TP3

Guillaume, Julio, Pierre-Yves

24 avril 2019

TP3

#Exercice 1

##1/1ecture du fichier

tab=read.csv("Advertising.csv",header=T,row.names=1)  
attach(tab)  
summary(tab)

## TV Radio Newspaper Sales   
## Min. : 0.70 Min. : 0.000 Min. : 0.30 Min. : 1.60   
## 1st Qu.: 74.38 1st Qu.: 9.975 1st Qu.: 12.75 1st Qu.:10.38   
## Median :149.75 Median :22.900 Median : 25.75 Median :12.90   
## Mean :147.04 Mean :23.264 Mean : 30.55 Mean :14.02   
## 3rd Qu.:218.82 3rd Qu.:36.525 3rd Qu.: 45.10 3rd Qu.:17.40   
## Max. :296.40 Max. :49.600 Max. :114.00 Max. :27.00

pub=data.frame(data=tab)  
summary(pub)

## data.TV data.Radio data.Newspaper data.Sales   
## Min. : 0.70 Min. : 0.000 Min. : 0.30 Min. : 1.60   
## 1st Qu.: 74.38 1st Qu.: 9.975 1st Qu.: 12.75 1st Qu.:10.38   
## Median :149.75 Median :22.900 Median : 25.75 Median :12.90   
## Mean :147.04 Mean :23.264 Mean : 30.55 Mean :14.02   
## 3rd Qu.:218.82 3rd Qu.:36.525 3rd Qu.: 45.10 3rd Qu.:17.40   
## Max. :296.40 Max. :49.600 Max. :114.00 Max. :27.00

##2/matrice de correlation

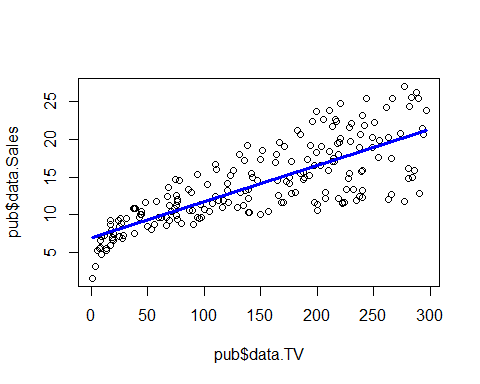
mcor=cor(pub)  
mcor

## data.TV data.Radio data.Newspaper data.Sales  
## data.TV 1.00000000 0.05480866 0.05664787 0.7822244  
## data.Radio 0.05480866 1.00000000 0.35410375 0.5762226  
## data.Newspaper 0.05664787 0.35410375 1.00000000 0.2282990  
## data.Sales 0.78222442 0.57622257 0.22829903 1.0000000

On remarque que : > les ventes et la pub sur la TV sont très corrélés (0,78); > les ventes et la pub sur la radio sont plutôt corrélés (0,57); > les ventes et la pub sur les journaux sont moins bien corrélés (0,22); > la corrélation entre la pub sur la radio et les journaux est assez peu corrélé (0,35).

##3/Régression linéaire simple

plot(pub$data.TV,pub$data.Sales)  
regTV=lm(data.Sales~data.TV,data=pub)  
lines(c(min(pub$data.TV),max(pub$data.TV)),c(min(regTV$fitted.values),max(regTV$fitted.values)),col="blue",lwd=3)



#plot(regTV)  
summary(regTV)#plot(X=data.TV,Y=data.Sales,data=pub)

##   
## Call:  
## lm(formula = data.Sales ~ data.TV, data = pub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.3860 -1.9545 -0.1913 2.0671 7.2124   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.032594 0.457843 15.36 <2e-16 \*\*\*  
## data.TV 0.047537 0.002691 17.67 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.259 on 198 degrees of freedom  
## Multiple R-squared: 0.6119, Adjusted R-squared: 0.6099   
## F-statistic: 312.1 on 1 and 198 DF, p-value: < 2.2e-16

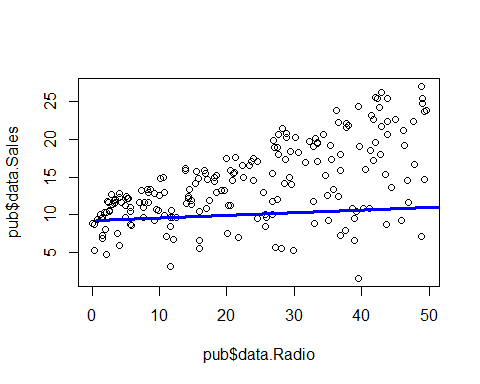
anova(regTV)

## Analysis of Variance Table  
##   
## Response: data.Sales  
## Df Sum Sq Mean Sq F value Pr(>F)   
## data.TV 1 3314.6 3314.6 312.14 < 2.2e-16 \*\*\*  
## Residuals 198 2102.5 10.6   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Le modèle n’est pas ajusté aux données (R²=0,61). La pub sur la TV a un effet sur la vente (Fvalue=312).

##4/

plot(pub$data.Radio,pub$data.Sales)  
regRadio=lm(data.Sales~data.Radio,data=pub)  
lines(c(min(pub$data.TV),max(pub$data.TV)),c(min(regRadio$fitted.values),max(regRadio$fitted.values)),col="blue",lwd=3)



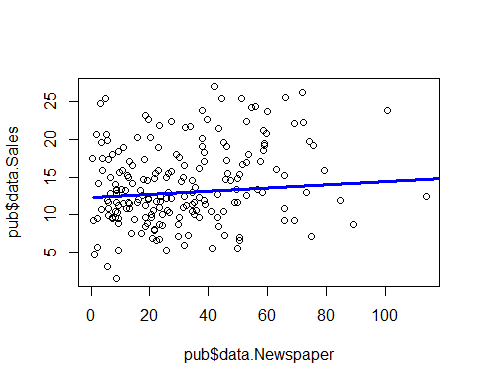
#plot(regTV)  
summary(regRadio)#plot(X=data.TV,Y=data.Sales,data=pub)

##   
## Call:  
## lm(formula = data.Sales ~ data.Radio, data = pub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.7305 -2.1324 0.7707 2.7775 8.1810   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.31164 0.56290 16.542 <2e-16 \*\*\*  
## data.Radio 0.20250 0.02041 9.921 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.275 on 198 degrees of freedom  
## Multiple R-squared: 0.332, Adjusted R-squared: 0.3287   
## F-statistic: 98.42 on 1 and 198 DF, p-value: < 2.2e-16

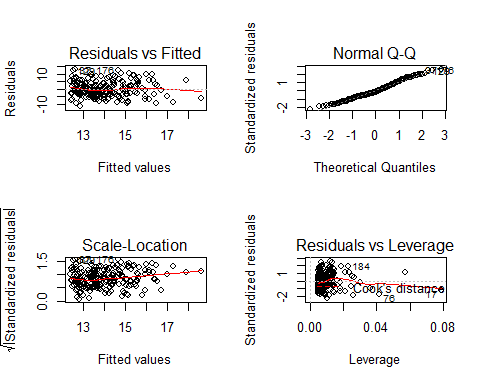
anova(regRadio)

## Analysis of Variance Table  
##   
## Response: data.Sales  
## Df Sum Sq Mean Sq F value Pr(>F)   
## data.Radio 1 1798.7 1798.67 98.422 < 2.2e-16 \*\*\*  
## Residuals 198 3618.5 18.28   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(pub$data.Newspaper,pub$data.Sales)  
regNP=lm(data.Sales~data.Newspaper,data=pub)  
lines(c(min(pub$data.TV),max(pub$data.TV)),c(min(regNP$fitted.values),max(regNP$fitted.values)),col="blue",lwd=3)



par(mfrow=c(2,2))  
plot(regNP)



summary(regNP)#plot(X=data.TV,Y=data.Sales,data=pub)

##   
## Call:  
## lm(formula = data.Sales ~ data.Newspaper, data = pub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.2272 -3.3873 -0.8392 3.5059 12.7751   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12.35141 0.62142 19.88 < 2e-16 \*\*\*  
## data.Newspaper 0.05469 0.01658 3.30 0.00115 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.092 on 198 degrees of freedom  
## Multiple R-squared: 0.05212, Adjusted R-squared: 0.04733   
## F-statistic: 10.89 on 1 and 198 DF, p-value: 0.001148

anova(regNP)

## Analysis of Variance Table  
##   
## Response: data.Sales  
## Df Sum Sq Mean Sq F value Pr(>F)   
## data.Newspaper 1 282.3 282.344 10.887 0.001148 \*\*  
## Residuals 198 5134.8 25.933   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##5/regression multiple

reg=lm(data.Sales~data.TV+data.Radio+data.Newspaper,data=pub)  
summary(reg)

##   
## Call:  
## lm(formula = data.Sales ~ data.TV + data.Radio + data.Newspaper,   
## data = pub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.8277 -0.8908 0.2418 1.1893 2.8292   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.938889 0.311908 9.422 <2e-16 \*\*\*  
## data.TV 0.045765 0.001395 32.809 <2e-16 \*\*\*  
## data.Radio 0.188530 0.008611 21.893 <2e-16 \*\*\*  
## data.Newspaper -0.001037 0.005871 -0.177 0.86   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.686 on 196 degrees of freedom  
## Multiple R-squared: 0.8972, Adjusted R-squared: 0.8956   
## F-statistic: 570.3 on 3 and 196 DF, p-value: < 2.2e-16

reg$coefficients

## (Intercept) data.TV data.Radio data.Newspaper   
## 2.938889369 0.045764645 0.188530017 -0.001037493

Concrètement, représente le nombre de ventes sans pub. Le poucentage de la variabilité des ventes est expliqué ici est de 89% (R²=0,89).

##6 et 7/régression multiple

anova(reg)

## Analysis of Variance Table  
##   
## Response: data.Sales  
## Df Sum Sq Mean Sq F value Pr(>F)   
## data.TV 1 3314.6 3314.6 1166.7308 <2e-16 \*\*\*  
## data.Radio 1 1545.6 1545.6 544.0501 <2e-16 \*\*\*  
## data.Newspaper 1 0.1 0.1 0.0312 0.8599   
## Residuals 196 556.8 2.8   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Au moins une des variables ici TV et Radio sont significatives.

8/

reg=lm(data.Sales~data.TV+data.Radio,data=pub)  
summary(reg)

##   
## Call:  
## lm(formula = data.Sales ~ data.TV + data.Radio, data = pub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.7977 -0.8752 0.2422 1.1708 2.8328   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.92110 0.29449 9.919 <2e-16 \*\*\*  
## data.TV 0.04575 0.00139 32.909 <2e-16 \*\*\*  
## data.Radio 0.18799 0.00804 23.382 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.681 on 197 degrees of freedom  
## Multiple R-squared: 0.8972, Adjusted R-squared: 0.8962   
## F-statistic: 859.6 on 2 and 197 DF, p-value: < 2.2e-16

reg$coefficients

## (Intercept) data.TV data.Radio   
## 2.92109991 0.04575482 0.18799423

0,19x10=1,9 Augmentation de 100 objets vendus.

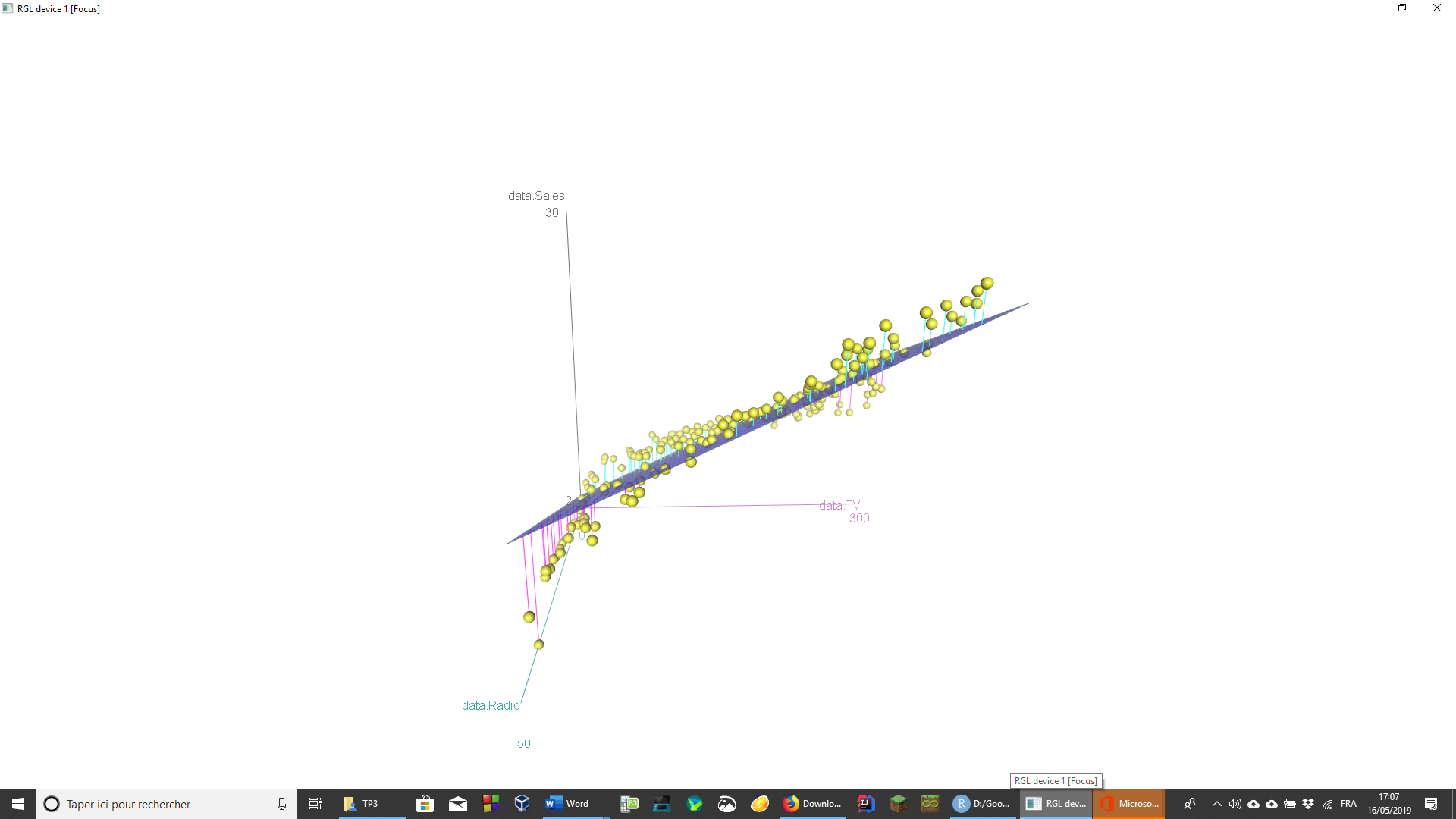
9/

library(rgl)  
library(car)

## Loading required package: carData

scatter3d(data.Sales~data.TV+data.Radio, data=pub)

## Loading required namespace: mgcv



10/

x0=data.frame(data.TV=100,data.Radio=20)  
borne=predict(reg,x0,interval='prediction')  
borne

## fit lwr upr  
## 1 11.25647 7.929616 14.58332

On s’attend à vendre environ 11256 produits. L’intervalle est entre 7929 produits et 14583 produits.

#Exercice 3:

1/

tab=read.table("cafe.txt",header=T)  
#summary(tab)  
cafe=data.frame(tab)  
summary(cafe)

## origine perte lumin xa xb   
## A:50 Min. :11.87 Min. :15.06 Min. : 5.730 Min. : 6.42   
## B:42 1st Qu.:14.72 1st Qu.:19.78 1st Qu.: 8.620 1st Qu.:12.18   
## C:42 Median :16.41 Median :23.41 Median : 9.980 Median :16.30   
## D:13 Mean :16.55 Mean :24.03 Mean : 9.938 Mean :16.86   
## E:15 3rd Qu.:18.16 3rd Qu.:27.51 3rd Qu.:11.460 3rd Qu.:21.03   
## F:14 Max. :22.81 Max. :41.15 Max. :13.130 Max. :30.57   
## G:13   
## xy xgn   
## Min. : 1.920 Min. :1.660   
## 1st Qu.: 2.890 1st Qu.:2.420   
## Median : 3.790 Median :3.140   
## Mean : 4.257 Mean :3.466   
## 3rd Qu.: 5.150 3rd Qu.:4.130   
## Max. :11.970 Max. :9.120   
##

2/ (a)

modele1=lm(perte~lumin+xa+xb+xy+xgn,data=cafe)  
summary(modele1)

##   
## Call:  
## lm(formula = perte ~ lumin + xa + xb + xy + xgn, data = cafe)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.67125 -0.54267 0.00579 0.48960 2.49762   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.8410 2.8769 10.720 < 2e-16 \*\*\*  
## lumin -0.1305 0.2634 -0.496 0.621   
## xa -1.3470 0.1723 -7.819 4.06e-13 \*\*\*  
## xb 0.2417 0.2117 1.142 0.255   
## xy -0.1623 0.1976 -0.821 0.413   
## xgn -0.3323 0.2216 -1.499 0.136   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8083 on 183 degrees of freedom  
## Multiple R-squared: 0.8919, Adjusted R-squared: 0.889   
## F-statistic: 302 on 5 and 183 DF, p-value: < 2.2e-16

modele1$coefficients

## (Intercept) lumin xa xb xy xgn   
## 30.8410003 -0.1305491 -1.3469711 0.2416761 -0.1623071 -0.3322836

anova(modele1)

## Analysis of Variance Table  
##   
## Response: perte  
## Df Sum Sq Mean Sq F value Pr(>F)   
## lumin 1 837.17 837.17 1281.2139 < 2e-16 \*\*\*  
## xa 1 145.14 145.14 222.1310 < 2e-16 \*\*\*  
## xb 1 2.90 2.90 4.4394 0.03648 \*   
## xy 1 0.01 0.01 0.0077 0.93005   
## xgn 1 1.47 1.47 2.2475 0.13555   
## Residuals 183 119.58 0.65   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Les variables qui ont un effet significatif sont lumin, xa et xb.

summary(modele1)

##   
## Call:  
## lm(formula = perte ~ lumin + xa + xb + xy + xgn, data = cafe)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.67125 -0.54267 0.00579 0.48960 2.49762   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.8410 2.8769 10.720 < 2e-16 \*\*\*  
## lumin -0.1305 0.2634 -0.496 0.621   
## xa -1.3470 0.1723 -7.819 4.06e-13 \*\*\*  
## xb 0.2417 0.2117 1.142 0.255   
## xy -0.1623 0.1976 -0.821 0.413   
## xgn -0.3323 0.2216 -1.499 0.136   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8083 on 183 degrees of freedom  
## Multiple R-squared: 0.8919, Adjusted R-squared: 0.889   
## F-statistic: 302 on 5 and 183 DF, p-value: < 2.2e-16

Donc pas significativement différent de (car pour xb : ).

3/

cor(cafe[,-1])

## perte lumin xa xb xy xgn  
## perte 1.0000000 -0.8699159 -0.9340574 -0.9058234 -0.7980422 -0.8035131  
## lumin -0.8699159 1.0000000 0.8643060 0.9895394 0.9717936 0.9661396  
## xa -0.9340574 0.8643060 1.0000000 0.9218400 0.7702314 0.7699353  
## xb -0.9058234 0.9895394 0.9218400 1.0000000 0.9405286 0.9350522  
## xy -0.7980422 0.9717936 0.7702314 0.9405286 1.0000000 0.9405116  
## xgn -0.8035131 0.9661396 0.7699353 0.9350522 0.9405116 1.0000000

On a xgn,xy et xa qui sont très corrélés avec lumin.

4/ (a)

modele2=step(modele1, direction="backward")

## Start: AIC=-74.52  
## perte ~ lumin + xa + xb + xy + xgn  
##   
## Df Sum of Sq RSS AIC  
## - lumin 1 0.161 119.74 -76.271  
## - xy 1 0.441 120.02 -75.829  
## - xb 1 0.851 120.43 -75.184  
## <none> 119.58 -74.525  
## - xgn 1 1.469 121.04 -74.218  
## - xa 1 39.953 159.53 -22.041  
##   
## Step: AIC=-76.27  
## perte ~ xa + xb + xy + xgn  
##   
## Df Sum of Sq RSS AIC  
## <none> 119.74 -76.271  
## - xb 1 1.584 121.32 -75.788  
## - xy 1 1.868 121.60 -75.345  
## - xgn 1 4.547 124.28 -71.227  
## - xa 1 48.311 168.05 -14.208

AIC plus petite que le START -> AIC la plus petite possible : arrêt de l’algo AIC=vraisemblance + pénalité liée au nombre de paramètre

anova(modele2)

## Analysis of Variance Table  
##   
## Response: perte  
## Df Sum Sq Mean Sq F value Pr(>F)   
## xa 1 965.17 965.17 1483.1955 < 2.2e-16 \*\*\*  
## xb 1 14.76 14.76 22.6862 3.855e-06 \*\*\*  
## xy 1 2.04 2.04 3.1408 0.07801 .   
## xgn 1 4.55 4.55 6.9867 0.00892 \*\*   
## Residuals 184 119.74 0.65   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Les variables significatives sont xa, xb et xgn au risque de 5%

summary(modele2)

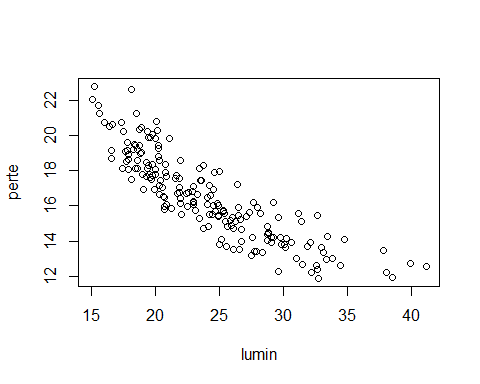
##   
## Call:  
## lm(formula = perte ~ xa + xb + xy + xgn, data = cafe)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.51027 -0.55664 0.01403 0.49402 2.48316   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 29.45636 0.68578 42.953 < 2e-16 \*\*\*  
## xa -1.30678 0.15166 -8.616 3.1e-15 \*\*\*  
## xb 0.14790 0.09481 1.560 0.12047   
## xy -0.23263 0.13730 -1.694 0.09189 .   
## xgn -0.41051 0.15530 -2.643 0.00892 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8067 on 184 degrees of freedom  
## Multiple R-squared: 0.8918, Adjusted R-squared: 0.8894   
## F-statistic: 379 on 4 and 184 DF, p-value: < 2.2e-16

Pour minimiser la perte en eau :

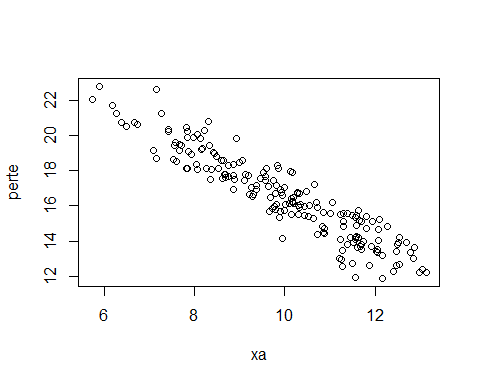
* xa =-1,3 > valeurs grande > maximiser xa
* xb =0,13 > valeurs petite > minimiser xb
* xgn =-0,41 > valeurs grande> maximiser xgn

5/ (a)

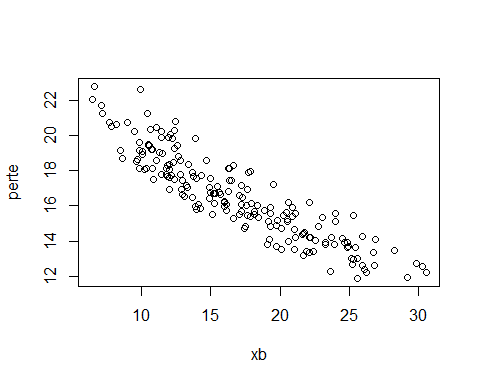
plot(perte~lumin, data=cafe)



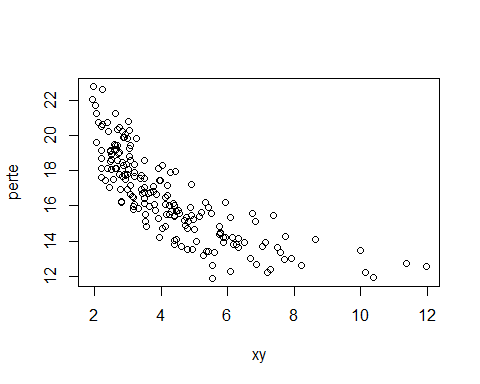
plot(perte~xa, data=cafe)



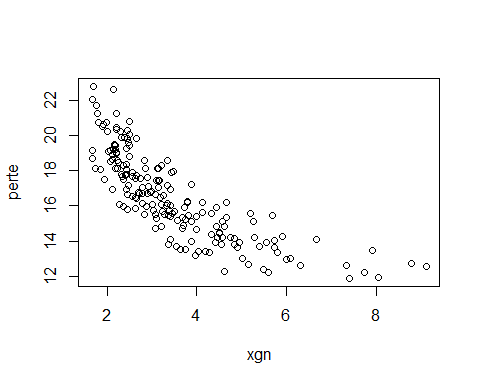
plot(perte~xb, data=cafe)



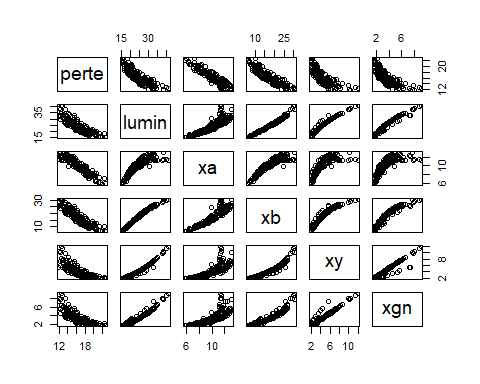
plot(perte~xy, data=cafe)



plot(perte~xgn, data=cafe)



pairs(cafe[,2:7])



Le choix du staticien est valable quant aux nouvelles variables pour voir si les variables expliquent mieux avec une courbe inverse(car pour lumin, xy et xgn on a l’impression de voir une fonction inverse).

cafe2=data.frame(cafe,Tlumin=1/cafe$lumin,Txy=1/cafe$xy,Txgn=1/cafe$xgn)

modele3=lm(perte~lumin+xa+xb+xy+xgn+Tlumin+Txy+Txgn,data=cafe2)

summary(modele3)

##   
## Call:  
## lm(formula = perte ~ lumin + xa + xb + xy + xgn + Tlumin + Txy +   
## Txgn, data = cafe2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.54030 -0.57527 -0.03807 0.53348 2.93653   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.45481 10.72459 2.187 0.0300 \*   
## lumin 0.02016 0.46078 0.044 0.9652   
## xa -1.29614 0.18244 -7.105 2.72e-11 \*\*\*  
## xb 0.15761 0.22649 0.696 0.4874   
## xy -0.06774 0.36779 -0.184 0.8541   
## xgn -0.64787 0.40310 -1.607 0.1098   
## Tlumin 293.39275 145.66016 2.014 0.0455 \*   
## Txy -8.59677 4.95488 -1.735 0.0845 .   
## Txgn -15.29300 4.31368 -3.545 0.0005 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.783 on 180 degrees of freedom  
## Multiple R-squared: 0.9003, Adjusted R-squared: 0.8958   
## F-statistic: 203.1 on 8 and 180 DF, p-value: < 2.2e-16

modele3$coefficients

## (Intercept) lumin xa xb xy   
## 23.45480635 0.02015892 -1.29613522 0.15761322 -0.06774019   
## xgn Tlumin Txy Txgn   
## -0.64786862 293.39274602 -8.59677287 -15.29300222

modele4=step(modele3, direction="backward")

## Start: AIC=-83.7  
## perte ~ lumin + xa + xb + xy + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## - lumin 1 0.0012 110.35 -85.698  
## - xy 1 0.0208 110.37 -85.665  
## - xb 1 0.2969 110.65 -85.192  
## <none> 110.35 -83.700  
## - xgn 1 1.5836 111.93 -83.007  
## - Txy 1 1.8454 112.19 -82.566  
## - Tlumin 1 2.4872 112.84 -81.487  
## - Txgn 1 7.7053 118.06 -72.943  
## - xa 1 30.9433 141.29 -38.983  
##   
## Step: AIC=-85.7  
## perte ~ xa + xb + xy + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## - xy 1 0.054 110.40 -87.606  
## - xb 1 0.889 111.24 -86.181  
## <none> 110.35 -85.698  
## - Txy 1 2.136 112.49 -84.075  
## - xgn 1 4.768 115.12 -79.704  
## - Tlumin 1 5.145 115.50 -79.085  
## - Txgn 1 9.080 119.43 -72.753  
## - xa 1 46.487 156.84 -21.255  
##   
## Step: AIC=-87.61  
## perte ~ xa + xb + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## - xb 1 0.926 111.33 -88.027  
## <none> 110.40 -87.606  
## - Txy 1 2.244 112.65 -85.803  
## - Tlumin 1 5.114 115.52 -81.047  
## - xgn 1 6.293 116.70 -79.129  
## - Txgn 1 10.255 120.66 -72.819  
## - xa 1 59.960 170.37 -7.619  
##   
## Step: AIC=-88.03  
## perte ~ xa + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## <none> 111.33 -88.027  
## - Txy 1 2.839 114.17 -85.267  
## - Tlumin 1 4.190 115.52 -83.044  
## - Txgn 1 9.352 120.68 -74.782  
## - xgn 1 10.258 121.59 -73.368  
## - xa 1 87.394 198.72 19.483

anova(modele4)

## Analysis of Variance Table  
##   
## Response: perte  
## Df Sum Sq Mean Sq F value Pr(>F)   
## xa 1 965.17 965.17 1586.5019 < 2.2e-16 \*\*\*  
## xgn 1 19.33 19.33 31.7724 6.467e-08 \*\*\*  
## Tlumin 1 0.77 0.77 1.2629 0.2625763   
## Txy 1 0.31 0.31 0.5067 0.4774790   
## Txgn 1 9.35 9.35 15.3731 0.0001246 \*\*\*  
## Residuals 183 111.33 0.61   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

modele variables stepwide = les 2 ajouts et suppressions de variables

modele5=step(modele3, direction="both")

## Start: AIC=-83.7  
## perte ~ lumin + xa + xb + xy + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## - lumin 1 0.0012 110.35 -85.698  
## - xy 1 0.0208 110.37 -85.665  
## - xb 1 0.2969 110.65 -85.192  
## <none> 110.35 -83.700  
## - xgn 1 1.5836 111.93 -83.007  
## - Txy 1 1.8454 112.19 -82.566  
## - Tlumin 1 2.4872 112.84 -81.487  
## - Txgn 1 7.7053 118.06 -72.943  
## - xa 1 30.9433 141.29 -38.983  
##   
## Step: AIC=-85.7  
## perte ~ xa + xb + xy + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## - xy 1 0.054 110.40 -87.606  
## - xb 1 0.889 111.24 -86.181  
## <none> 110.35 -85.698  
## - Txy 1 2.136 112.49 -84.075  
## + lumin 1 0.001 110.35 -83.700  
## - xgn 1 4.768 115.12 -79.704  
## - Tlumin 1 5.145 115.50 -79.085  
## - Txgn 1 9.080 119.43 -72.753  
## - xa 1 46.487 156.84 -21.255  
##   
## Step: AIC=-87.61  
## perte ~ xa + xb + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## - xb 1 0.926 111.33 -88.027  
## <none> 110.40 -87.606  
## - Txy 1 2.244 112.65 -85.803  
## + xy 1 0.054 110.35 -85.698  
## + lumin 1 0.034 110.37 -85.665  
## - Tlumin 1 5.114 115.52 -81.047  
## - xgn 1 6.293 116.70 -79.129  
## - Txgn 1 10.255 120.66 -72.819  
## - xa 1 59.960 170.37 -7.619  
##   
## Step: AIC=-88.03  
## perte ~ xa + xgn + Tlumin + Txy + Txgn  
##   
## Df Sum of Sq RSS AIC  
## <none> 111.33 -88.027  
## + xb 1 0.926 110.40 -87.606  
## + lumin 1 0.416 110.91 -86.735  
## + xy 1 0.091 111.24 -86.181  
## - Txy 1 2.839 114.17 -85.267  
## - Tlumin 1 4.190 115.52 -83.044  
## - Txgn 1 9.352 120.68 -74.782  
## - xgn 1 10.258 121.59 -73.368  
## - xa 1 87.394 198.72 19.483

anova(modele5)

## Analysis of Variance Table  
##   
## Response: perte  
## Df Sum Sq Mean Sq F value Pr(>F)   
## xa 1 965.17 965.17 1586.5019 < 2.2e-16 \*\*\*  
## xgn 1 19.33 19.33 31.7724 6.467e-08 \*\*\*  
## Tlumin 1 0.77 0.77 1.2629 0.2625763   
## Txy 1 0.31 0.31 0.5067 0.4774790   
## Txgn 1 9.35 9.35 15.3731 0.0001246 \*\*\*  
## Residuals 183 111.33 0.61   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

6/

extractAIC(modele1)

## [1] 6.00000 -74.52464

extractAIC(modele2)

## [1] 5.0000 -76.2711

extractAIC(modele3)

## [1] 9.00000 -83.70015

extractAIC(modele4)

## [1] 6.00000 -88.02695

Le meilleur modèle en terme de AIC est le modèle 4.

res1=summary(modele1)  
res1$adj.r.squared

## [1] 0.8889571

res2=summary(modele2)  
res2$adj.r.squared

## [1] 0.8894124

res3=summary(modele3)  
res3$adj.r.squared

## [1] 0.8958167

res4=summary(modele4)  
res4$adj.r.squared

## [1] 0.8966134

Le meilleur modèle en termes de R² est le modèle 4.

Dans R, ajuster c’est pour la régression multiple.