16

Bernat Chiva

12/6/2020

```r
library(car)

## Loading required package: carData

library(HH)

## Loading required package: lattice

## Loading required package: grid

## Loading required package: latticeExtra

## Loading required package: multcomp

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

## Loading required package: gridExtra

##
## Attaching package: 'HH'

## The following objects are masked from 'package:car':
##
##     logit, vif

library(tables)
library(RcmdrMisc)

## Loading required package: sandwich

library(doBy)
library(emmeans)
```

```
##
## Attaching package: 'emmeans'

## The following object is masked from 'package:HH':
##
##     as.glht

library(qwraps2)

##
## Attaching package: 'qwraps2'

## The following object is masked from 'package:HH':
##
##     logit

## The following object is masked from 'package:car':
##
##     logit

library(readr)
PRACOVAR <- read_delim("C:/Users/berna/OneDrive/Desktop/UPC/S1/5. Models Line
als/datasets/PRACOVAR.csv",
    ";", escape_double = FALSE, locale = locale(decimal_mark = ","),
    trim_ws = TRUE)

##
## -- Column specification ---------------------------------------------------
------
## cols(
##   FACTOR = col_double(),
##   X = col_double(),
##   Y1 = col_double(),
##   Y2 = col_double(),
##   Y3 = col_double(),
##   Y4 = col_double(),
##   Y5 = col_double(),
##   Y6 = col_double(),
##   Y7 = col_double(),
##   Y8 = col_double()
## )

PRACOVAR$FACTOR<-as.factor(PRACOVAR$FACTOR)
summary(PRACOVAR)

##  FACTOR       X                Y1               Y2               Y3
##  1:10   Min.   :10.30   Min.   :27.90   Min.   :26.60   Min.   :20.50
##  2:10   1st Qu.:13.00   1st Qu.:32.60   1st Qu.:30.52   1st Qu.:22.25
##  3:10   Median :15.75   Median :34.40   Median :34.80   Median :23.85
##  4:10   Mean   :15.40   Mean   :34.73   Mean   :33.90   Mean   :23.89
##         3rd Qu.:18.25   3rd Qu.:37.00   3rd Qu.:37.02   3rd Qu.:25.43
##         Max.   :19.90   Max.   :41.20   Max.   :40.60   Max.   :28.30
```

```
##          Y4              Y5              Y6              Y7
##   Min.   :28.00   Min.   :14.60   Min.   :30.80   Min.   :20.20
##   1st Qu.:30.95   1st Qu.:21.85   1st Qu.:33.58   1st Qu.:22.10
##   Median :32.50   Median :24.00   Median :35.25   Median :23.25
##   Mean   :33.33   Mean   :22.92   Mean   :35.25   Mean   :23.96
##   3rd Qu.:36.12   3rd Qu.:25.82   3rd Qu.:37.00   3rd Qu.:26.23
##   Max.   :39.30   Max.   :29.10   Max.   :39.90   Max.   :29.20
##          Y8
##   Min.   :17.70
##   1st Qu.:21.00
##   Median :22.95
##   Mean   :22.96
##   3rd Qu.:25.07
##   Max.   :27.50
```

```
dim(PRACOVAR)
```

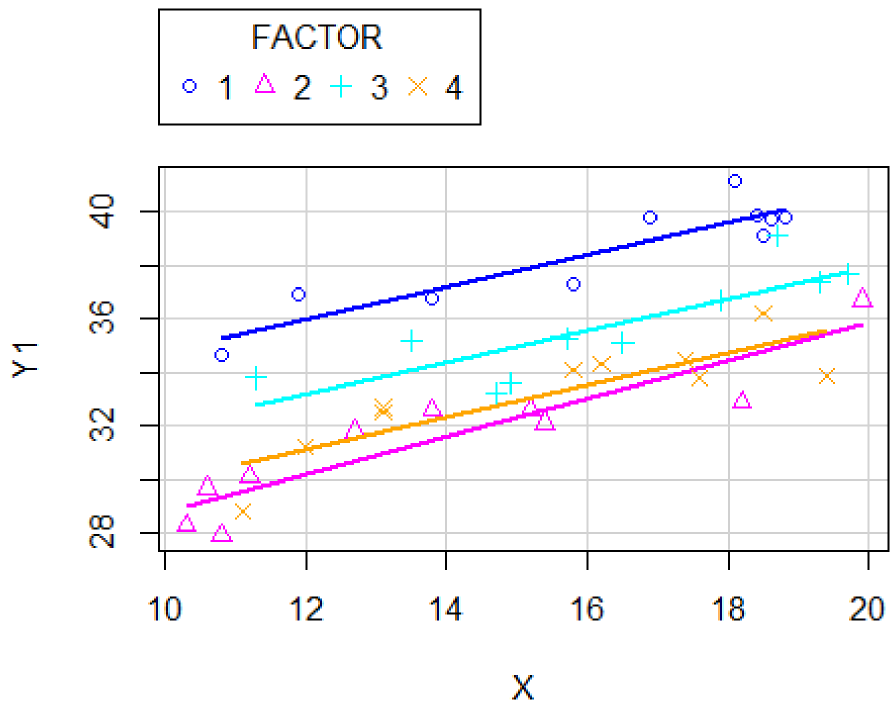
```
## [1] 40 10
```

1. Y1

```
#Mean ans sd table
aa<-tabular((FACTOR+1)~(X+Y1)*((n=1)+mean+sd),PRACOVAR)
pander(aa, style='multiline', plan.ascii=F)
```

	X			Y1		
FACTOR	n	mean	sd	n	mean	sd
1	10	16.16	2.987	10	38.52	1.993
2	10	13.81	3.342	10	31.47	2.588
3	10	16.22	2.720	10	35.71	1.958
4	10	15.42	2.904	10	33.20	2.046
All	40	15.40	3.043	40	34.73	3.401

```
scatterplot(Y1~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

Checking the scatterplot and the tables, we see seemingly parallel lines, so we assume there is no interaction between the covariate and the factor.

The response variable could be the punctuation in an exam, the covariate could be the age and the factor the school the observations attend to. With this model we would like to conclude if the school a student attend to has an effect on the punctuation of this exact time, controlling by the age. We can think of the exam something like an English official exam for example.

To collect the data we would randomly select 10 tests from students of each of the four school, this way we have a balanced design and the observations are independent.

1.1. Model with interaction

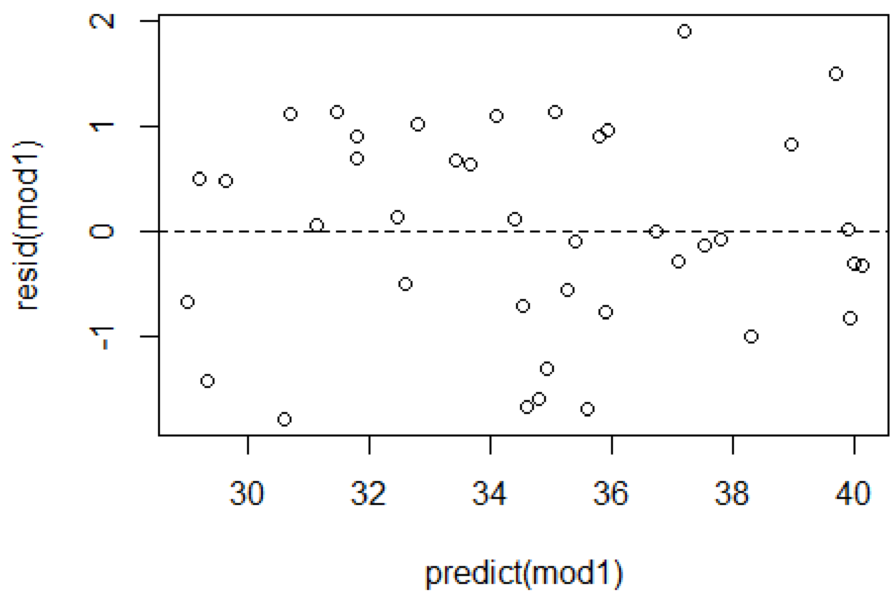
```
mod1<-lm(Y1~X*FACTOR, PRACOVAR)
summary(mod1)

##
## Call:
## lm(formula = Y1 ~ X * FACTOR, data = PRACOVAR)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -1.7958 -0.6853  0.0054  0.8479  1.9138 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 28.714206   1.959631  14.653 9.59e-16 ***
## X            0.606794   0.119442   5.080 1.57e-05 ***
## FACTOR2     -7.057988   2.475422  -2.851  0.00756 ** 
## FACTOR3     -2.658829   2.912380  -0.913  0.36810    
## FACTOR4     -4.809778   2.746794  -1.751  0.08952 .  
## X:FACTOR2    0.103834   0.160188   0.648  0.52148    
## X:FACTOR3   -0.011565   0.177410  -0.065  0.94843    
## X:FACTOR4   -0.003969   0.171363  -0.023  0.98167    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.07 on 32 degrees of freedom
## Multiple R-squared:  0.9187, Adjusted R-squared:  0.901 
## F-statistic: 51.68 on 7 and 32 DF,  p-value: 1.184e-15
```

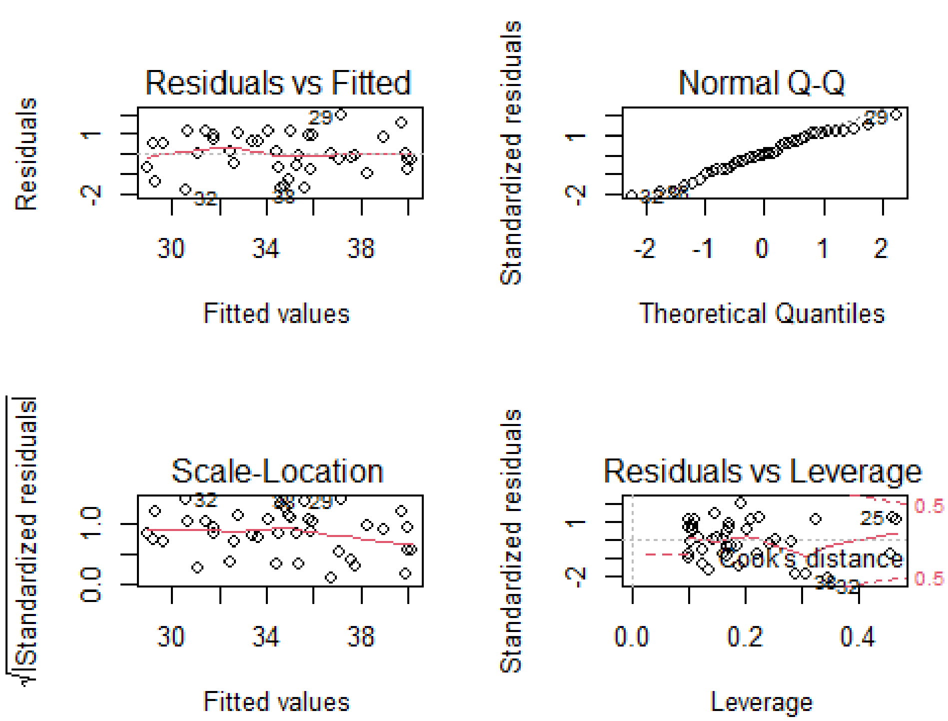
```
anova(mod1)
```

```
## Analysis of Variance Table
## 
## Response: Y1
##           Df  Sum Sq Mean Sq  F value    Pr(>F)    
## X          1 239.241 239.241 208.8268 1.413e-15 ***
## FACTOR     3 174.371  58.124  50.7344 2.906e-12 ***
## X:FACTOR   3   0.823   0.274   0.2394    0.8682    
## Residuals 32  36.661   1.146                       
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(predict(mod1),resid(mod1))
abline(h=0,lty=2)
```




```
oldpar <- par( mfrow=c(2,2))
plot(mod1, ask=F)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

Yes, both the factor and the covariate are significant.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

No, the significance of the interaction in the Ancova table is not significant, then we assume the lines are parallel.

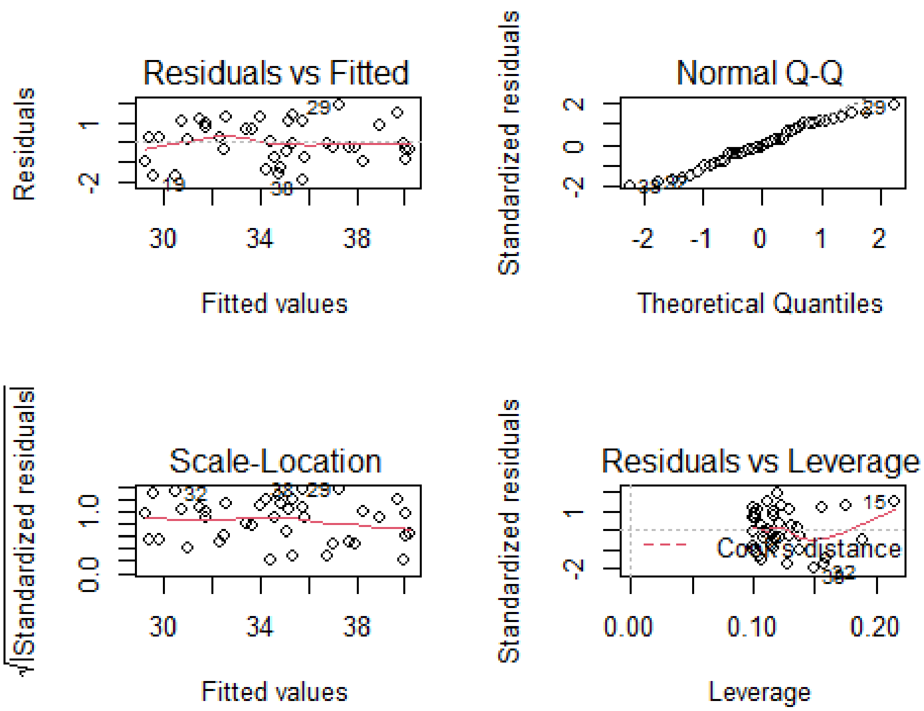
• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

By checking the plots we can assume homoscedasticity and normality, but our model predicts an interaction that is not significative, so an additive model would fit better.

1.2. Model without interaction

```
#model without interaction
mod12<-lm(Y1~X+FACTOR, PRACOVAR)
summary(mod12)

##
## Call:
## lm(formula = Y1 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.83038 -0.78650 -0.06112  0.85089  1.81328
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 28.24589    0.98597  28.648  < 2e-16 ***
## X            0.63577    0.05755  11.046 5.89e-13 ***
## FACTOR2     -5.55593    0.48217 -11.523 1.83e-13 ***
## FACTOR3     -2.84815    0.46282  -6.154 4.84e-07 ***
## FACTOR4     -4.84953    0.46476 -10.434 2.75e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.035 on 35 degrees of freedom
## Multiple R-squared:  0.9169, Adjusted R-squared:  0.9074
## F-statistic: 96.55 on 4 and 35 DF,  p-value: < 2.2e-16
```

```r
anova(mod12)

## Analysis of Variance Table
##
## Response: Y1
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X          1 239.241 239.241 223.392 < 2.2e-16 ***
## FACTOR     3 174.371  58.124  54.273 3.019e-13 ***
## Residuals 35  37.483   1.071
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

oldpar <- par( mfrow=c(2,2))
plot(mod12, ask=F)
```




```r
par(oldpar)
```

• Is there any significant variable?

Yes, again, both the covariate and the factor, each level, are significant to the response variable.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

By means of the residual plots we can assume homoscedasticity and normality.

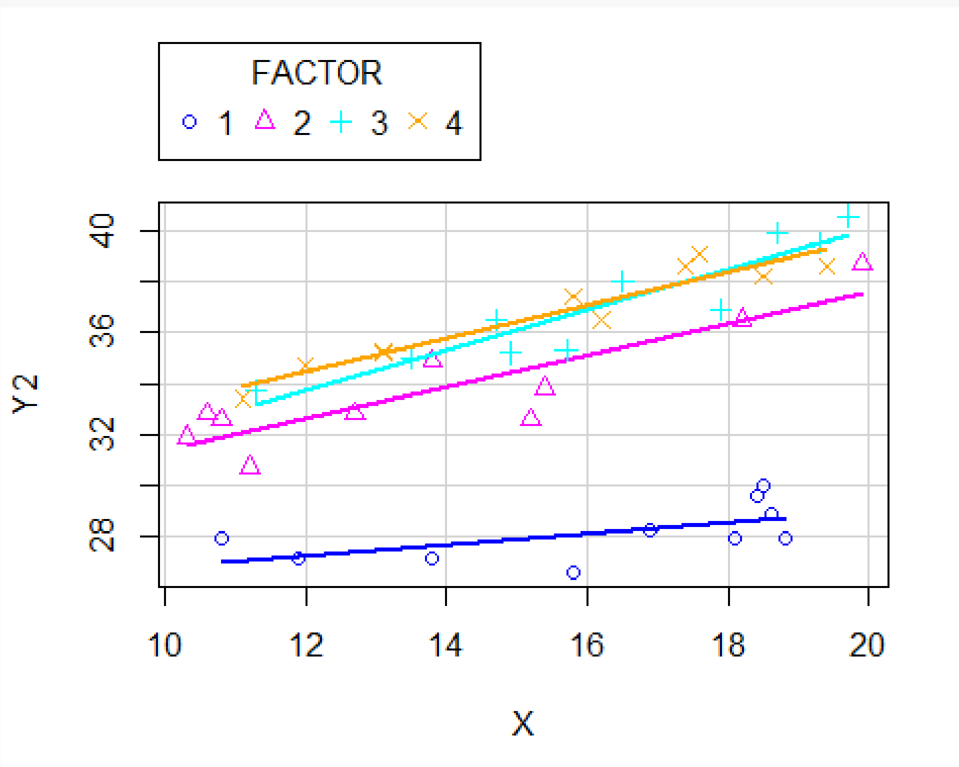
• Which model do you think is more appropriate?

The additive model is better than the factorial one, as we have less parameters and the R squared is really similar as the factor model.

2. Y2

```
scatterplot(Y2~X|FACTOR,smooth=F,dat=PRACOVAR)

#Mean ans sd table
a<-tabular((FACTOR+1)~(X+Y2)*((n=1)+mean+sd),PRACOVAR)
pander(a, style='multiline', plan.ascii=F)
```

	X			Y2		
FACTOR	n	mean	sd	n	mean	sd
1	10	16.16	2.987	10	28.12	1.099
2	10	13.81	3.342	10	33.73	2.370
3	10	16.22	2.720	10	37.06	2.354
4	10	15.42	2.904	10	36.70	1.968
All	40	15.40	3.043	40	33.90	4.110



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

Data is very similar to the previous one, but one level of the factor performs worst than the rest.

We could think of the response variable to be the high jump value of an sport event of one person, the factor could be the type of shoes they use and the covariate could be the age. We would see how one type of shoe performs worst than the rest, but the rest performs very similar, and how the older you get the better the high jump you perform.

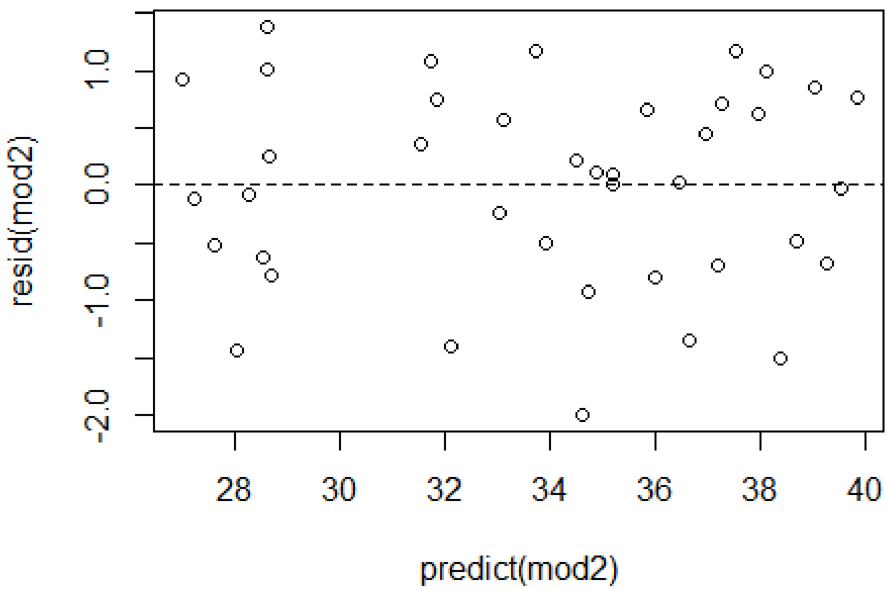
We can collect the data by observing a high jump competition and writing down the age, the type of shoes and the high jump score of each participant.

2.1. Model with interaction

```
#Model with interaction
mod2<-lm(Y2~X*FACTOR, PRACOVAR)
summary(mod2)

##
## Call:
## lm(formula = Y2 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9969 -0.6447  0.0669  0.7239  1.3803
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.6692     1.7444  14.142 2.58e-15 ***
## X             0.2135     0.1063   2.008 0.053105 .
## FACTOR2       0.4482     2.2036   0.203 0.840101
## FACTOR3      -0.5772     2.5925  -0.223 0.825223
## FACTOR4       2.0503     2.4451   0.839 0.407946
## X:FACTOR2     0.4101     0.1426   2.876 0.007109 **
## X:FACTOR3     0.5860     0.1579   3.710 0.000784 ***
## X:FACTOR4     0.4337     0.1525   2.843 0.007718 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9528 on 32 degrees of freedom
## Multiple R-squared:  0.9559, Adjusted R-squared:  0.9463
## F-statistic:  99.1 on 7 and 32 DF,  p-value: < 2.2e-16

anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: Y2
##             Df Sum Sq Mean Sq  F value    Pr(>F)
## X            1  77.64  77.643  85.5272 1.481e-10 ***
## FACTOR       3 537.66 179.219 197.4179 < 2.2e-16 ***
## X:FACTOR     3  14.44   4.813   5.3018  0.004417 **
## Residuals 32  29.05   0.908
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(predict(mod2),resid(mod2))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

Yes, both the covariate and the factor are significant to the response variable.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

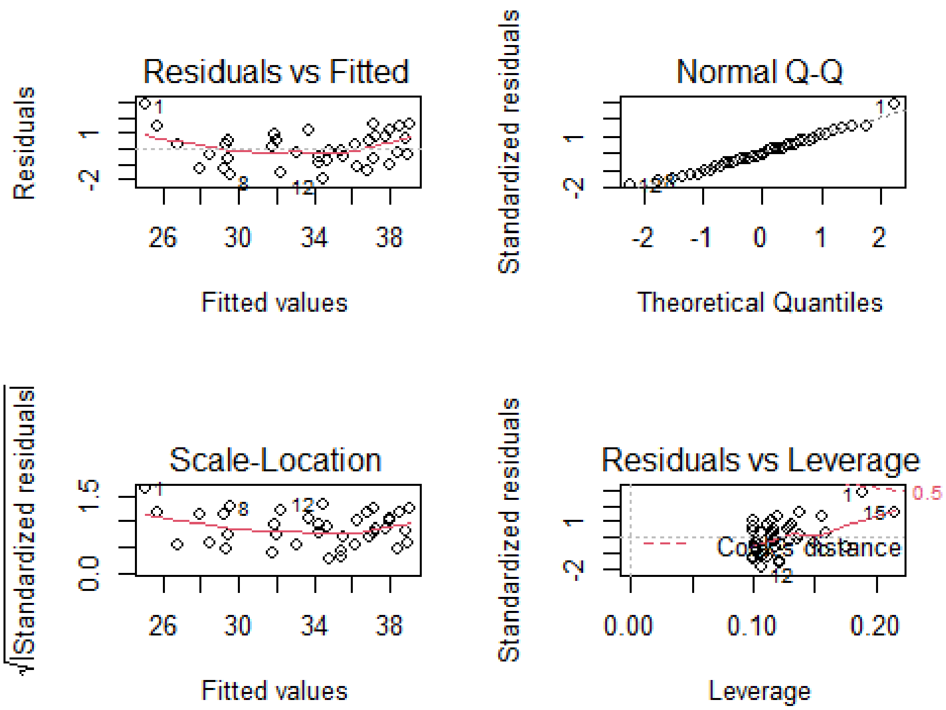
The lines are not all parallel and we checked that the interaction term is significant.

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

Yes, we can assume that the assumptions for the ancova are verified.

2.2. Model without interaction

```
#model without interaction
mod22<-lm(Y2~X+FACTOR, PRACOVAR)
summary(mod22)

##
## Call:
## lm(formula = Y2 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.91331 -0.68617 -0.08266  0.72379  2.80054
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.01331    1.06203   17.90  < 2e-16 ***
## X            0.56353    0.06199    9.09 9.68e-11 ***
## FACTOR2      6.93430    0.51936   13.35 2.68e-15 ***
## FACTOR3      8.90619    0.49852   17.86  < 2e-16 ***
## FACTOR4      8.99701    0.50061   17.97  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.115 on 35 degrees of freedom
## Multiple R-squared:  0.934,  Adjusted R-squared:  0.9264
## F-statistic: 123.8 on 4 and 35 DF,  p-value: < 2.2e-16

anova(mod22)

## Analysis of Variance Table
##
## Response: Y2
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X          1  77.64  77.643  62.487 2.702e-09 ***
## FACTOR     3 537.66 179.219 144.235 < 2.2e-16 ***
## Residuals 35  43.49   1.243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

oldpar <- par( mfrow=c(2,2))
plot(mod22, ask=F)
```




```r
par(oldpar)
```

• Is there any significant variable?

Both the covariate and the factor are significant to the response variable.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

We see that we could assume homoscedasticity and normality but as well we see some observations that might be influential, high leverage and residual that could make the model worst.

• Which model do you think is more appropriate?

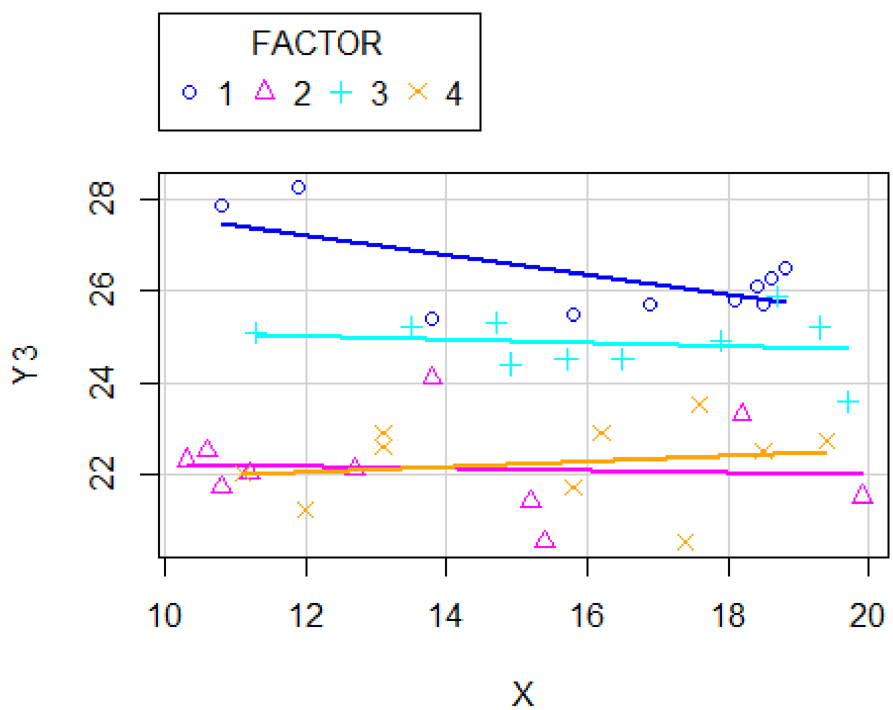
The model with the interaction term has better results when checking the residuals. As well the R squared of the model with interaction is higher than the additive one, because the interaction is significant.

3. Y3

```
#Mean ans sd table
b<-tabular((FACTOR+1)~(X+Y3)*((n=1)+mean+sd),PRACOVAR)
pander(b, style='multiline', plan.ascii=F)
```

	X			Y3		
FACTOR	n	mean	sd	n	mean	sd
1	10	16.16	2.987	10	26.32	1.0031
2	10	13.81	3.342	10	22.14	1.0113
3	10	16.22	2.720	10	24.86	0.6346
4	10	15.42	2.904	10	22.25	0.9022
All	40	15.40	3.043	40	23.89	1.9948

```
scatterplot(Y3~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

From the scatterplot it seems there is no relation between the covariate and the response variable and the factor besides the first level of the factor.

By this non-relation we could think of the response variable as the temperature at a room, the covariate as the level of humidity and the factor as the type of light in the room.

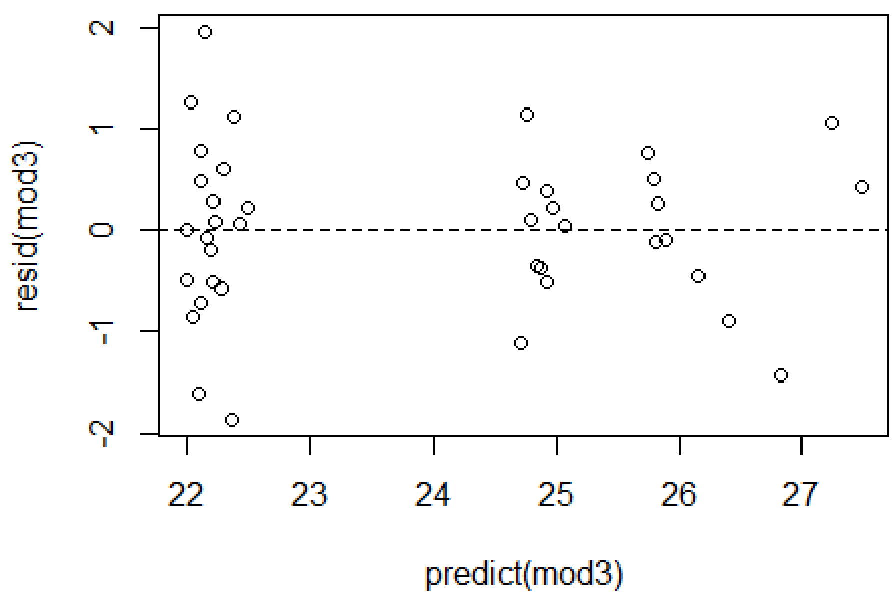
To collect the data we would need to experiment and measure the temperature of the room using different type of light bulbs ten times each, and checking at the same time the humidity. We could randomize the order of what experiment is first so we can control any other variable that might effect the temperature.

3.1. Model with interaction

```
#Model with interaction
mod3<-lm(Y3~X*FACTOR, PRACOVAR)
summary(mod3)

##
## Call:
## lm(formula = Y3 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.86665 -0.49727  0.05185  0.47285  1.95976
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.82592    1.61796  18.434  < 2e-16 ***
## X           -0.21695    0.09862  -2.200 0.035150 *
## FACTOR2     -7.35025    2.04382  -3.596 0.001072 **
## FACTOR3     -4.29058    2.40459  -1.784 0.083856 .
## FACTOR4     -8.48434    2.26788  -3.741 0.000721 ***
## X:FACTOR2    0.19264    0.13226   1.457 0.154977
## X:FACTOR3    0.17531    0.14648   1.197 0.240152
## X:FACTOR4    0.27586    0.14148   1.950 0.060012 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8837 on 32 degrees of freedom
## Multiple R-squared:  0.839,  Adjusted R-squared:  0.8037
## F-statistic: 23.82 on 7 and 32 DF,  p-value: 5.409e-11

anova(mod3)

## Analysis of Variance Table
##
## Response: Y3
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X          1   3.540   3.540  4.5323   0.04105 *
## FACTOR     3 123.460  41.153 52.6950 1.763e-12 ***
```

```
## X:FACTOR   3   3.197   1.066  1.3645   0.27125
## Residuals 32  24.991   0.781
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(predict(mod3),resid(mod3))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

Both the covariate and the factor are significant to the response variable.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

The interaction term is not significant, so we assume the lines are parallel.

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

The model does not look quite good, we cannot assume homoscedasticity.

3.2. Model without interaction

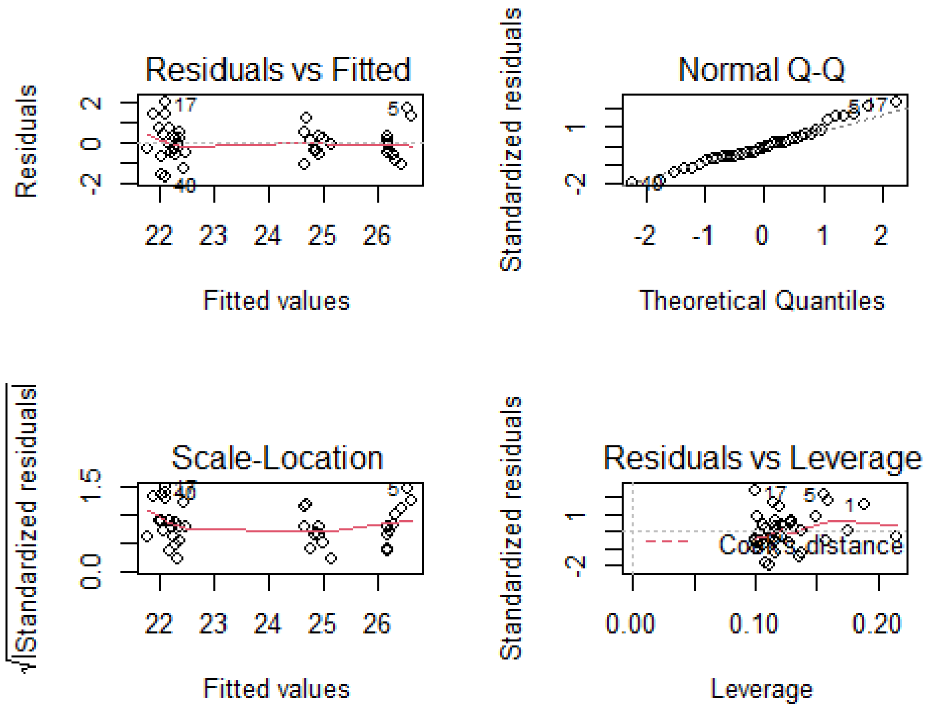
```r
#model without interaction
mod32<-lm(Y3~X+FACTOR, PRACOVAR)
summary(mod32)
```

```
##
## Call:
## lm(formula = Y3 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63873 -0.53003 -0.06568  0.44558  1.95944
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.22810    0.85502  31.845  < 2e-16 ***
## X           -0.05619    0.04991  -1.126 0.267866
## FACTOR2     -4.31206    0.41813 -10.313 3.76e-12 ***
## FACTOR3     -1.45663    0.40135  -3.629 0.000899 ***
## FACTOR4     -4.11158    0.40304 -10.202 5.02e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8974 on 35 degrees of freedom
## Multiple R-squared:  0.8184, Adjusted R-squared:  0.7976
## F-statistic: 39.42 on 4 and 35 DF,  p-value: 1.665e-12
```

```r
anova(mod32)
```

```
## Analysis of Variance Table
##
## Response: Y3
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X          1   3.540   3.540   4.395   0.04334 *
## FACTOR     3 123.460  41.153  51.099 7.122e-13 ***
## Residuals 35  28.188   0.805
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
oldpar <- par( mfrow=c(2,2))
plot(mod32, ask=F)
```




```
par(oldpar)
```

• Is there any significant variable?

Both the factor and the covariate are significant to the response variable.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

We can assume homoscedasticity and normality for the residuals.

• Which model do you think is more appropriate?

As the interaction term is not significant at 5% is better to stick to the additive model, so we have less parameters. The R squared of both models are similar.

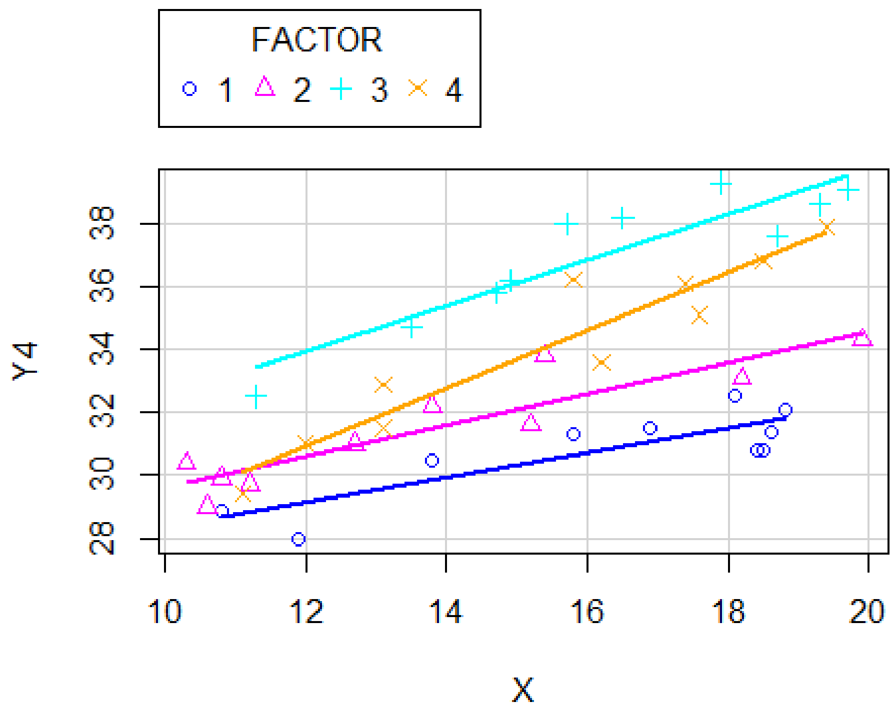
4. Y4

```
#Mean ans sd table
c<-tabular((FACTOR+1)~(X+Y4)*((n=1)+mean+sd),PRACOVAR)
pander(c, style='multiline', plan.ascii=F)
```

	X			Y4		
FACTOR	n	mean	sd	n	mean	sd

1	10	16.16	2.987	10	30.78	1.383
2	10	13.81	3.342	10	31.50	1.817
3	10	16.22	2.720	10	37.00	2.178
4	10	15.42	2.904	10	34.05	2.814
All	40	15.40	3.043	40	33.33	3.200

```
scatterplot(Y4~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

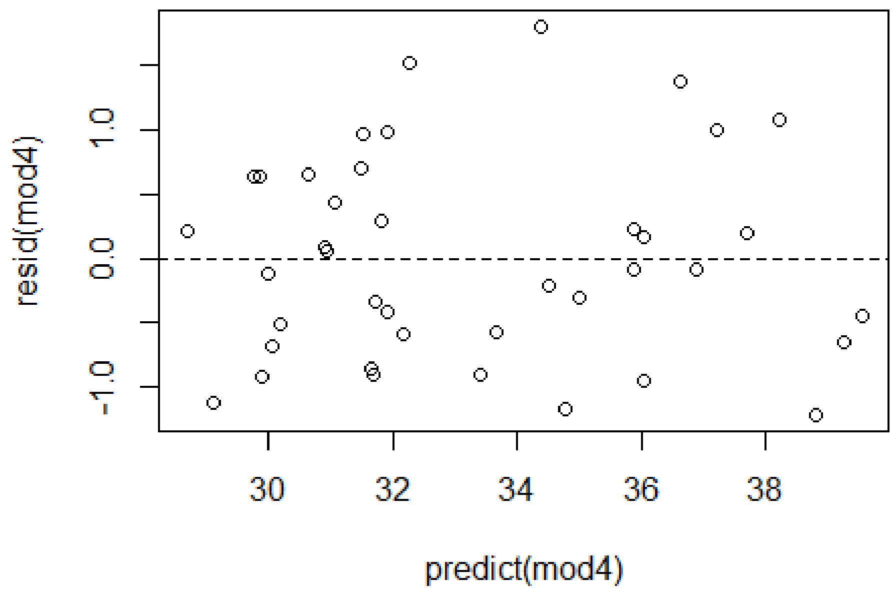
From the scatterplot we see a positive relation between the covariate and the dependent variable.

We could think of the response variable to be the salary of a person in a given company, the covariate to be the age and the factor the department he/she is working. With this, we would construct a model to predict the salary of a person in a company and compare it between departments.

To collect the data we would need to select randomly ten people from each department and write down their age and salary. This way we would have forty independent observations to construct our table and start analyzing.

4.1. Model with interaction

```
#Model with interaction
mod4<-lm(Y4~X*FACTOR, PRACOVAR)
summary(mod4)

##
## Call:
## lm(formula = Y4 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21336 -0.60456 -0.08415  0.63970  1.80088
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.47087    1.59531  15.339 2.65e-16 ***
## X            0.39042    0.09724   4.015 0.000336 ***
## FACTOR2      0.18655    2.01520   0.093 0.926820
## FACTOR3      0.66916    2.37093   0.282 0.779582
## FACTOR4     -4.58778    2.23612  -2.052 0.048461 *
## X:FACTOR2    0.10506    0.13041   0.806 0.426386
## X:FACTOR3    0.34078    0.14443   2.360 0.024565 *
## X:FACTOR4    0.52832    0.13950   3.787 0.000634 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8714 on 32 degrees of freedom
## Multiple R-squared:  0.9391, Adjusted R-squared:  0.9258
## F-statistic: 70.55 on 7 and 32 DF,  p-value: < 2.2e-16

anova(mod4)

## Analysis of Variance Table
##
## Response: Y4
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## X          1 158.875 158.875 209.2510 1.374e-15 ***
## FACTOR     3 202.692  67.564  88.9870 1.300e-15 ***
## X:FACTOR   3  13.384   4.461   5.8759  0.002583 **
## Residuals 32  24.296   0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(predict(mod4),resid(mod4))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

Both the covariate and the factor are significant to the response variable

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

The interaction term is significant, so we assume that the lines are not parallel.

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

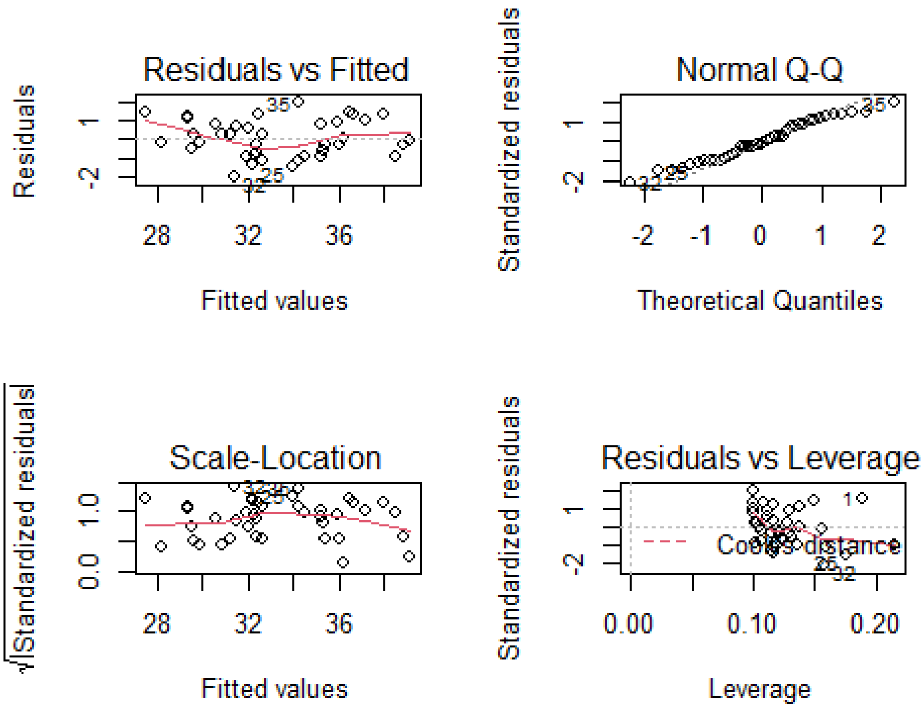
From the scatterplot there seems to be a significant effect

We could think of the response variable as the score of a test, the covariate of the number of hours slept the week before and the factor as the years at the university.

We should need access to the scores of the test and select randomly ten students from each year of university and check their numbers of hours slept, we should ask them, and check their test score.

4.2. Model without interaction

```
#model without interaction
mod42<-lm(Y4~X+FACTOR, PRACOVAR)
summary(mod42)

##
## Call:
## lm(formula = Y4 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.98345 -0.89728 -0.09927  0.83309  1.91544
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.80514    0.98856  21.046  < 2e-16 ***
## X            0.61726    0.05771  10.697 1.41e-12 ***
## FACTOR2      2.17055    0.48343   4.490 7.41e-05 ***
## FACTOR3      6.18296    0.46403  13.324 2.85e-15 ***
## FACTOR4      3.72677    0.46598   7.998 2.07e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.038 on 35 degrees of freedom
## Multiple R-squared:  0.9056, Adjusted R-squared:  0.8948
## F-statistic: 83.96 on 4 and 35 DF,  p-value: < 2.2e-16

anova(mod42)

## Analysis of Variance Table
##
## Response: Y4
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X          1 158.88 158.875 147.575 4.133e-14 ***
## FACTOR     3 202.69  67.564  62.758 3.671e-14 ***
## Residuals 35  37.68   1.077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

oldpar <- par( mfrow=c(2,2))
plot(mod42, ask=F)
```




```r
par(oldpar)
```

• Is there any significant variable?

Both the covariate and the factor are significant.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

We can assume normality and homoscedasticity.

• Which model do you think is more appropriate?

The interaction is significant so the interaction model should be better. The difference of R squares is 0.03, so we would stick with the interaction model.

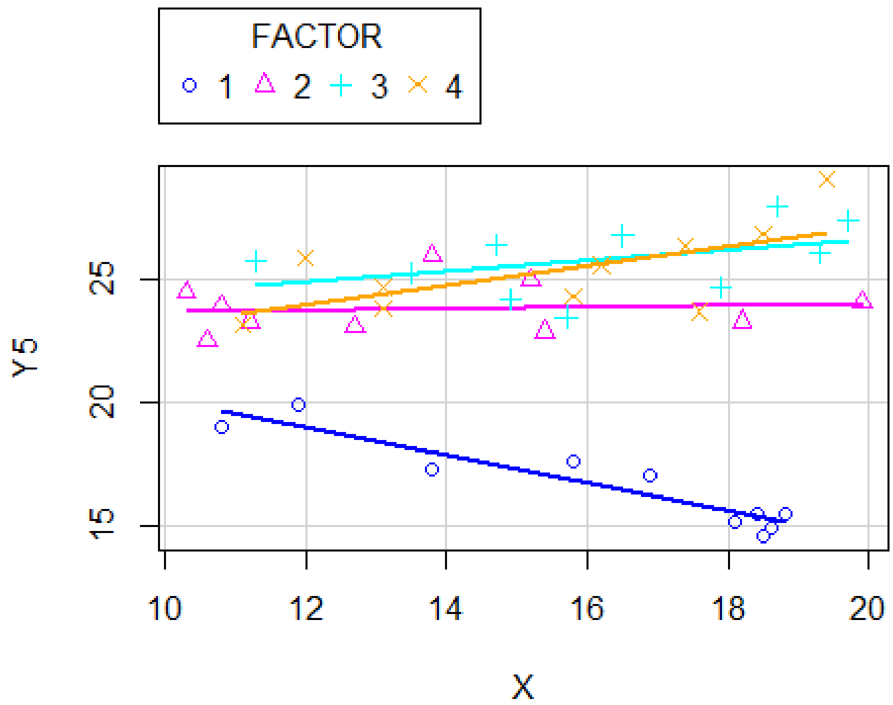
5. Y5

```r
#Mean ans sd table
d<-tabular((FACTOR+1)~(X+Y5)*((n=1)+mean+sd),PRACOVAR)
pander(d, style='multiline', plan.ascii=F)
```

	X			Y5		
FACTOR	n	mean	sd	n	mean	sd

1	10	16.16	2.987	10	16.64	1.821
2	10	13.81	3.342	10	23.86	1.071
3	10	16.22	2.720	10	25.82	1.420
4	10	15.42	2.904	10	25.36	1.801
All	40	15.40	3.043	40	22.92	4.033

```
scatterplot(Y5~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

It seems there is no effect from the factor or the covariate to the response variable, except from one level of the factor that seems to interact with the covariate.

We could think the response variable to be the symptom reduction of an illness, the covariate the amount in mg of medicine and the factor the type of medicine.

As a medical experimentation we need a control group and a group that we facilitate the medicine. We should pool all the volunteers and randomly select 1o for each medicament and amount of medicine to avoid confounding.

5.1. Model with interaction

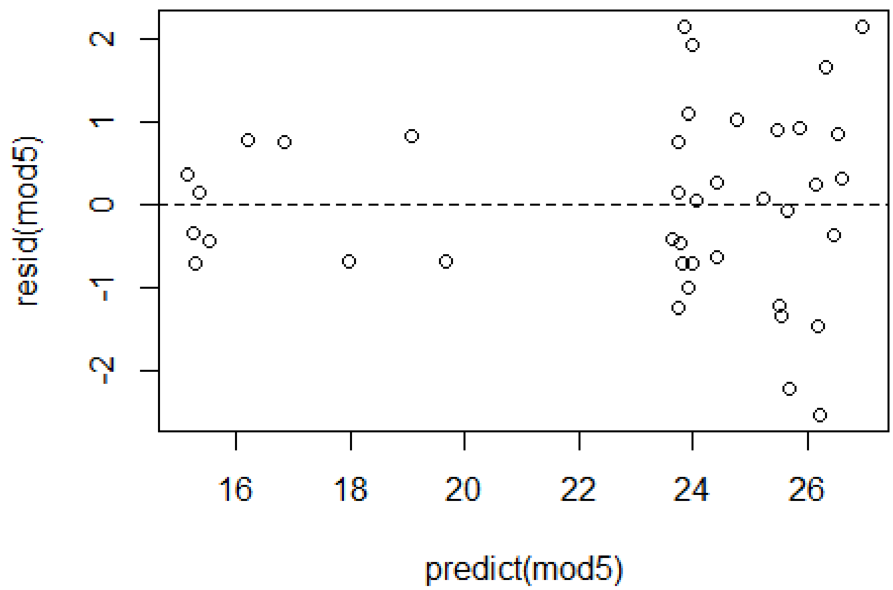
```
#Model with interaction
mod5<-lm(Y5~X*FACTOR, PRACOVAR)
summary(mod5)
```

```
##
## Call:
## lm(formula = Y5 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5360 -0.6964  0.0475  0.7945  2.1408
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.8453     2.1977  11.760 3.72e-13 ***
## X            -0.5696     0.1339  -4.253 0.000171 ***
## FACTOR2      -2.4446     2.7761  -0.881 0.385108
## FACTOR3      -3.4693     3.2661  -1.062 0.296096
## FACTOR4      -6.6813     3.0804  -2.169 0.037627 *
## X:FACTOR2     0.6029     0.1796   3.356 0.002050 **
## X:FACTOR3     0.7820     0.1990   3.930 0.000426 ***
## X:FACTOR4     0.9714     0.1922   5.055 1.69e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.2 on 32 degrees of freedom
## Multiple R-squared:  0.9273, Adjusted R-squared:  0.9114
## F-statistic: 58.33 on 7 and 32 DF,  p-value: < 2.2e-16
```

```
anova(mod5)
```

```
## Analysis of Variance Table
##
## Response: Y5
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## X          1   3.63   3.626   2.5162 0.1225139
## FACTOR     3 543.25 181.082 125.6766 < 2.2e-16 ***
## X:FACTOR   3  41.41  13.802   9.5789 0.0001156 ***
## Residuals 32  46.11   1.441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(predict(mod5),resid(mod5))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

In the ancova table we get that the only significant effects are the factor and the interaction, but the coefficients of the covariate seem to be different from 0. This might be because we only observe a significance of the covariate respect the response variable when checking for the first level of the factor, while the other levels seem to not have any effect.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

All the lines seem to be parallel, horizontal, except for one, level 1 of the factor. This is because this level is the only one that has an effect over the response variable.

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

The residual plot is not very good for our model, we cannot assume homoscedasticity and hence our model is not valid.

5.2. Model without interaction

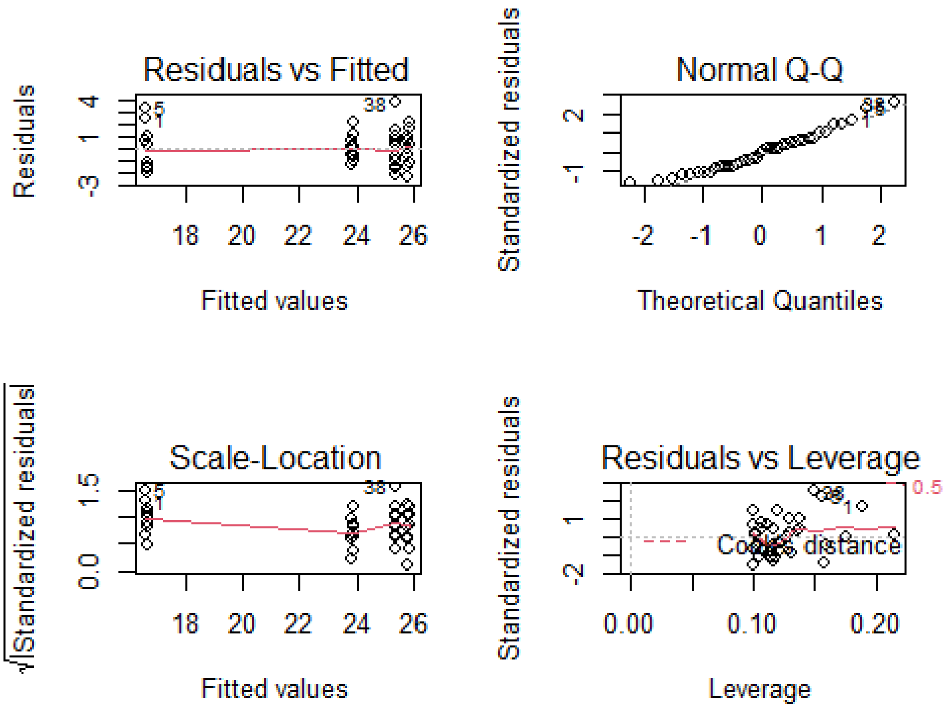
```r
#model without interaction
mod52<-lm(Y5~X+FACTOR, PRACOVAR)
summary(mod52)
```

```
##
## Call:
## lm(formula = Y5 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3164 -1.1561  0.0373  0.9664  3.7126
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.528836   1.506544  10.971 7.10e-13 ***
## X            0.006879   0.087942   0.078    0.938
## FACTOR2      7.236166   0.736738   9.822 1.35e-11 ***
## FACTOR3      9.179587   0.707178  12.981 6.12e-15 ***
## FACTOR4      8.725090   0.710146  12.286 2.99e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.581 on 35 degrees of freedom
## Multiple R-squared:  0.8621, Adjusted R-squared:  0.8463
## F-statistic: 54.68 on 4 and 35 DF,  p-value: 1.418e-14
```

```r
anova(mod52)
```

```
## Analysis of Variance Table
##
## Response: Y5
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X          1   3.63   3.626   1.450    0.2366
## FACTOR     3 543.25 181.082  72.422 4.377e-15 ***
## Residuals 35  87.51   2.500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
oldpar <- par( mfrow=c(2,2))
plot(mod52, ask=F)
```




```
par(oldpar)
```

• Is there any significant variable?

In this model, only the factor has a significant effect, both in the table of coefficients and the ancova table.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

We can assume homoscedasticity and normality of the residuals.

• Which model do you think is more appropriate?

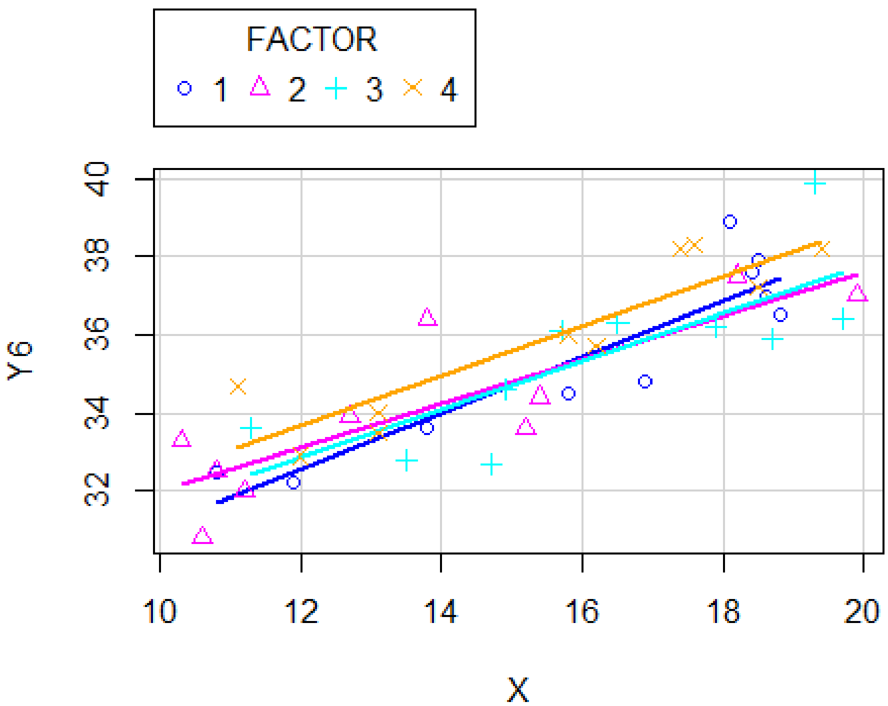
The model with interaction seems to explain better, the R squared is quite higher compared to the additive model, but the analysis of residuals is not good for it. Hence, we would stick with the additive model that yields good results in the analysis of residuals.

6. Y6

```
#Mean ans sd table
e<-tabular((FACTOR+1)~(X+Y6)*((n=1)+mean+sd),PRACOVAR)
pander(e, style='multiline', plan.ascii=F)
```

FACTOR	X n	mean	sd	Y6 n	mean	sd
1	10	16.16	2.987	10	35.55	2.356
2	10	13.81	3.342	10	34.14	2.214
3	10	16.22	2.720	10	35.45	2.143
4	10	15.42	2.904	10	35.87	2.052
All	40	15.40	3.043	40	35.25	2.212

```
scatterplot(Y6~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

From the scatterplot we can assume a positive relation between variables.

We could think of the response variable to be the shop sales of a store, the covariate the number of people who enters the store and the factor the type of weather, rainy, sunny, cloudy and snowy. We want to understand if these variables explain the variability in our sales.

To collect the data we should need access to our sales, and measure the number of visitors our store gets in one day, as well we should write down the type of weather it is making in that day.

6.1. Model with interaction

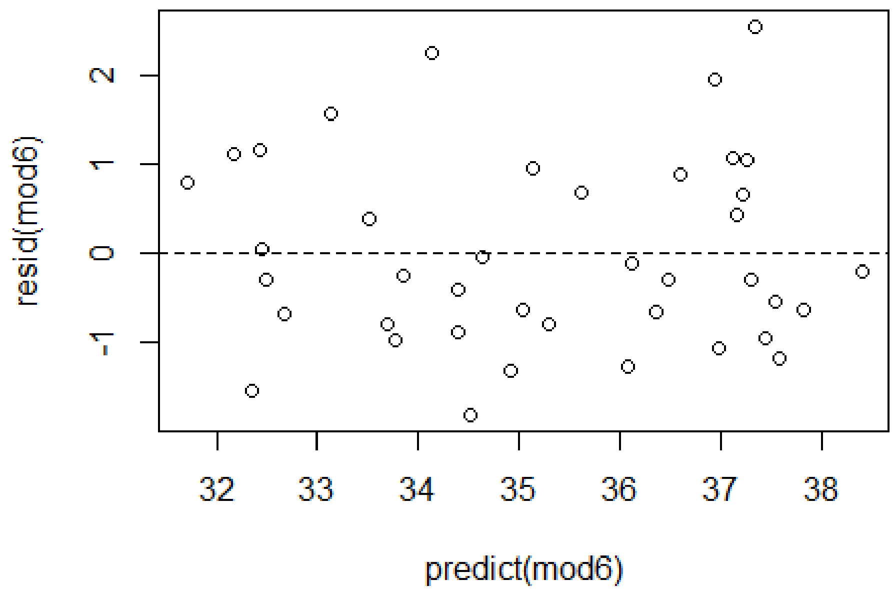
```r
#Model with interaction
mod6<-lm(Y6~X*FACTOR, PRACOVAR)
summary(mod6)
```

```
##
## Call:
## lm(formula = Y6 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8162 -0.7924 -0.2685  0.8258  2.5579
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.94275    2.16084  11.080 1.74e-12 ***
## X            0.71827    0.13171   5.454 5.31e-06 ***
## FACTOR2      2.46934    2.72960   0.905    0.372
## FACTOR3      1.54273    3.21142   0.480    0.634
## FACTOR4      2.12240    3.02883   0.701    0.489
## X:FACTOR2   -0.15868    0.17664  -0.898    0.376
## X:FACTOR3   -0.10394    0.19563  -0.531    0.599
## X:FACTOR4   -0.08242    0.18896  -0.436    0.666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.18 on 32 degrees of freedom
## Multiple R-squared:  0.7663, Adjusted R-squared:  0.7152
## F-statistic: 14.99 on 7 and 32 DF,  p-value: 1.694e-08
```

```r
anova(mod6)
```

```
## Analysis of Variance Table
##
## Response: Y6
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## X          1 139.932 139.932 100.4547 2.145e-11 ***
## FACTOR     3   5.110   1.703   1.2227    0.3174
## X:FACTOR   3   1.142   0.381   0.2733    0.8442
## Residuals 32  44.576   1.393
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(predict(mod6),resid(mod6))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

Only the covariate has a significant effect on the response variable.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

The interaction term is not significant, we assume the lines are parallel.

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

We can assume homoscedasticity.

6.2. Model without interaction

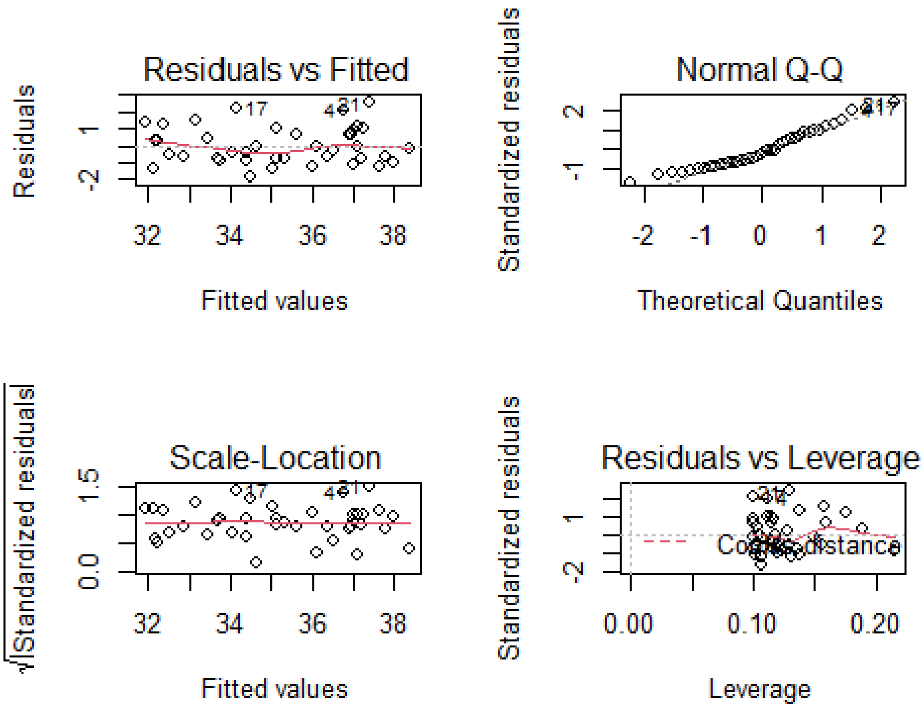
```
#model without interaction
mod62<-lm(Y6~X+FACTOR, PRACOVAR)
summary(mod62)
```

```
##
## Call:
## lm(formula = Y6 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7952 -0.8222 -0.2377  0.7256  2.5152
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 25.39870    1.08890  23.325  < 2e-16 ***
## X            0.62817    0.06356   9.883 1.15e-11 ***
## FACTOR2      0.06621    0.53250   0.124    0.902
## FACTOR3     -0.13769    0.51113  -0.269    0.789
## FACTOR4      0.78485    0.51328   1.529    0.135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.143 on 35 degrees of freedom
## Multiple R-squared:  0.7603, Adjusted R-squared:  0.7329
## F-statistic: 27.76 on 4 and 35 DF,  p-value: 1.988e-10
```

```
anova(mod62)
```

```
## Analysis of Variance Table
##
## Response: Y6
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## X          1 139.932 139.932 107.1277 3.418e-12 ***
## FACTOR     3   5.110   1.703   1.3039    0.2886
## Residuals 35  45.718   1.306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
oldpar <- par( mfrow=c(2,2))
plot(mod62, ask=F)
```




```
par(oldpar)
```

• Is there any significant variable?

Only the covariate is significant.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

We can assume homoscedasticity and normality of the residuals.

• Which model do you think is more appropriate?

As the interaction term is not significant, the additive model performs better in this case. The R squared is almost similar but with the additive model we use one parameter less.

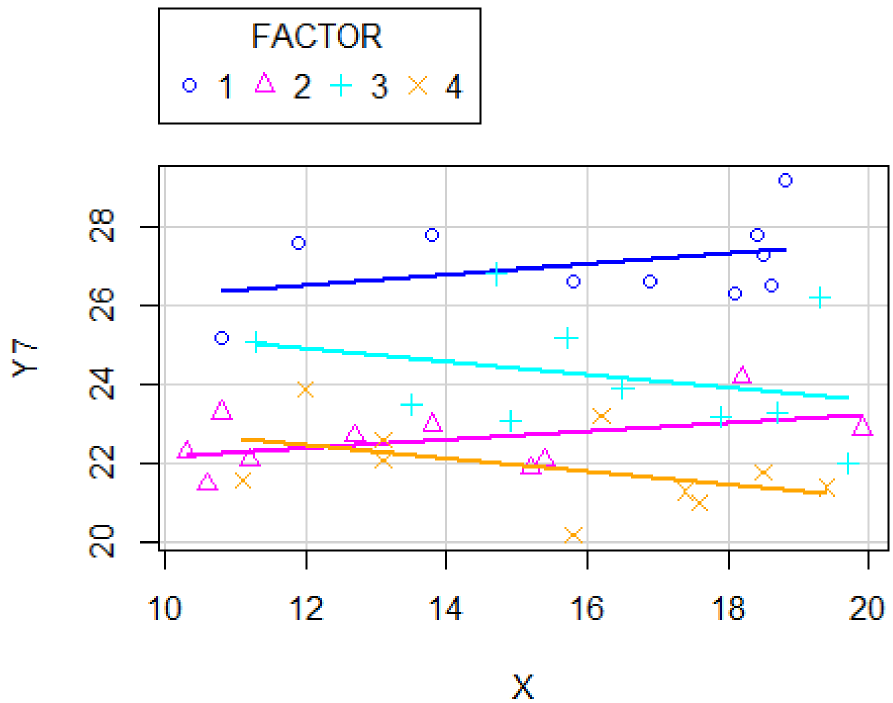
7. Y7

```
#Mean ans sd table
f<-tabular((FACTOR+1)~(X+Y7)*((n=1)+mean+sd),PRACOVAR)
pander(f, style='multiline', plan.ascii=F)
```

	X			Y7		
FACTOR	n	mean	sd	n	mean	sd

1	10	16.16	2.987	10	27.09	1.0949
2	10	13.81	3.342	10	22.60	0.7888
3	10	16.22	2.720	10	24.23	1.5276
4	10	15.42	2.904	10	21.91	1.0867
All	40	15.40	3.043	40	23.96	2.3052

```
scatterplot(Y7~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

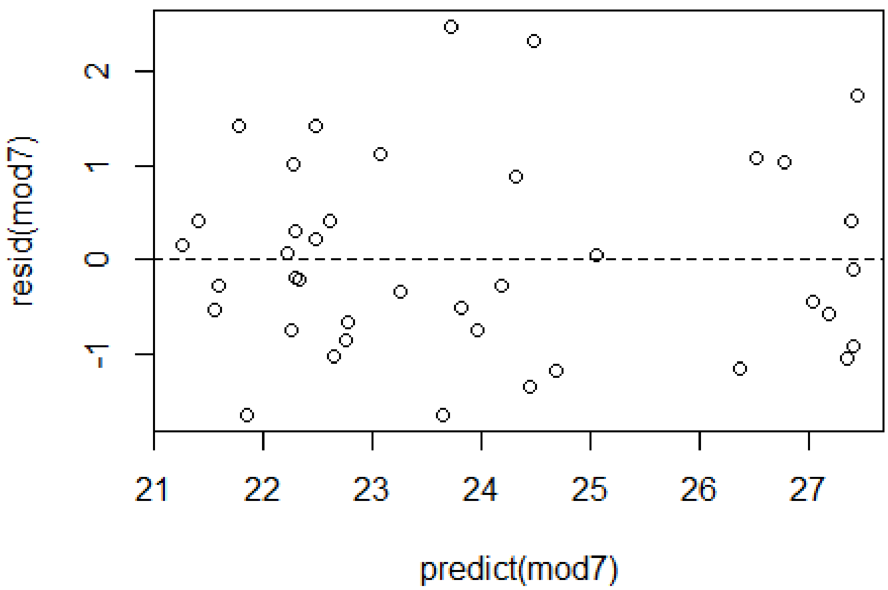
From the scatterplot seems nor the factor or the covariate has an effect on the response variable.

We could think of the response variable to be the number of car accidents in a road, the covariate the age of the drivers of that road and the factor the type of his/her vehicle, car, moto, van or wagon.

We should record the number of accident in that road over a period of fixed time, e.g. a year. And from those accidents check the type of vehicle and the age of the driver. To get a balanced design is difficult this way but is possible that we get 10 observations of each type of vehicle.

7.1. Model with interaction

```
#Model with interaction
mod7<-lm(Y7~X*FACTOR, PRACOVAR)
summary(mod7)

##
## Call:
## lm(formula = Y7 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6531 -0.7528 -0.2098  0.5280  2.4806
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.92149    2.08407  11.958  2.4e-13 ***
## X            0.13419    0.12703   1.056    0.299
## FACTOR2     -3.79796    2.63262  -1.443    0.159
## FACTOR3      1.99723    3.09732   0.645    0.524
## FACTOR4     -0.42808    2.92122  -0.147    0.884
## X:FACTOR2   -0.02728    0.17036  -0.160    0.874
## X:FACTOR3   -0.29996    0.18868  -1.590    0.122
## X:FACTOR4   -0.30173    0.18224  -1.656    0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.138 on 32 degrees of freedom
## Multiple R-squared:  0.7999, Adjusted R-squared:  0.7562
## F-statistic: 18.28 on 7 and 32 DF,  p-value: 1.562e-09

anova(mod7)

## Analysis of Variance Table
##
## Response: Y7
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X          1   5.607   5.607  4.3269   0.04561 *
## FACTOR     3 153.627  51.209 39.5203 7.162e-11 ***
## X:FACTOR   3   6.539   2.180  1.6822   0.19044
## Residuals 32  41.465   1.296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(predict(mod7),resid(mod7))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

In the ancova table the factor seems to be significant, but in the coefficient table there is no significant variable. This is similar to what happened in the las dataset. It seems only one level of the factor has a significant effect.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

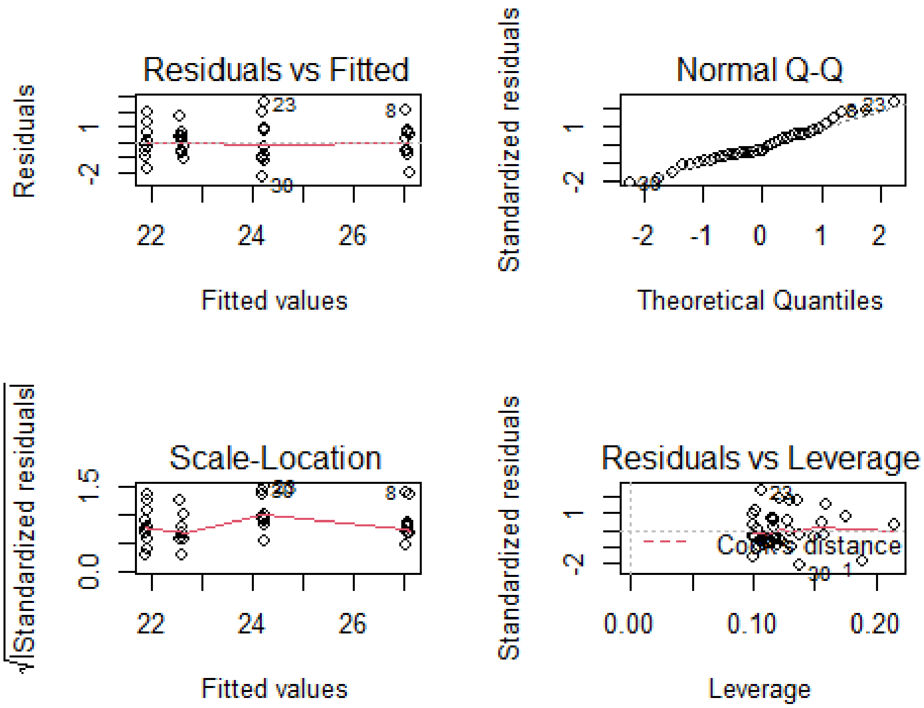
The interaction term is not significant, so we assume that the lines are parallel.

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

We can assume homoscedasticity.

7.2. Model without interaction

```r
#model without interaction
mod72<-lm(Y7~X+FACTOR, PRACOVAR)
summary(mod72)

##
## Call:
## lm(formula = Y7 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2061 -0.7050 -0.3261  0.6829  2.5596
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.201064   1.115793  24.378  < 2e-16 ***
## X           -0.006873   0.065132  -0.106    0.917
## FACTOR2     -4.506151   0.545651  -8.258 9.83e-10 ***
## FACTOR3     -2.859588   0.523758  -5.460 3.98e-06 ***
## FACTOR4     -5.185086   0.525956  -9.858 1.23e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.171 on 35 degrees of freedom
## Multiple R-squared:  0.7684, Adjusted R-squared:  0.7419
## F-statistic: 29.02 on 4 and 35 DF,  p-value: 1.106e-10

anova(mod72)

## Analysis of Variance Table
##
## Response: Y7
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X          1   5.607   5.607  4.0878   0.05089 .
## FACTOR     3 153.627  51.209 37.3371 5.251e-11 ***
## Residuals 35  48.004   1.372
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

oldpar <- par( mfrow=c(2,2))
plot(mod72, ask=F)
```




```
par(oldpar)
```

• Is there any significant variable?

The factor variable has a significant effect over the response variable.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

We can assume homoscedasticity and normality.

• Which model do you think is more appropriate?

The interaction coefficient is not significant, and we obtain a similar R squared, we it is better to use the additive model as we have less parameters.

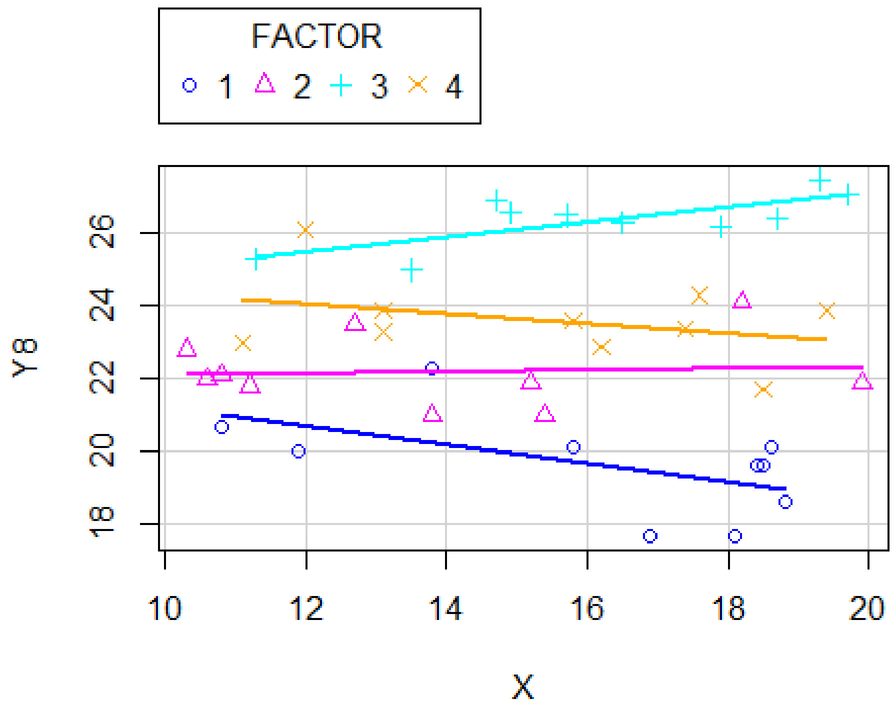
8. Y8

```
#Mean ans sd table
j<-tabular((FACTOR+1)~(X+Y8)*((n=1)+mean+sd),PRACOVAR)
pander(j, style='multiline', plan.ascii=F)
```

	X			Y8		
FACTOR	n	mean	sd	n	mean	sd

1	10	16.16	2.987	10	19.64	1.3890
2	10	13.81	3.342	10	22.21	0.9960
3	10	16.22	2.720	10	26.38	0.7613
4	10	15.42	2.904	10	23.61	1.1308
All	40	15.40	3.043	40	22.96	2.6797

```
scatterplot(Y8~X|FACTOR,smooth=F,dat=PRACOVAR)
```



a) Describe a real situation that agrees with the experimental data. Explain how do you propose to collect the data.

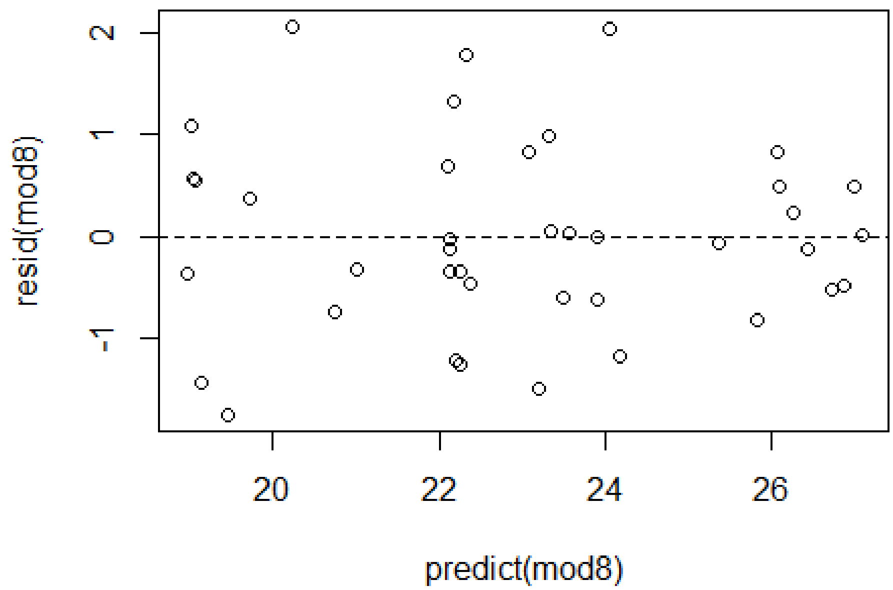
From the scatterplot we see how two levels of the factor interact in a different way with the response variable. The third level has a positive relation, and the first level has a negative relation.

We can think of the response variable to be the score in a weight in competition, the covariate the age, and the factor the race, white, afroamerican, latino and Asian. We might want to explain if some races their strength starts to decay earlier or later.

In order to collect the data we should access to the event records and select randomly ten people from each race and measure their weight-in and their age.

8.1. Model with interaction

```
#Model with interaction
mod8<-lm(Y8~X*FACTOR, PRACOVAR)
summary(mod8)

##
## Call:
## lm(formula = Y8 ~ X * FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.74958 -0.54418 -0.05283  0.54283  2.05272
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.7983     1.8740  12.699 4.86e-14 ***
## X            -0.2573     0.1142  -2.253   0.0313 *
## FACTOR2      -1.9508     2.3672  -0.824   0.4160
## FACTOR3      -0.7327     2.7851  -0.263   0.7942
## FACTOR4       1.8504     2.6267   0.704   0.4862
## X:FACTOR2     0.2836     0.1532   1.851   0.0734 .
## X:FACTOR3     0.4617     0.1697   2.721   0.0104 *
## X:FACTOR4     0.1251     0.1639   0.763   0.4508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.024 on 32 degrees of freedom
## Multiple R-squared:  0.8803, Adjusted R-squared:  0.8541
## F-statistic: 33.62 on 7 and 32 DF,  p-value: 5.26e-13

anova(mod8)

## Analysis of Variance Table
##
## Response: Y8
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## X          1   0.000   0.000  0.0005  0.98315
## FACTOR     3 237.684  79.228 75.6234 1.29e-14 ***
## X:FACTOR   3   8.847   2.949  2.8147  0.05486 .
## Residuals 32  33.525   1.048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(predict(mod8),resid(mod8))
abline(h=0,lty=2)
```



b) Propose a factorial ANCOVA model (ANCOVA with interaction), and after fitting it, answer the following questions:

• Is there any significant variable?

Only the factor has a significant effect.

• Do the factor and the continuous covariate interact? (Are the lines parallel?)

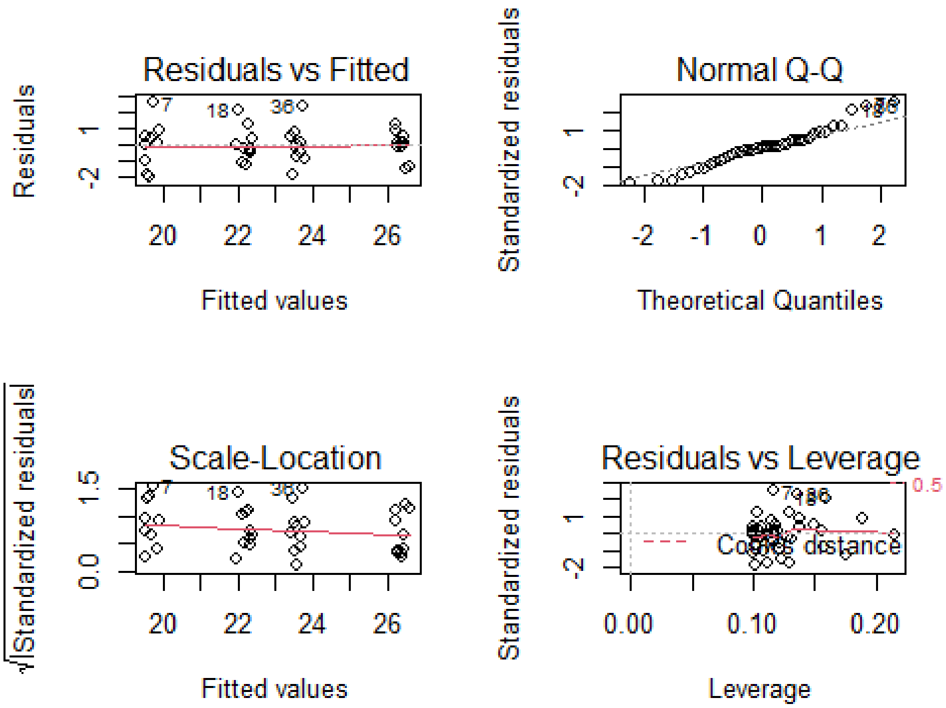
The interaction term is not significant, although the p-value is very small. At the same time, we do not see parallelism in the scatterplot, but there is no significant differentiation, so we assume they are parallel

• Is your model appropriate? Decide based on the plot of the residuals as a function of the predictions.

It seems the variability of the residuals is greater in the middle, but I think is fair to assume homoscedasticity.

8.2. Model without interaction

```r
#model without interaction
mod82<-lm(Y8~X+FACTOR, PRACOVAR)
summary(mod82)

##
## Call:
## lm(formula = Y8 ~ X + FACTOR, data = PRACOVAR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.90692 -0.56379  0.03356  0.45602  2.55450
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.36242    1.04830  19.424  < 2e-16 ***
## X           -0.04470    0.06119  -0.731     0.47
## FACTOR2      2.46495    0.51264   4.808 2.86e-05 ***
## FACTOR3      6.74268    0.49208  13.703 1.25e-15 ***
## FACTOR4      3.93692    0.49414   7.967 2.26e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.1 on 35 degrees of freedom
## Multiple R-squared:  0.8487, Adjusted R-squared:  0.8314
## F-statistic: 49.08 on 4 and 35 DF,  p-value: 7.033e-14

anova(mod82)

## Analysis of Variance Table
##
## Response: Y8
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## X          1   0.000   0.000  0.0004   0.9843
## FACTOR     3 237.684  79.228 65.4438 1.98e-14 ***
## Residuals 35  42.372   1.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

oldpar <- par( mfrow=c(2,2))
plot(mod82, ask=F)
```




```r
par(oldpar)
```

• Is there any significant variable?

The factor has a significant effect.

• Is the model appropriate? Decide based on the residual versus predicted plot (perform this plot distinguishing and without distinguishing the factor levels).

It seems normality might be a little bit out of terms on the extremes, although homoscedasticity could be assumed.

• Which model do you think is more appropriate?

Although the interaction term is not significant, we have seen one of the levels interaction is. The difference of R squared is about 0.04 between models, so I would stick with the model with interaction.