Homework7

Rohan Sadale

March 10, 2016

```
library(alr4)
```

```
## Warning: package 'alr4' was built under R version 3.2.3
```

```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 3.2.3
```

```
## Loading required package: effects
```

```
## Warning: package 'effects' was built under R version 3.2.3
```

```
##
## Attaching package: 'effects'
```

```
## The following object is masked from 'package:car':
##
##     Prestige
```

6.1

```
m1 <- lm(lifeExpF ~ 1, UN11)
m2 <- lm(lifeExpF ~ group, UN11)
anova(m1, m2)
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ 1
## Model 2: lifeExpF ~ group
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    198 20293.2
## 2    196  7730.2  2     12563 159.27 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The summary shows that p-value is almost 0. Thus group should not be removed.

6.2

For F-tests to be done, NH should be special case of AH (such as setting one of the β parameter to zero).

6.3

```
m1 <- lm(lifeExpF ~ group, UN11)
m2 <- lm(lifeExpF ~ group + log(ppgdp), UN11)
anova(m1, m2)
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ group
## Model 2: lifeExpF ~ group + log(ppgdp)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    196 7730.2
## 2    195 5090.4  1    2639.8 101.12 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The summary says that as p-value is almost 0, we should not remove log(ppgdp).

6.4

6.4.1 Common intercept and different slopes in each group. The absence of group removes all the different intercepts.

```
m1 <- lm(lifeExpF ~ log(ppgdp) + group:log(ppgdp), UN11)
m2 <- lm(lifeExpF ~ group + log(ppgdp) + group:log(ppgdp), UN11)
anova(m1, m2)
```

6.4.2

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ log(ppgdp) + group:log(ppgdp)
## Model 2: lifeExpF ~ group + log(ppgdp) + group:log(ppgdp)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    195 5232.0
## 2    193 5077.7  2    154.31 2.9326 0.05564 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As p-value isn't 0, we should remove the interaction term.

6.5

```
m1 <- lm(lifeExpF ~ group + log(ppgdp), UN11)
summary(m1)
```

6.5.1

```
##
## Call:
## lm(formula = lifeExpF ~ group + log(ppgdp), data = UN11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.6348  -2.1741   0.2441   2.3537  14.6539
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   49.529      3.400  14.569  < 2e-16 ***
## groupother    -1.535      1.174  -1.308    0.193
## groupafrica  -12.170      1.557  -7.814 3.35e-13 ***
## log(ppgdp)     3.177      0.316  10.056  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.109 on 195 degrees of freedom
## Multiple R-squared:  0.7492, Adjusted R-squared:  0.7453
## F-statistic: 194.1 on 3 and 195 DF,  p-value: < 2.2e-16
```

From the p-value of groupother we can say that it isn't significantly different from the intercept(which is in this case oecd)

```
linearHypothesis(m1, "groupother - groupafrica")
```

6.5.2

```
## Linear hypothesis test
##
## Hypothesis:
## groupother - groupafrica = 0
##
## Model 1: restricted model
## Model 2: lifeExpF ~ group + log(ppgdp)
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    196 8170.2
## 2    195 5090.4  1    3079.8 117.98 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the p-value we can say that intercept of other and africa aren't different.

6.6

```
data(fuel2001)
fuel2001$Dlic = 1000*fuel2001$Drivers/fuel2001$Pop
fuel2001$Fuel = 1000*fuel2001$FuelC/fuel2001$Pop
```

```
fuel2001$Income = fuel2001$Income/1000

m1 <- lm(Fuel ~ 1, fuel2001)
m2 <- lm(Fuel ~ Tax+Dlic+Income+log(Miles), fuel2001)
anova(m2, m1)
```

```
## Analysis of Variance Table
##
## Model 1: Fuel ~ Tax + Dlic + Income + log(Miles)
## Model 2: Fuel ~ 1
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     46 193700
## 2     50 395694 -4   -201994 11.992 9.331e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the summary we can say that model with all the regressors has p-value of almost 0 (significant). Thus we don't need to do any futher testing such as removing regressors.

6.7

```
data(fuel2001)
fuel2001$Dlic = 1000*fuel2001$Drivers/fuel2001$Pop
fuel2001$Fuel = 1000*fuel2001$FuelC/fuel2001$Pop
fuel2001$Income = fuel2001$Income/1000

m1 <- lm(Fuel ~ Tax+Dlic+Income+log(Miles), fuel2001)
m2 <- lm(Fuel ~ log(Miles)+Tax+Dlic+Income, fuel2001)

anova(m1)
```

6.7.1

```
## Analysis of Variance Table
##
## Response: Fuel
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Tax         1  26635   26635  6.3254 0.0154602 *
## Dlic        1  79378   79378 18.8506 7.692e-05 ***
## Income      1  61408   61408 14.5833 0.0003997 ***
## log(Miles)  1  34573   34573  8.2104 0.0062592 **
## Residuals  46 193700    4211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(m2)
```

```
## Analysis of Variance Table
##
```

```
## Response: Fuel
##             Df Sum Sq Mean Sq F value    Pr(>F)
## log(Miles)   1  70478   70478 16.7371 0.0001711 ***
## Tax          1  23024   23024  5.4678 0.0237724 *
## Dlic         1  75553   75553 17.9423 0.0001081 ***
## Income       1  32940   32940  7.8225 0.0075078 **
## Residuals   46 193700    4211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As anova() provides tests in sequential order, there is difference between the anova for m1 and m2.

```
data(fuel2001)
fuel2001$Dlic = 1000*fuel2001$Drivers/fuel2001$Pop
fuel2001$Fuel = 1000*fuel2001$FuelC/fuel2001$Pop
fuel2001$Income = fuel2001$Income/1000

m1 <- lm(Fuel ~ Tax+Dlic+Income+log(Miles), fuel2001)
m2 <- lm(Fuel ~ log(Miles)+Tax+Dlic+Income, fuel2001)

Anova(m1)
```

6.7.2

```
## Anova Table (Type II tests)
##
## Response: Fuel
##             Sum Sq Df F value    Pr(>F)
## Tax          18264  1  4.3373 0.0428733 *
## Dlic         56770  1 13.4819 0.0006256 ***
## Income       32940  1  7.8225 0.0075078 **
## log(Miles)   34573  1  8.2104 0.0062592 **
## Residuals   193700 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(m2)
```

```
## Anova Table (Type II tests)
##
## Response: Fuel
##             Sum Sq Df F value    Pr(>F)
## log(Miles)   34573  1  8.2104 0.0062592 **
## Tax          18264  1  4.3373 0.0428733 *
## Dlic         56770  1 13.4819 0.0006256 ***
## Income       32940  1  7.8225 0.0075078 **
## Residuals   193700 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The regressors which are added at the last in Type 1 anova have same estimates as respective those in Type 2 anova.