GU4205/5205-Linear Regression Models-Lab2a

Author: Dr. Banu Baydil

Fall 2022

Section 1: Confidence Intervals for Regression Coefficients

In this section we will continue to work with the 2011 UN data.

```
library(alr4)
attach(UN11)
dim(UN11)
```

```
## [1] 199   6
```

```
names(UN11)
```

```
## [1] "region"    "group"     "fertility" "ppgdp"     "lifeExpF"  "pctUrban"
```

```
help(UN11)
```

Let us recall from last Lab the regression of lifeExpF on log(ppgdp):

```
m1 <- lm(lifeExpF ~ log(ppgdp), data=UN11)
summary(m1)
```

```
##
## Call:
## lm(formula = lifeExpF ~ log(ppgdp), data = UN11)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.749  -2.879   1.280   3.987  12.345
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  29.8148     2.5314   11.78   <2e-16 ***
## log(ppgdp)    5.0188     0.2942   17.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.448 on 197 degrees of freedom
## Multiple R-squared:  0.5964, Adjusted R-squared:  0.5943
## F-statistic: 291.1 on 1 and 197 DF,  p-value: < 2.2e-16
```

Note that beta1.hat = 5.02, beta0.hat = 29.8, sigma.hat = 6.45, and R.squared = .596. About 60% of variability in lifeExpF is explained by log(ppgdp).

We can obtain a 95% CI for beta1.hat and beta0.hat as:

```
confint(m1, level = 0.95)
```

```
##                  2.5 %    97.5 %
## (Intercept) 24.822750 34.806855
## log(ppgdp)   4.438718  5.598953
```

We can alternatively obtain only the model coefficient of determintaion, R-squared, as:

```
summary(m1)$r.squared
```

```
## [1] 0.5963835
```

or, as:

```
r=cor(lifeExpF,log(ppgdp))
Rsquared=r^2
Rsquared
```

```
## [1] 0.5963835
```

We can conduct model F-test as:

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: lifeExpF
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## log(ppgdp)   1 12102.5 12102.5  291.09 < 2.2e-16 ***
## Residuals  197  8190.7    41.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that the F-statistic = 291.09 is square of the t-statistic (= 17.06) in the summary(m1) of the previous computation.

Section 2: Effects Plots

Next, let us recall from last Lab the regression of y=lifeExpF on x1=log(ppgdp) and x2=fertility:

```
m12 <- lm(lifeExpF ~ log(ppgdp) + fertility, data=UN11)
summary(m12)
```

```
##
## Call:
## lm(formula = lifeExpF ~ log(ppgdp) + fertility, data = UN11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.6362  -1.6854   0.4221   2.7301  11.7978
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.4484     3.7446  16.944  < 2e-16 ***
## log(ppgdp)    2.4150     0.3386   7.132 1.86e-11 ***
## fertility    -4.1991     0.3938 -10.664  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 5.142 on 196 degrees of freedom
## Multiple R-squared:  0.7446, Adjusted R-squared:  0.742
## F-statistic: 285.7 on 2 and 196 DF,  p-value: < 2.2e-16
```

Alternatively, we can obtain:

```
coef(m12)
```

```
## (Intercept)  log(ppgdp)    fertility
##   63.448396    2.415015    -4.199124
```

```
summary(m12)$coef[,2]    #Standard errors of coefficients
```

```
## (Intercept)  log(ppgdp)    fertility
##   3.7446417   0.3386055    0.3937541
```

```
vcov(m12)       #Estimated covariance matrix of beta.hat, note the diagonal elements
```

```
##              (Intercept)  log(ppgdp)    fertility
## (Intercept)    14.022341 -1.23588358 -1.24183722
## log(ppgdp)     -1.235884  0.11465368  0.09613964
## fertility      -1.241837  0.09613964  0.15504226
```

```
summary(m12)$coef[3,]   # Test of NH: beta2 = 0 vs AH: beta2 ^= 0
```

```
##      Estimate     Std. Error        t value       Pr(>|t|)
## -4.199124e+00  3.937541e-01  -1.066433e+01   3.119790e-21
```

```
summary(m12)$r.squared
```
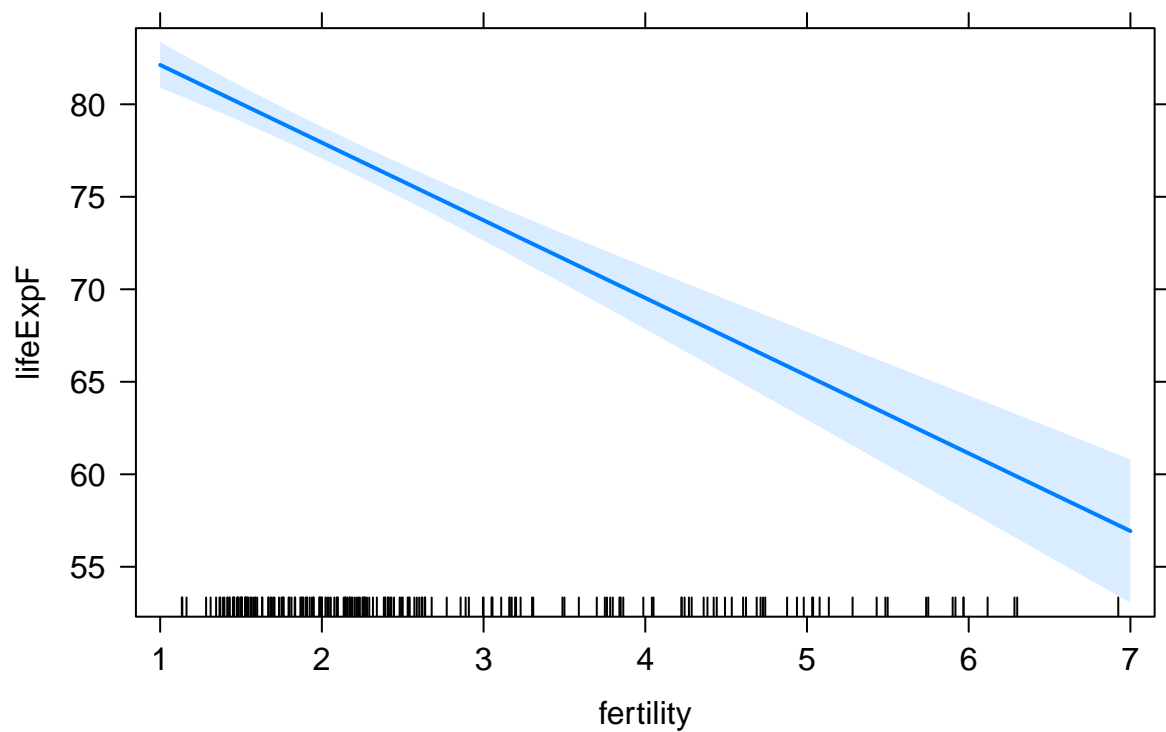
```
## [1] 0.7445861
```

```
summary(m12)$sigma
```

```
## [1] 5.142446
```

We can obtain the effects plots of fertility, log(ppgdp), and ppgdp as:
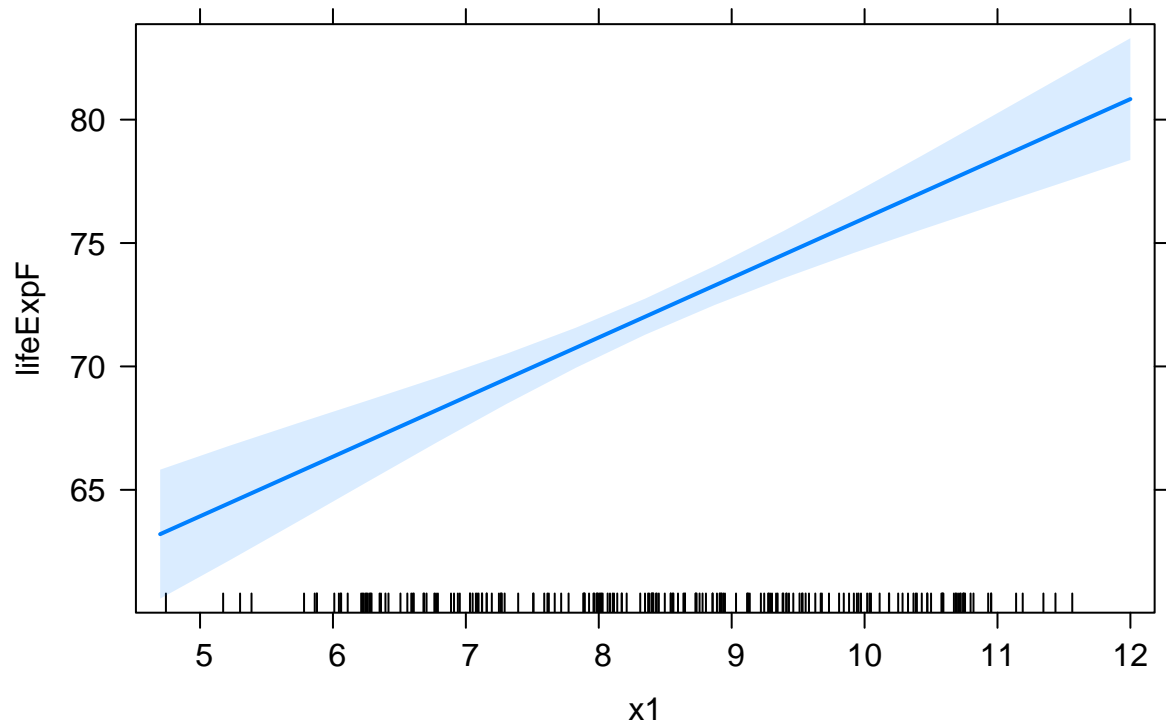
```
plot(Effect("fertility", m12))
```
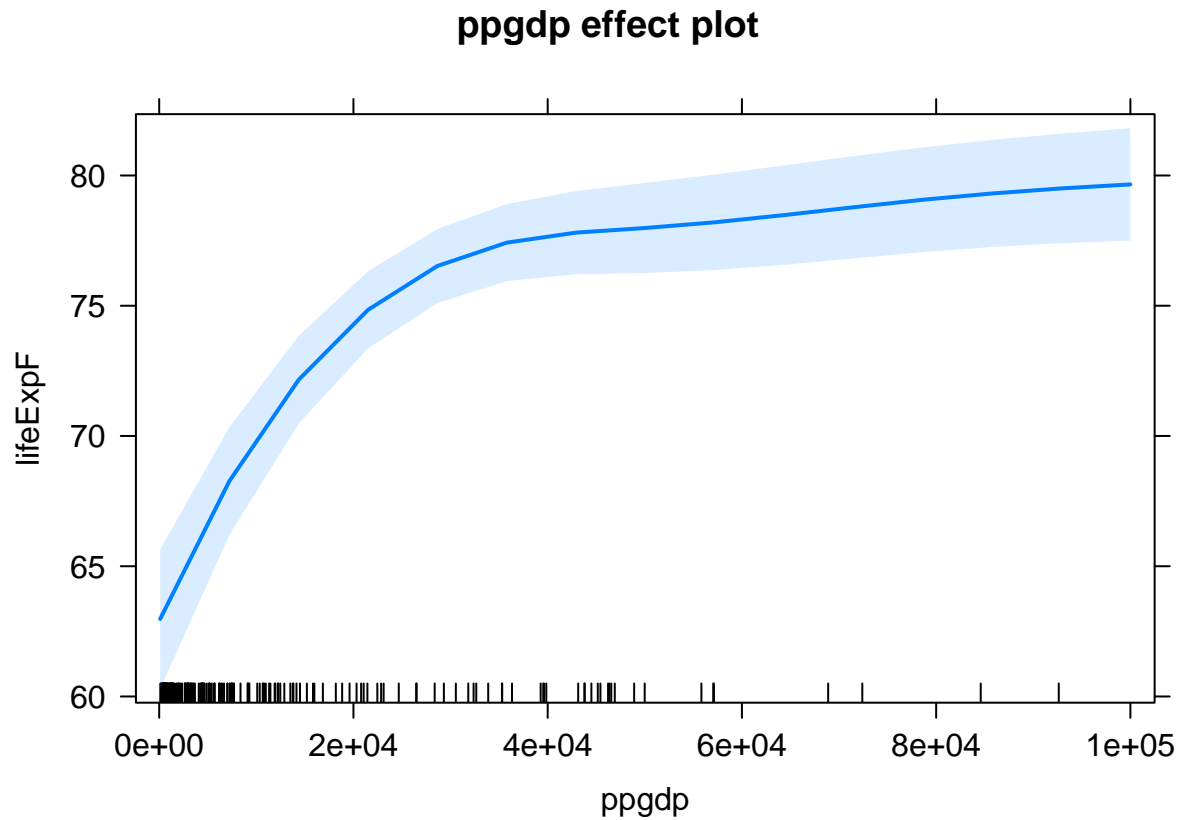
fertility effect plot


```
x1=log(ppgdp)
m12x1 <- lm(lifeExpF ~ x1 + fertility, data=UN11)
plot(Effect("x1", m12x1))   #Note, this is linear
```

x1 effect plot


```r
plot(Effect("ppgdp", m12)) #Note, this is nonlinear
```
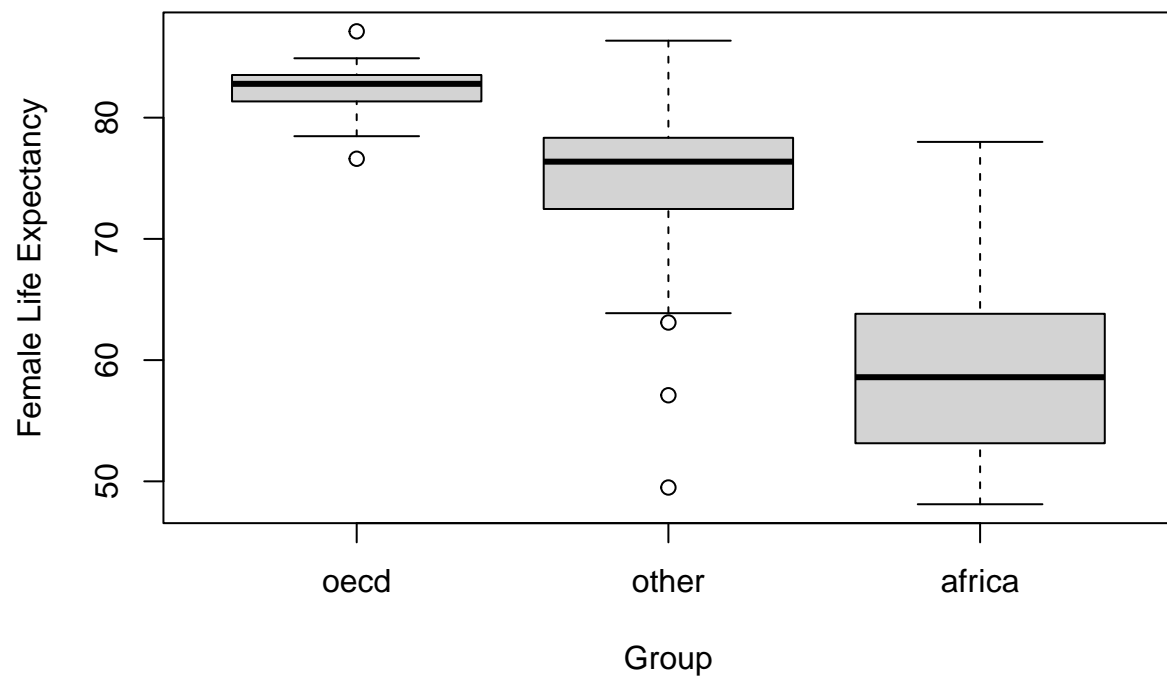
ppgdp effect plot

Section 3: One Categorical and One Continuous Predictor

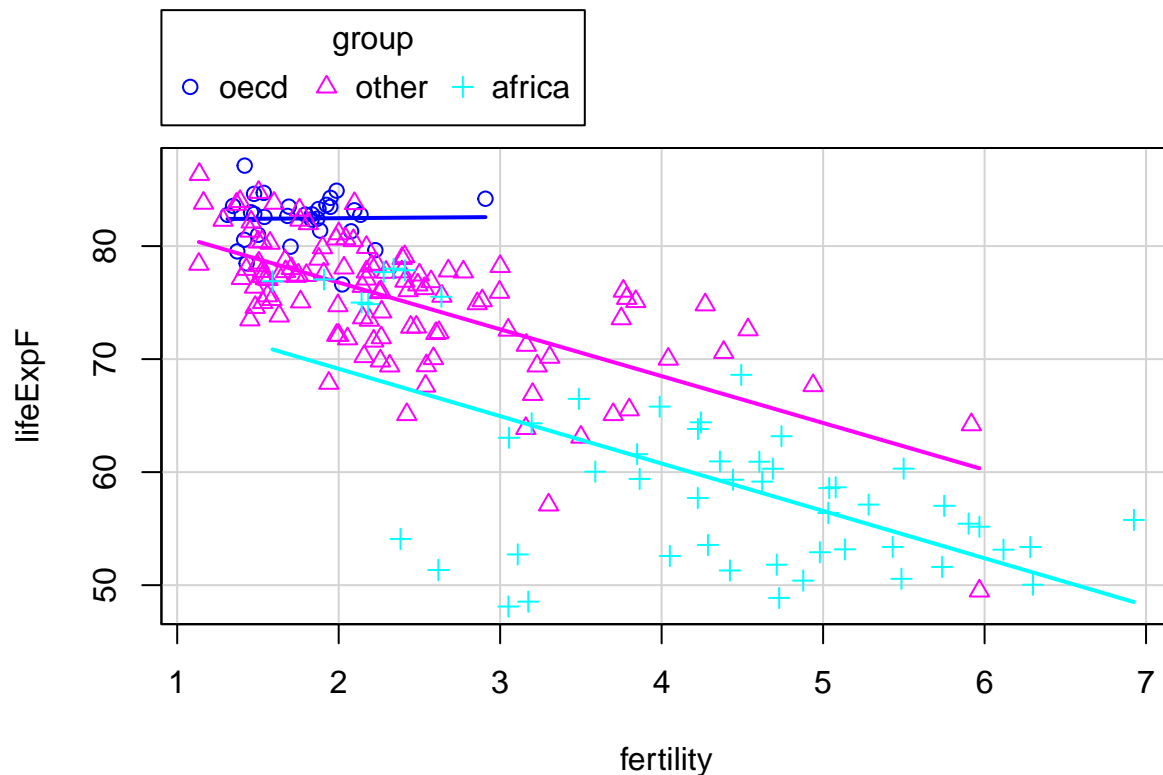Let us consider regressing lifeExpF on the predictors fertility and group.

```r
#Marginal effect of the predictor 'group' can be visualized as:
boxplot(lifeExpF ~ group, data=UN11, xlab="Group", ylab="Female Life Expectancy")
```

Let us plot lifeExpF vs fertility grouped by the predictor 'group':

```
scatterplot(lifeExpF ~ fertility | group, data=UN11, smooth=F)
```

```
# Coincident mean function (no 'group' variable effect)
m4a <- lm(lifeExpF ~ fertility, data=UN11);
summary(m4a)
```

```
##
## Call:
## lm(formula = lifeExpF ~ fertility, data = UN11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.3806  -2.6855   0.5826   3.6434  12.8156
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  89.4805     0.9369   95.51   <2e-16 ***
## fertility    -6.2242     0.3054  -20.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.757 on 197 degrees of freedom
## Multiple R-squared:  0.6783, Adjusted R-squared:  0.6767
## F-statistic: 415.4 on 1 and 197 DF,  p-value: < 2.2e-16
```

```
# Parallel mean function
m4b <- lm(lifeExpF ~ group + fertility, data=UN11);
summary(m4b)
```

```
## 
## Call:
## lm(formula = lifeExpF ~ group + fertility, data = UN11)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -16.506  -2.435   0.399   3.038  10.624 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  89.6607     1.0668  84.049  < 2e-16 ***
## groupother   -4.7285     0.9999  -4.729 4.32e-06 ***
## groupafrica -12.5751     1.3955  -9.011  < 2e-16 ***
## fertility    -4.0870     0.3510 -11.645  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 4.835 on 195 degrees of freedom
## Multiple R-squared:  0.7753, Adjusted R-squared:  0.7719 
## F-statistic: 224.3 on 3 and 195 DF,  p-value: < 2.2e-16
# Coincident at zero
m4c <- lm(lifeExpF ~ fertility + group:fertility, data=UN11);  summary(m4c)
```

```
## 
## Call:
## lm(formula = lifeExpF ~ fertility + group:fertility, data = UN11)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -18.5975  -2.2032   0.6361   3.0917  10.1266 
## 
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           83.2968     1.0723  77.681  < 2e-16 ***
## fertility             -0.4615     0.7646  -0.604    0.547    
## fertility:groupother  -3.0300     0.5569  -5.441 1.58e-07 ***
## fertility:groupafrica -4.9758     0.6244  -7.969 1.31e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 4.908 on 195 degrees of freedom
## Multiple R-squared:  0.7686, Adjusted R-squared:  0.765 
## F-statistic: 215.8 on 3 and 195 DF,  p-value: < 2.2e-16
# Separate mean functions for each country group, full model
m4d <- lm(lifeExpF ~ group + fertility + group:fertility, data=UN11);
summary(m4d)
```

```
## 
## Call:
## lm(formula = lifeExpF ~ group + fertility + group:fertility, 
##     data = UN11)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
```

```
## -16.6279  -2.3652   0.3772   2.7183  10.4365
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            82.2675     4.6604  17.652   <2e-16 ***
## groupother              2.7947     4.8204   0.580    0.563
## groupafrica            -4.7467     5.1856  -0.915    0.361
## fertility               0.1014     2.5941   0.039    0.969
## groupother:fertility   -4.2437     2.6396  -1.608    0.110
## groupafrica:fertility  -4.2911     2.6445  -1.623    0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.827 on 193 degrees of freedom
## Multiple R-squared:  0.7784, Adjusted R-squared:  0.7726
## F-statistic: 135.6 on 5 and 193 DF,  p-value: < 2.2e-16
```

Let us test for the main effect of the predictor 'group' and group:fertility interaction simultaneously:

We will use F-test with the full model m4d and the null model m4a. The latter corresponds to NH: beta02 = beta 03 = beta12 = beta13 = 0.

```
anova(m4a, m4d)
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ fertility
## Model 2: lifeExpF ~ group + fertility + group:fertility
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1    197 6528.4
## 2    193 4497.3  4      2031 21.79 7.466e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let us test only for the interaction effect, i.e are separate slopes necessary?

We will use F-test with the full model m4d and the null model m4b. The latter corresponds to NH: beta12 = beta13 = 0.

```
anova(m4b, m4d)
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ group + fertility
## Model 2: lifeExpF ~ group + fertility + group:fertility
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    195 4559.3
## 2    193 4497.3  2    61.977 1.3298 0.2669
```