

# Homework 3

Gitanshu Munjal

October 28, 2014

## 1 Model Development

### 1.1

Write a model to describe the  $\ln wt$  of filaree plants as a function of plant age (days) and temperature as a continuous variable. Assume that temperature treatments were 10, 15 and 20 C for treatments 1-3. Make sure you incorporate terms that will allow you to estimate RGR for different temperatures, even if the temperatures were not in the treatments. Interpret the biological meaning of the symbols in the model. [10]

$$\ln wt = \beta_0 + \beta_1 * \text{days} + \beta_2 * \text{temperature} + \beta_{12} * \text{days} * \text{temperature} + \epsilon \quad (1)$$

The above model can be rearranged and reduced to correspond to an intuitive slope-intercept form, for any given temperature  $T$ , as:

$$\ln wt = (\beta_0 + \beta_2 * T) + (\beta_1 + \beta_{12} * T) * \text{days} + \epsilon \quad (2)$$

Where,

$\ln wt$  is the natural log of weight of filaree plants

$\beta_0$  is the effect of initial weight

$\beta_1$  is the effect of days (time)

$\beta_2$  is the effect of temperature

$\beta_{12}$  is the effect of the temperature\*days interaction

$T$  is the test temperature (a continuous variable)

$\text{days}$  is number of days elapsed since application of treatments (plant age, continuous variable)

$\epsilon$  is a random variable with mean 0, variance  $\sigma^2$ .

Then, the slope and intercept here are linear functions of temperature. It is possible that RGR first increases and then decreases with increasing temperature; so we consider including quadratic temperature terms that affect the coefficients in our model. Logically, initial weight of plants before the temperature treatments are applied should not be treatment dependent and so we would only need to include a quadratic term for the coefficient of  $\text{days}$

$$\ln wt = (\beta_0 + \beta_2 * T) + (\beta_1 + \beta_{12} * T + \beta_{13} * T^2) * \text{days} + \epsilon \quad (3)$$

$$\ln wt = \beta_0 + \beta_1 * \text{days} + \beta_2 * T + \beta_{12} * \text{days} * T + \beta_{13} * \text{days} * T^2 + \epsilon \quad (4)$$

### 1.2

Run a linear model in R to get estimated parameters for your model. Use the principle of the extra sum of squares to determine if it is necessary to have the quadratic effect of temperature in the model. [10]

#### 1.2.1 Code and Relevant Output

```
library(car)
filaree<-read.csv("C:\\Users\\uglysweaters\\Downloads\\xmpl_filaree.csv", header=TRUE)
options(contrasts =c("contr.sum", "contr.poly"))
filaree$Tgroup<-factor(filaree$Tgroup)

filaree$t <- NA
filaree$t[filaree$Tgroup==1] <- 10
filaree$t[filaree$Tgroup==2] <- 15
filaree$t[filaree$Tgroup==3] <- 20

FullModel<-lm(lnwt ~ day + t + day:t + day:I(t^2), filaree)
```

```
summary(FullModel)
```

```

Call:
lm(formula = lnwt ~ day + t + day:t + day:I(t^2), data = filaree)

Residuals:
    Min       1Q   Median       3Q      Max
-0.32633 -0.12405 -0.02737  0.11139  0.34032

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.652e+00  2.439e-01   6.772  3.9e-08 ***
day          4.707e-02  2.123e-02   2.218  0.03232 *
t           5.177e-02  1.547e-02   3.348  0.00178 **
day:t       5.456e-03  2.781e-03   1.962  0.05673 .
day:I(t^2)  -1.646e-04  9.064e-05  -1.816  0.07684 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1773 on 40 degrees of freedom
Multiple R-squared:  0.9683,    Adjusted R-squared:  0.9652
F-statistic: 305.8 on 4 and 40 DF,  p-value: < 2.2e-16

```

```
Anova(FullModel,type=3)
```

Anova Table (Type III tests)

```

Response: lnwt
            Sum Sq Df F value    Pr(>F)
(Intercept) 1.44102  1 45.8628 3.898e-08 ***
day          0.15454  1  4.9184  0.032315 *
t           0.35214  1 11.2075  0.001783 **
day:t       0.12096  1  3.8499  0.056734 .
day:I(t^2)  0.10364  1  3.2985  0.07684 .
Residuals   1.25681 40
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The type III sum of squares tests each term as if it were entered last. The result of the above type-III tests, [non-significant](#) for quadratic term, indicates that it is not necessary to include a quadratic term for temperature affecting the slope with respect to days and so our full model can be reduced further.

### 1.3

Create a model with only those terms that you think are necessary and obtain estimated parameters. Report the formula, summary, R2, PRESS and diagnostic plots. [10]

Based on non-significant findings in the section above, our conceptual full model can be reduced to:

$$\ln wt = (\beta_0 + \beta_2 * T) + (\beta_1 + \beta_{12} * T) * days + \epsilon$$

#### 1.3.1 Code and Relevant Output

```
ReducedModel <- lm(lnwt ~ day + t + day:t, filaree)
summary(ReducedModel)
```

Call:

```
lm(formula = lnwt ~ day + t + day:t, data = filaree)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.33446	-0.12462	-0.04405	0.14016	0.40251

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.6557959	0.2506489	6.606	5.95e-08 ***
day	0.0811292	0.0102187	7.939	8.04e-10 ***
t	0.0514323	0.0158915	3.236	0.0024 **
day:t	0.0005393	0.0006527	0.826	0.4134

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1822 on 41 degrees of freedom

Multiple R-squared: 0.9657, Adjusted R-squared: 0.9632

F-statistic: 385 on 3 and 41 DF, p-value: < 2.2e-16

```
PRESS.statistic <- sum( (resid(ReducedModel)/(1-hatvalues(ReducedModel)))^2 )
PRESS.statistic
```

```
[1] 1.661572
```

```
library(car)
outlierTest(ReducedModel) # Bonferonni p-value for most extreme obs
```

No Studentized residuals with Bonferonni p < 0.05

Largest |rstudent|:

	rstudent	unadjusted	p-value	Bonferonni	p
22	2.35546		0.023491		NA

```
qqPlot(ReducedModel, main="QQPlot") #qq plot for studentized resid
```

The Q-Q plot demonstrates that the residuals from the data are not normally distributed, assuming the model is correct, as more than a few points from the double s-shaped pattern of the plot lie outside the confidence intervals. The alternative explanation for the pattern observed here could be an incorrect model. To this end, we look at the residual plots to make further comments.

```
residualPlots(ReducedModel)
```

	Test stat	Pr(> t )
day	-7.376	0.000
t	-2.036	0.048
Tukey test	-7.884	0.000

The Diagnostic residual plots and supporting tests presented above indicate that there is indeed a significant quadratic response to *days* that needs to be included in the model in order to meet normality of residuals.

```
influenceIndexPlot(ReducedModel)
```

## 2 Model Diagnostics

### 2.1

Interpret the diagnostic plots and report any problems or potential problems. [10]

- See Page 4

### 2.2

Perform a test of Lack of Fit and interpret the results. [20]

```
filaree$dayg <- NA
filaree$dayg[filaree$day<10] <- "A"
filaree$dayg[filaree$day>10&filaree$day<20] <- "B"
filaree$dayg[filaree$day>20&filaree$day<25] <- "C"
filaree$dayg[filaree$day>25&filaree$day<35] <- "D"
filaree$dayg[filaree$day>35] <- "E"

FullestModel <- lm(lnwt ~ Tgroup*dayg, filaree)
anova(ReducedModel, FullestModel)
```

#### Analysis of Variance Table

```
Model 1: lnwt ~ day + t + day:t
Model 2: lnwt ~ Tgroup * dayg
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1     41 1.3604
2     30 0.2413 11     1.1192 12.649 1.887e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our results indicate a significant lack of fit meaning the deviations from prediction made by the reduced model to the mean of each set of observations with equal values for all predictors are significantly larger than expected on the basis of the MSE from the fullest model (each combination of days and temperature becomes a parameter.). Biologically speaking, this finding points out that RGR is not constant as plants age. Conclusively, our prognosis in response to the diagnostics would be to add a quadratic term or create a better model than the current one. or a better mechanistic model should be constructed.