(1.77, 2.10)
2	ave	1.74	1.96	2.51	(2.44, 2.58)
	max	2.30	3.00	3.13	(2.95, 3.30)
	min	2.60	2.00	3.70	(3.56, 3.84)
3	ave	3.13	2.70	4.17	(4.10, 4.24)
	max	3.90	3.30	4.63	(4.47, 4.80)

Table 1: Predictions of average `adg`

Statistical Appendix

I started by fitting the regression model that contained all three two-factor interactions in addition to the main effects. The output from the `anova` command suggested that it was not necessary to retain all three two-factor interactions. I used the added variable F-test to try dropping the two-factor interactions one at a time. The test for dropping `dadg:breed` gave the highest P-value (0.73) and so I eliminated this term. I then tried eliminating each of the two remaining interactions and found that `sadg:dadg` could also be dropped (P-value= 0.14). This leaves a model that contains the three main effects and the `sadg:breed` interaction. An F-test that considers eliminating `sadg:breed` give a very small P-value (0.004) which means that `sadg`, `breed`, and `sadg:breed` should all be kept. I then tried eliminating the `dadg` term but this gives an extremely small P-value. Thus I decided to use this model.

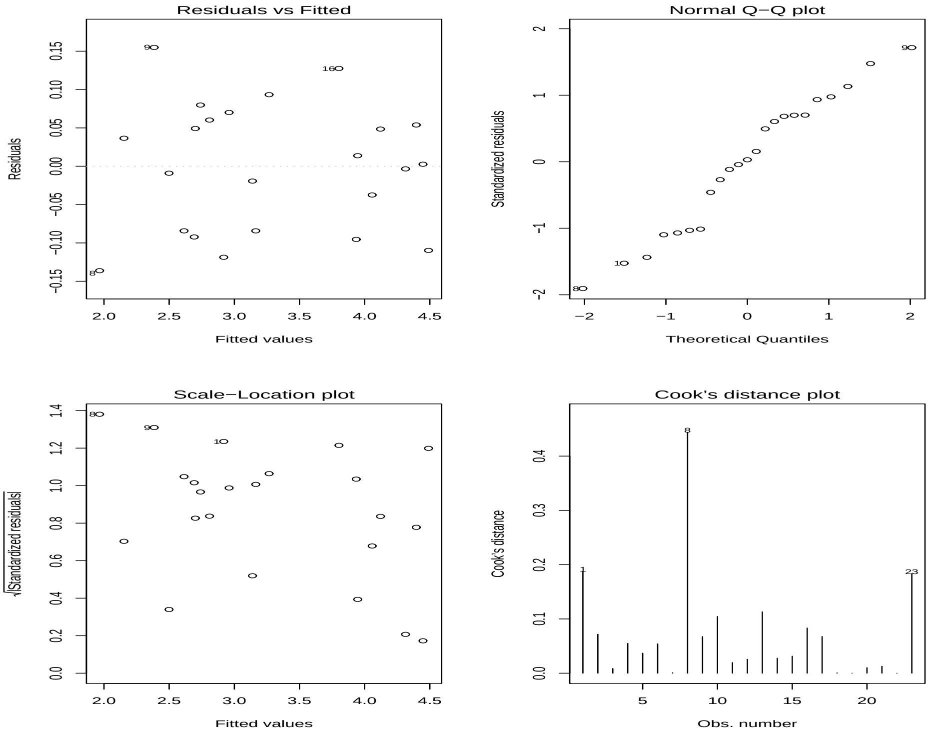
The output from `summary` and `dummy.coef` for this model is:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.33907    0.28834   1.176 0.256809
sadg         0.64331    0.10450   6.156 1.38e-05 ***
breed2       0.90342    0.30171   2.994 0.008580 **
breed3       1.79835    0.36201   4.968 0.000140 ***
dadg         0.51196    0.04318  11.855 2.45e-09 ***
sadg:breed2 -0.49207    0.13342  -3.688 0.001992 **
sadg:breed3 -0.43648    0.12971  -3.365 0.003941 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.09729 on 16 degrees of freedom
Multiple R-Squared: 0.9892, Adjusted R-squared: 0.9852
F-statistic: 244.9 on 6 and 16 DF,  p-value: 7.994e-15

> dummy.coef(calf.fit3)
Full coefficients are

(Intercept):      0.339073
sadg:            0.6433073
breed:                   1         2         3
                 0.0000000 0.9034225 1.7983496
dadg:            0.5119633
sadg:breed:              1         2         3
                 0.0000000 -0.4920690 -0.4364815
```

Note that almost 99% of the variability in the response is explained by this model. Diagnostic plots for this model are:



There doesn't seem to be clear evidence of either curvature or a funnel effect in the plot of residuals versus fitted values and the Normal probability plot follows a reasonably straight line. About the only cause for concern is that observation 8 has a somewhat high value of Cook's Distance. I tried deleting this point and re-fitting the model to see what impact this would have on the fitted model. The output from `dummy.coef` for the fitted model with observation 8 deleted is:

```
> dummy.coef(calf.fit3A)
Full coefficients are

(Intercept):     0.4951132
sadg:            0.6266756
breed:                   1          2          3
                  0.000000   1.047317   1.850831
dadg:            0.4556711
sadg:breed:              1          2          3
                 0.0000000 -0.5663055 -0.4378921
```

Although, the fitted coefficients have changed somewhat there is not drastic changed in what the model is indicating about differences between the breeds. Thus I did not discuss the effect of dropping this observation in my report.