

Homework 2

Gitanshu Munjal

October 28, 2014

1 Question 1

Write out the full model (equation). State the null and alternative hypotheses in terms of values of parameters (hypothesis above: all RGR's are the same). [20]

1.1 Full Model(Equation)

$$Y = \beta_0 + \Delta\beta_{01}G_1 + \Delta\beta_{02}G_2 + \Delta\beta_{03}G_3 + (\beta_1 + \Delta\beta_{11}G_1 + \Delta\beta_{12}G_2 + \Delta\beta_{13}G_3)X + \epsilon$$

Where,

β_0 and β_1 are the intercept and slope respectively (for all groups pooled),

$\Delta\beta_{0i}$ is the effect of group i on the intercept (i=1,2,3; 3 temperature groups), $\sum \Delta\beta_{0i} = 0$

$\Delta\beta_{1i}$ is the effect of group i on the slope (i=1,2,3; 3 temperature groups), $\sum \Delta\beta_{1i} = 0$

G_i is a "dummy" variable that has a value of 1 for i= 1,2,3 and 0 for all other values of i

X is a continuous variable (with units of time(days) for our model)

ϵ is a random variable with mean 0, variance σ^2 .

1.2 Hypotheses

We know that Relative Growth Rate (RGR) is given by slope of the regression line that relates natural logarithm of plant size (size = dry weight for our purposes here) to time. (Evans 1972, Plant growth analysis, p.197). Our null hypothesis (H_0) stating that RGR is the same for all groups, independent of temperature can thus be mathematically stated as:

$$H_0 : \beta_{11} = \beta_{12} = \beta_{13}$$

Given, $\sum \Delta\beta_{1i} = 0$ and $\beta_{1i} = \beta_1 + \Delta\beta_{1i}$ The null hypothesis can be rewritten as:

$$H_0 : \Delta\beta_{11} = \Delta\beta_{12} = \Delta\beta_{13} = 0$$

Where symbols assume their meanings described in section 1.1 above. The alternative hypothesis (H_1) can then be stated as:

$H_1 : \text{at least one of the slopes(RGRs) is significantly different from the others}$

or equivalently (Given, $\sum \Delta\beta_{1i} = 0$ and $\beta_{1i} = \beta_1 + \Delta\beta_{1i}$)

$H_1 : \text{at least one of the group effects } (\Delta\beta_{1i}) \text{ is significantly different from 0}$

2 Question 2

Perform a test of this hypothesis by analyzing the data with R, and interpret the corresponding F test. [20]

2.1 Test of Hypothesis: Code and Relevant Output

```
library(multcomp)
library(car)

clover<- read.table("C:\\Users\\gitanshu\\Desktop\\Clover.dat",header=T,sep=",")

clover$group<-as.factor(clover$group)
options(contrasts=c("contr.sum", "contr.poly"))

model1 <- lm(lnwt~group*days, data=clover)
summary(model1)
```

Call:

```
lm(formula = lnwt ~ group * days, data = clover)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.2851	-0.1037	-0.0346	0.1613	0.2764

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.583504	0.062411	25.372	< 2e-16 ***
group1	-0.089557	0.088364	-1.014	0.317
group2	-0.048508	0.090520	-0.536	0.595
days	0.091338	0.002540	35.955	< 2e-16 ***
group1:days	-0.022139	0.003585	-6.176	2.95e-07 ***
group2:days	0.002504	0.003655	0.685	0.497

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

Residual standard error: 0.1705 on 39 degrees of freedom

Multiple R-squared: 0.9769, Adjusted R-squared: 0.9739

F-statistic: 329.2 on 5 and 39 DF, p-value: < 2.2e-16

Above is the result of the correct F-test for our overall model as it uses 5 degrees of freedom for the model and 39 degrees of freedom for the error term. A [highly significant](#) p-value here (p-value: < 2.2e-16) at the 95% confidence level encourages us to partition the model in order to better understand the source of significance.

```
anova(model1)
```

Analysis of Variance Table

Response: lnwt

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	8.549	4.275	146.988	< 2.2e-16 ***
days	1	37.953	37.953	1305.062	< 2.2e-16 ***
group:days	2	1.368	0.684	23.519	1.992e-07 ***
Residuals	39	1.134	0.029		

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

Above are the results of sequential (Type 1) tests for our model. A highly [significant interaction term](#) (p-value: 1.992e-07) here for the interaction between temperature groups and days indicates that the RGRs for the temperature groups are [NOT](#) all the same and so we [reject our null hypothesis](#) from Question 2 at the 95% confidence level.

3 Question 3

Determine which group, if any, differs in RGR. Use the Bonferroni correction. [20]

3.1 Simultaneous test of differences between relative growth rates (slopes)

```
b11<-as.vector(cbind(0,0,0,1,1,0))
b12<-as.vector(cbind(0,0,0,1,0,1))
b13<-as.vector(cbind(0,0,0,1,-1,-1))

H01s <- rbind(B12_B11 = b12-b11,
              B13_B11 = b13-b11,
              B13_B12 = b13-b12)
tH01s <- glht(model1, linfct = H01s)
summary(tH01s, test = adjusted("bonferroni"))
```

Simultaneous Tests for General Linear Hypotheses

Fit: `lm(formula = lnwt ~ group * days, data = clover)`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
B12_B11 == 0	0.024643	0.006317	3.901	0.0011	**
B13_B11 == 0	0.041773	0.006113	6.834	1.09e-07	***
B13_B12 == 0	0.017131	0.006236	2.747	0.0272	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- bonferroni method)

The results from the above simultaneous tests of hypotheses (controlled for familywise error rate using the Bonferroni method) indicate significant differences in RGR between all groups as all comparisons were found to have a significant p value ($\alpha_{critical} = \frac{\alpha}{3} = \frac{0.05}{3} = 0.0167$). In other words, the RGR for each group is significantly different from the RGRs for all other groups at the 95% confidence level.

4 Question 4

Determine if the weight of plants grown in the lowest temperature for 27 days is significantly different from the weight of plants grown at the highest temperature for 11 days: First, (a) state H_0 and H_a in words and with equations, or using statements about parameter values; then, (b) to test this hypothesis, build an L vector and use it in a linear hypothesis in R, and state your conclusion. [20]

4.1 Hypotheses

Let $Y_{1,27}$ denote the weight of plants grown at the lowest temperature (group 1) for 27 days and let $Y_{3,11}$ denote the weight of plants grown at the highest temperature (group 3) for 11 days. Then the null hypothesis is that there is no significant difference between the weights of group 1 plants grown for 27 days in comparison to the weights of group 3 plants grown for 11 days. Mathematically, this can be stated as:

$$H_0 : Y_{1,27} = Y_{3,11}$$

or equivalently,

$$H_0 : Y_{1,27} - Y_{3,11} = 0$$

The alternative hypothesis would then be that there is indeed a significant difference between the weights of group 1 plants grown for 27 days in comparison to the weights of group 3 plants grown for 11 days. Mathematically, this can be stated as:

$$H_0 : Y_{1,27} \neq Y_{3,11}$$

or equivalently,

$$H_0 : Y_{1,27} - Y_{3,11} \neq 0$$

Then using the model matrix and our previously defined full model (with described restrictions), we can calculate L vectors for $Y_{1,27}$ and $Y_{3,11}$. The L vector for the equivalent null hypothesis can then be calculated simply by subtracting the two previously calculated vectors.

4.2 Test of Hypothesis: Code and Relevant Output

```
Y1.27<-as.vector(cbind(1,1,0,27,27,0))
Y3.11<-as.vector(cbind(1,-1,-1,11,-11,-11))

H0s <- rbind(Hypothesis=Y1.27-Y3.11)

tH0s <- glht(model1, lmfct = H0s)
summary(tH0s, test = univariate())
```

Simultaneous Tests for General Linear Hypotheses

```
Fit: lm(formula = lnwt ~ group * days, data = clover)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
Hypothesis == 0	0.42006	0.07932	5.296	4.91e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Univariate p values reported)

The results from the above test of hypothesis (not controlled for familywise error rate because we are only interested in one test) indicate that there is indeed a significant difference (p-value=4.91e-06) at the 95% confidence level between the weight of plants grown in the lowest temperature for 27 days and the weight of plants grown at the highest temperature for 11 days.

5 Question 5

Why would the above analysis be incorrect if the same 3 plants per temperature treatment had been used throughout the experiment as depicted below? What typical assumption of regression would probably be violated? Why? Assume that the plants are not destroyed when measured (if you have trouble assuming this, imagine that you are measuring height instead of mass). Assume that the plants are not in the same greenhouses. [20]

5.1 Violation of Independence of Error

The analysis would be incorrect because the independence of error assumption would be violated and so our usual tests might not hold. One of the most important assumptions of the tests we used in our analysis was independence of errors which we achieved in the past by having true replicates. By making measurements on the same 3 plants over time we are introducing a repeated measures type design and violating the independence of errors assumption adding some structure to the error (consider 1 of the 3 plants, the error for the second measurement would be related to the first measurement as the experimental unit is the same, so on and so forth, we cannot consider them as true replicates.)