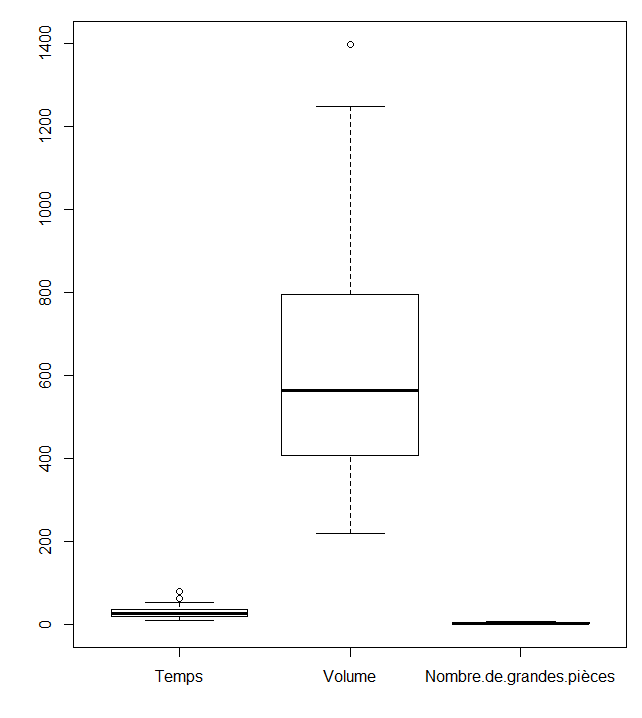
Les travaux de l'évaluation 1: Data mining avancé et Android

1. La dépendance de plusieurs variables
2. Déménagement

boxplot(demenagement)



summary(demenagement)

Temps Volume Nombre.de.grandes.pièces

Min. : 9.00 Min. : 220.0 Min. :1.000

1st Qu.:19.88 1st Qu.: 411.2 1st Qu.:2.000

Median :25.00 Median : 563.0 Median :3.000

Mean :28.96 Mean : 625.6 Mean :3.056

3rd Qu.:34.75 3rd Qu.: 793.8 3rd Qu.:4.000

Max. :79.50 Max. :1397.0 Max. :7.000

model<-lm(formula = demenagement$Temps ~ demenagement$Volume + demenagement$Nombre.de.grandes.pièces)

Coefficients:

(Intercept) demenagement$Volume demenagement$Nombre.de.grandes.pièces

-3.91522 0.03192 4.22283

summary(model)

Call:

lm(formula = demenagement$Temps ~ demenagement$Volume + demenagement$Nombre.de.grandes.pièces)

Residuals:

Min 1Q Median 3Q Max

-9.2921 -2.1574 0.3798 2.6174 9.2571

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.915221 1.673790 -2.339 0.0255 \*

demenagement$Volume 0.031924 0.004604 6.934 6.36e-08 \*\*\*

demenagement$Nombre.de.grandes.pièces 4.222834 0.914190 4.619 5.64e-05 \*\*\*

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.98 on 33 degrees of freedom

Multiple R-squared: 0.9327, Adjusted R-squared: 0.9287

F-statistic: 228.8 on 2 and 33 DF, p-value: < 2.2e-16

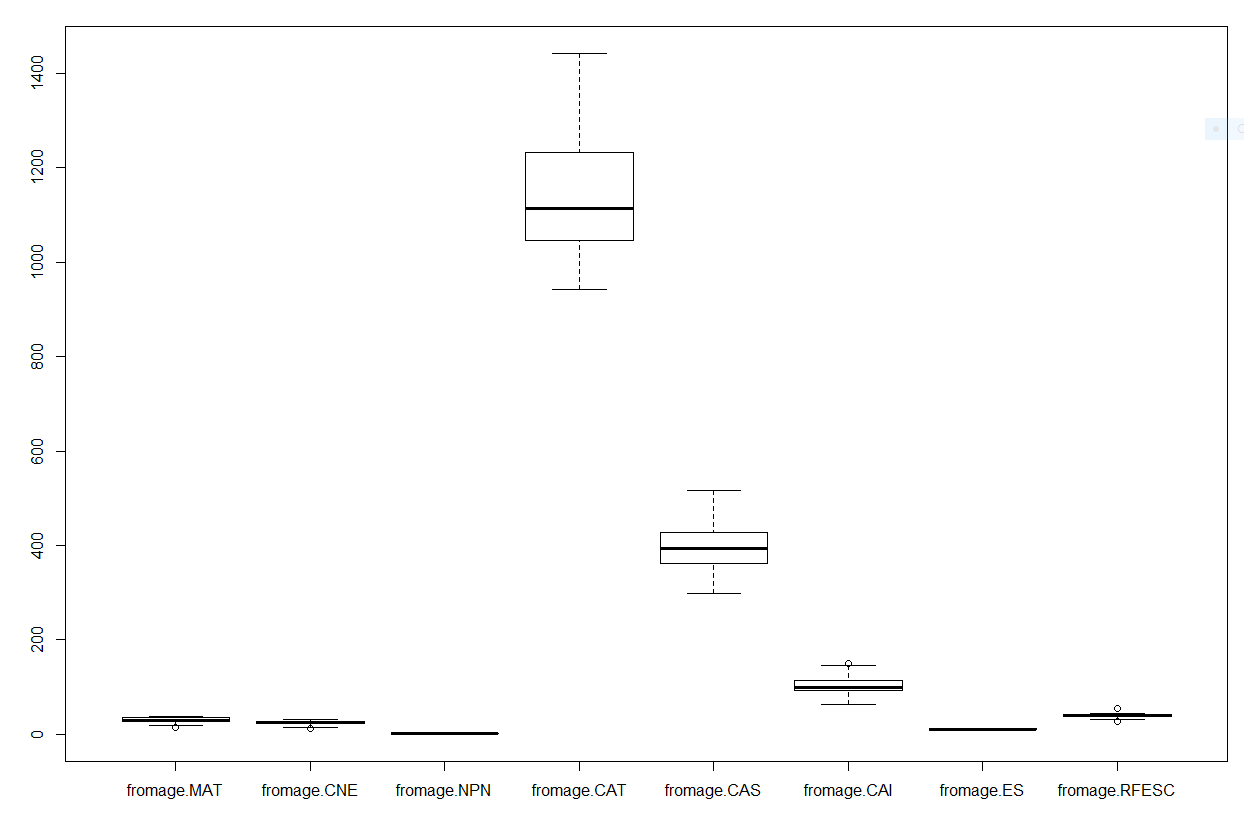
Il y a une corrélation entre le temp et le volume + le nombre de grande pièce.

1. Le fromage boursoulavien

fromage <-read.table ("C:/Users/dries/Documents/école/3eme supérieur/ECommerce /RdtFromage.txt", sep=";", header = TRUE)

fromageReduit <- data.frame(fromage$MAT, fromage$CNE, fromage$NPN, fromage$CAT, fromage$CAS, fromage$CAI, fromage$ES, fromage$RFESC)

boxplot(fromageReduit)



summary(fromageReduit)

fromage.MAT fromage.CNE fromage.NPN fromage.CAT

Min. :14.80 Min. :12.10 Min. :1.060 Min. : 942

1st Qu.:27.50 1st Qu.:22.30 1st Qu.:1.510 1st Qu.:1046

Median :29.80 Median :24.50 Median :1.730 Median :1114

Mean :30.38 Mean :24.76 Mean :1.771 Mean :1140

3rd Qu.:34.70 3rd Qu.:27.90 3rd Qu.:2.030 3rd Qu.:1232

Max. :38.10 Max. :32.40 Max. :2.710 Max. :1441

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

fromage.CAS fromage.CAI fromage.ES fromage.RFESC

Min. :298.0 Min. : 63.0 Min. : 9.75 Min. :26.40

1st Qu.:362.0 1st Qu.: 92.0 1st Qu.:10.76 1st Qu.:37.20

Median :394.0 Median : 99.0 Median :11.15 Median :39.30

Mean :398.4 Mean :103.2 Mean :11.11 Mean :39.27

3rd Qu.:427.0 3rd Qu.:115.0 3rd Qu.:11.51 3rd Qu.:41.60

Max. :516.0 Max. :150.0 Max. :11.96 Max. :54.60

model2<-lm(formula = fromageReduit$fromage.RFESC ~ fromageReduit$fromage.MAT + fromageReduit$fromage.CNE + fromageReduit$fromage.NPN + fromageReduit$fromage.CAT + fromageReduit$fromage.CAS + fromageReduit$fromage.CAI + fromageReduit$fromage.ES)

Call:

lm(formula = fromageReduit$fromage.RFESC ~ fromageReduit$fromage.MAT +

fromageReduit$fromage.CNE + fromageReduit$fromage.NPN + fromageReduit$fromage.CAT +

fromageReduit$fromage.CAS + fromageReduit$fromage.CAI + fromageReduit$fromage.ES)

Coefficients:

(Intercept) fromage.MAT fromage.CNE fromage.NPN fromage.CAT fromage.CAS

45.866869 -1.474130 2.616518 -4.388583 0.012730 0.010961

fromage.CAI fromage.ES

-0.001198 -3.383349

summary(model2)

Call:

lm(formula = fromageReduit$fromage.RFESC ~ fromageReduit$fromage.MAT +

fromageReduit$fromage.CNE + fromageReduit$fromage.NPN + fromageReduit$fromage.CAT +

fromageReduit$fromage.CAS + fromageReduit$fromage.CAI + fromageReduit$fromage.ES)

Residuals:

Min 1Q Median 3Q Max

-4.454 -1.969 -0.336 2.409 5.969

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 45.866869 12.147498 3.776 0.000633 \*\*\*

fromageReduit$fromage.MAT -1.474130 0.419369 -3.515 0.001301 \*\*

fromageReduit$fromage.CNE 2.616518 0.497517 5.259 8.59e-06 \*\*\*

fromageReduit$fromage.NPN -4.388583 1.367767 -3.209 0.002965 \*\*

fromageReduit$fromage.CAT 0.012730 0.004710 2.703 0.010783 \*

fromageReduit$fromage.CAS 0.010961 0.011374 0.964 0.342224

fromageReduit$fromage.CAI -0.001198 0.025440 -0.047 0.962722

fromageReduit$fromage.ES -3.383349 1.103702 -3.065 0.004313 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.684 on 33 degrees of freedom

Multiple R-squared: 0.6819, Adjusted R-squared: 0.6145

F-statistic: 10.11 on 7 and 33 DF, p-value: 1.086e-06

Plot(fromageReduit);

Une image contenant table

Description générée automatiquement

fromageReduit2<-data.frame(fromageReduit$fromage.CNE, fromageReduit$fromage.NPN, fromageReduit$fromage.CAT, fromageReduit$fromage.CAS, fromageReduit$fromage.CAI, fromageReduit$fromage.ES, fromageReduit$fromage.RFESC)

On retire fromage.MAT!

model3<-lm(data=fromageReduit2,fromageReduit.fromage.RFESC~fromageReduit.fromage.CNE + fromageReduit.fromage.NPN + fromageReduit.fromage.CAT +

+ fromageReduit.fromage.CAS + fromageReduit.fromage.CAI + fromageReduit.fromage.ES)

Call:

lm(formula = fromageReduit.fromage.RFESC ~ fromageReduit.fromage.CNE +

fromageReduit.fromage.NPN + fromageReduit.fromage.CAT + fromageReduit.fromage.CAS +

fromageReduit.fromage.CAI + fromageReduit.fromage.ES, data = fromageReduit2)

Coefficients:

(Intercept) fromageReduit.fromage.CNE fromageReduit.fromage.NPN fromageReduit.fromage.CAT

41.154117 0.938755 -3.120777 0.006905

fromageReduit.fromage.CAS fromageReduit.fromage.CAI fromageReduit.fromage.ES

0.010309 0.017914 -3.009186

summary(model3)

Call:

lm(formula = fromageReduit.fromage.RFESC ~ fromageReduit.fromage.CNE +

fromageReduit.fromage.NPN + fromageReduit.fromage.CAT + fromageReduit.fromage.CAS +

fromageReduit.fromage.CAI + fromageReduit.fromage.ES, data = fromageReduit2)

Residuals:

Min 1Q Median 3Q Max

-9.0666 -2.0184 -0.3284 2.2995 6.1930

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 41.154117 13.944517 2.951 0.0057 \*\*

fromageReduit.fromage.CNE 0.938755 0.162145 5.790 1.62e-06 \*\*\*

fromageReduit.fromage.NPN -3.120777 1.523842 -2.048 0.0483 \*

fromageReduit.fromage.CAT 0.006905 0.005092 1.356 0.1840

fromageReduit.fromage.CAS 0.010309 0.013136 0.785 0.4380

fromageReduit.fromage.CAI 0.017914 0.028703 0.624 0.5367

fromageReduit.fromage.ES -3.009186 1.268821 -2.372 0.0235 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.1 on 34 degrees of freedom

Multiple R-squared: 0.5629, Adjusted R-squared: 0.4857

F-statistic: 7.296 on 6 and 34 DF, p-value: 4.59e-05

model4<-lm(data=fromageReduit2,fromageReduit.fromage.RFESC~fromageReduit.fromage.CNE + fromageReduit.fromage.NPN + fromageReduit.fromage.CAT + fromageReduit.fromage.ES)

summary(model4)

Call:

lm(formula = fromageReduit.fromage.RFESC ~ fromageReduit.fromage.CNE +

fromageReduit.fromage.NPN + fromageReduit.fromage.CAT + fromageReduit.fromage.ES,

data = fromageReduit2)

Residuals:

Min 1Q Median 3Q Max

-10.1253 -1.8641 -0.3326 2.2601 5.6954

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 49.864250 11.796949 4.227 0.000155 \*\*\*

fromageReduit.fromage.CNE 0.896263 0.149118 6.010 6.73e-07 \*\*\*

fromageReduit.fromage.NPN -3.448164 1.440109 -2.394 0.021979 \*

fromageReduit.fromage.CAT 0.009524 0.004444 2.143 0.038940 \*

fromageReduit.fromage.ES -3.378920 1.209792 -2.793 0.008312 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.077 on 36 degrees of freedom

Multiple R-squared: 0.544, Adjusted R-squared: 0.4933

F-statistic: 10.74 on 4 and 36 DF, p-value: 7.842e-06

**La p-value est petite donc nous pouvons faire confiance à nos régresseurs. Il y a une corrélation entre le rendement et le reste mais qui n'est pas élevé!**

1. ANOVA
2. Les civilisations précolombiennes

summary(civilisation)

nbrRecipient civilisation

Min. : 28.00 Cuacuacomeqiqi:15

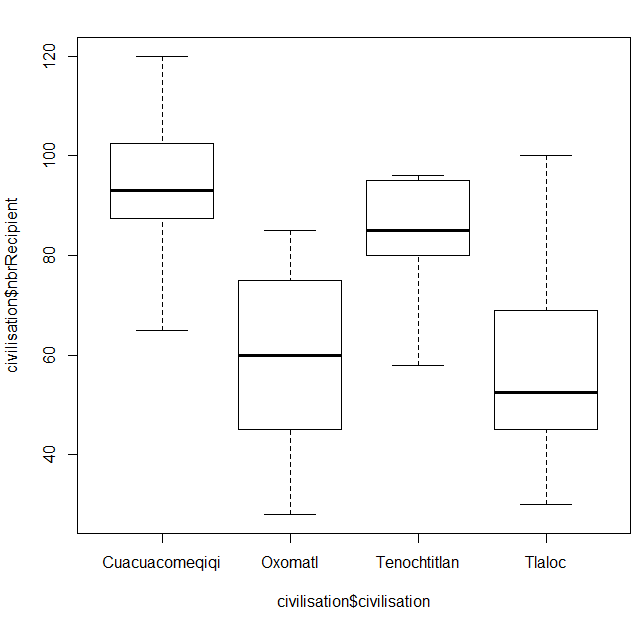
1st Qu.: 55.75 Oxomatl :10

Median : 79.00 Tenochtitlan : 9

Mean : 75.04 Tlaloc :12

3rd Qu.: 93.00

Max. :120.00

plot(civilisation$nbrRecipient~civilisation$civilisation)

model<-lm(civilisation$nbrRecipient~civilisation$civilisation)

model

Call:

lm(formula = civilisation$nbrRecipient ~ civilisation$civilisation)

Coefficients:

(Intercept) civilisation$civilisationOxomatl civilisation$civilisationTenochtitlan civilisation$civilisationTlaloc 94.67 -35.37 -11.78 -36.92

anova(model)

Analysis of Variance Table

Response: civilisation$nbrRecipient

Df Sum Sq Mean Sq F value Pr(>F)

civilisation$civilisation 3 12397 4132.4 15.139 7.991e-07 \*\*\*

Residuals 42 11465 273.0

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

pairwise.t.test(civilisation$nbrRecipient,civilisation$civilisation,p.adjust.method = "none",pool.sd = TRUE)

Pairwise comparisons using t tests with pooled SD

data: civilisation$nbrRecipient and civilisation$civilisation

Cuacuacomeqiqi Oxomatl Tenochtitlan

Oxomatl 4.8e-06 - -

Tenochtitlan 0.0983 0.0034 -

Tlaloc 8.5e-07 0.8276 0.0013

P value adjustment method: none

1. Médicaments

summary(medicaments)

molecules coefficient administration

AlphaVictoire:20 Min. : 5.00 Injection :30

BetaTriomphe :20 1st Qu.: 8.00 Voie Orale:30

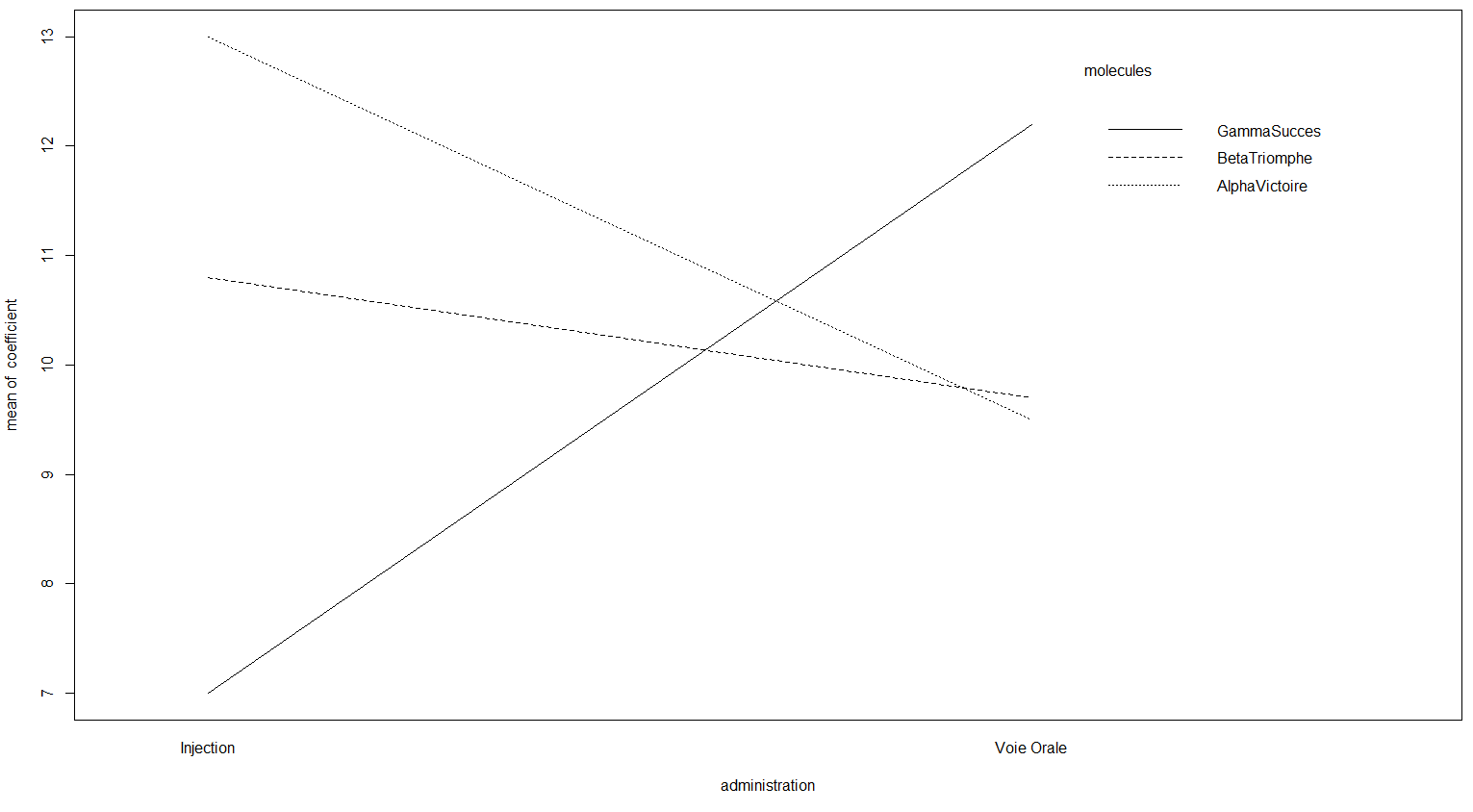
GammaSucces :20 Median :10.00

Mean :10.37

3rd Qu.:12.00

Max. :27.00

with(medicaments,interaction.plot(administration,molecules,coefficient))



model<-lm(medicaments$coefficient~medicaments$molecules \* medicaments$administration)

model

Call:

lm(formula = medicaments$coefficient ~ medicaments$molecules \* medicaments$administration)

Coefficients:

(Intercept) medicaments$moleculesBetaTriomphe medicaments$moleculesGammaSucces

13.0 -2.2 -6.0

medicaments$administrationVoie Orale

-3.5

medicaments$moleculesBetaTriomphe:medicaments$administrationVoie Orale

2.4

medicaments$moleculesGammaSucces:medicaments$administrationVoie Orale

8.7

anova(model)

Analysis of Variance Table

Response: medicaments$coefficient

Df /Sum Sq /Mean Sq /F value/ Pr(>F)

medicaments$molecules 2 / 27.63 /13.817 /1.4030/0.2546829

medicaments$administration 1 / 0.60 / 0.600 /0.0609/0.8059756

medicaments$molecules:medicaments$administration 2 /201.90/ 100.950/10.2507/0.0001683

Residuals 54 531.80 9.848

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

summary(model)

Call:

lm(formula = medicaments$coefficient ~ medicaments$molecules \*

medicaments$administration)

Residuals:

Min 1Q Median 3Q Max

-5.200 -1.575 -0.100 1.300 14.800

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 13.0000 0.9924 13.100 < 2e-16 \*\*\*

medicaments$moleculesBetaTriomphe -2.2000 1.4034 -1.568 0.1228

medicaments$moleculesGammaSucces -6.0000 1.4034 -4.275 7.82e-05 \*\*\*

medicaments$administrationVoie Orale -3.5000 1.4034 -2.494 0.0157 \*

medicaments$moleculesBetaTriomphe:medicaments$administrationVoie Orale

2.4000 1.9848 1.209 0.2318

medicaments$moleculesGammaSucces:medicaments$administrationVoie Orale

8.7000 1.9848 4.383 5.43e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.138 on 54 degrees of freedom

Multiple R-squared: 0.302, Adjusted R-squared: 0.2374

F-statistic: 4.674 on 5 and 54 DF, p-value: 0.001285