02441: Case 1

Zuzanna Czelusniak Thomas Aamand Witting Zoltán György Varga

## Read data, clean data

library(car)

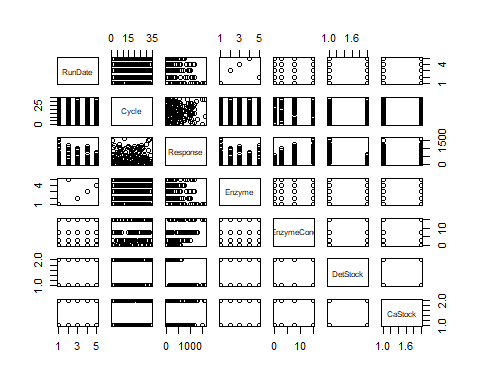
## Warning: package 'car' was built under R version 4.0.5

## Loading required package: carData

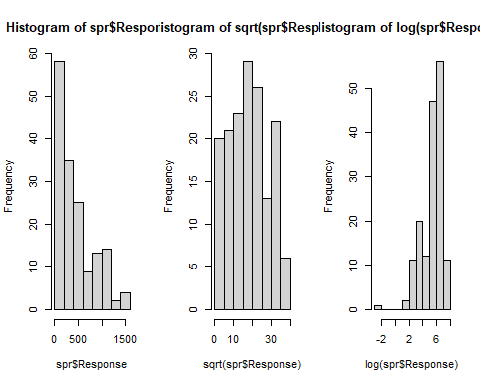
spr <- read.table("spr.txt", header=TRUE)  
#spr <- spr[-c(1,2)]  
spr$RunDate <- as.factor(spr$RunDate)  
levels(spr$RunDate) <- c(1,3,5,7,9)  
  
spr$Enzyme <- as.factor(spr$Enzyme)  
spr$DetStock <- as.factor(spr$DetStock)  
spr$CaStock <- as.factor(spr$CaStock)  
str(spr)

## 'data.frame': 160 obs. of 7 variables:  
## $ RunDate : Factor w/ 5 levels "1","3","5","7",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Cycle : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Response : num 323 614 326 162 545 ...  
## $ Enzyme : Factor w/ 5 levels "A","B","C","D",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ EnzymeConc: num 2.5 7.5 15 7.5 2.5 7.5 0 15 2.5 15 ...  
## $ DetStock : Factor w/ 2 levels "Det+","Det0": 1 1 2 2 1 2 2 2 2 1 ...  
## $ CaStock : Factor w/ 2 levels "Ca+","Ca0": 1 2 1 2 2 1 1 2 1 2 ...

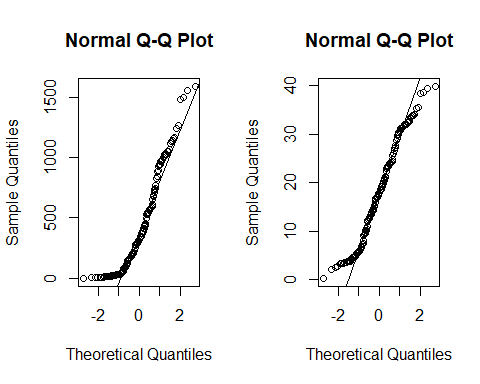
pairs(spr)



# Response variable doesn't look normal distributed  
par(mfrow=c(1,3))  
hist(spr$Response, 10)  
hist(sqrt(spr$Response), 10)  
hist(log(spr$Response), 10)



# Looks to be decent with sqrt transformation, check with boxCox which transformation would work the best  
lm.a <- lm(Response~EnzymeConc, spr)  
  
# Sqrt transformation suggested by boxCox - same result as the visual histogram inspection  
  
# Perform the sqrt transformation  
par(mfrow=c(1,2))  
qqnorm(spr$Response)  
qqline(spr$Response)  
spr$Response <- sqrt(spr$Response)  
spr$EnzymeConc <- spr$EnzymeConc  
  
qqnorm(spr$Response)  
qqline(spr$Response)

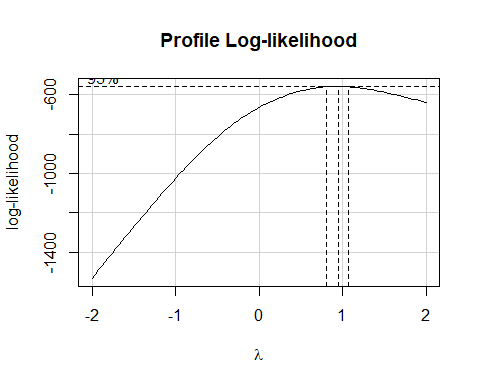
 ## Built complete model, then reduce it to a minimal model by utilizing the step function

lm.complete <- lm(Response~(EnzymeConc+Enzyme+DetStock+CaStock)^2, spr)  
step(lm.complete)

## Start: AIC=342.87  
## Response ~ (EnzymeConc + Enzyme + DetStock + CaStock)^2  
##   
## Df Sum of Sq RSS AIC  
## - Enzyme:CaStock 4 25.796 1048.9 338.85  
## - EnzymeConc:DetStock 1 2.447 1025.6 341.25  
## - DetStock:CaStock 1 5.522 1028.6 341.73  
## - EnzymeConc:CaStock 1 9.697 1032.8 342.38  
## <none> 1023.1 342.87  
## - Enzyme:DetStock 4 64.790 1087.9 344.69  
## - EnzymeConc:Enzyme 4 205.185 1228.3 364.11  
##   
## Step: AIC=338.85  
## Response ~ EnzymeConc + Enzyme + DetStock + CaStock + EnzymeConc:Enzyme +   
## EnzymeConc:DetStock + EnzymeConc:CaStock + Enzyme:DetStock +   
## DetStock:CaStock  
##   
## Df Sum of Sq RSS AIC  
## - EnzymeConc:DetStock 1 2.447 1051.4 337.23  
## - DetStock:CaStock 1 5.522 1054.4 337.69  
## - EnzymeConc:CaStock 1 9.697 1058.6 338.33  
## <none> 1048.9 338.85  
## - Enzyme:DetStock 4 64.790 1113.7 340.44  
## - EnzymeConc:Enzyme 4 205.185 1254.1 359.44  
##   
## Step: AIC=337.23  
## Response ~ EnzymeConc + Enzyme + DetStock + CaStock + EnzymeConc:Enzyme +   
## EnzymeConc:CaStock + Enzyme:DetStock + DetStock:CaStock  
##   
## Df Sum of Sq RSS AIC  
## - DetStock:CaStock 1 5.522 1056.9 336.06  
## - EnzymeConc:CaStock 1 9.697 1061.1 336.70  
## <none> 1051.4 337.23  
## - Enzyme:DetStock 4 64.790 1116.2 338.79  
## - EnzymeConc:Enzyme 4 205.185 1256.5 357.75  
##   
## Step: AIC=336.06  
## Response ~ EnzymeConc + Enzyme + DetStock + CaStock + EnzymeConc:Enzyme +   
## EnzymeConc:CaStock + Enzyme:DetStock  
##   
## Df Sum of Sq RSS AIC  
## - EnzymeConc:CaStock 1 9.697 1066.6 335.53  
## <none> 1056.9 336.06  
## - Enzyme:DetStock 4 64.790 1121.7 337.58  
## - EnzymeConc:Enzyme 4 205.185 1262.1 356.45  
##   
## Step: AIC=335.53  
## Response ~ EnzymeConc + Enzyme + DetStock + CaStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock  
##   
## Df Sum of Sq RSS AIC  
## - CaStock 1 1.118 1067.7 333.69  
## <none> 1066.6 335.53  
## - Enzyme:DetStock 4 64.790 1131.4 336.96  
## - EnzymeConc:Enzyme 4 205.185 1271.8 355.68  
##   
## Step: AIC=333.69  
## Response ~ EnzymeConc + Enzyme + DetStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock  
##   
## Df Sum of Sq RSS AIC  
## <none> 1067.7 333.69  
## - Enzyme:DetStock 4 64.79 1132.5 335.12  
## - EnzymeConc:Enzyme 4 205.19 1272.9 353.82

##   
## Call:  
## lm(formula = Response ~ EnzymeConc + Enzyme + DetStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock, data = spr)  
##   
## Coefficients:  
## (Intercept) EnzymeConc EnzymeB   
## 23.3813 1.1614 -6.0535   
## EnzymeC EnzymeD EnzymeE   
## -2.2309 -5.7076 -2.1342   
## DetStockDet0 EnzymeConc:EnzymeB EnzymeConc:EnzymeC   
## -14.7138 -0.1390 -0.3571   
## EnzymeConc:EnzymeD EnzymeConc:EnzymeE EnzymeB:DetStockDet0   
## -0.5826 -0.2617 1.2640   
## EnzymeC:DetStockDet0 EnzymeD:DetStockDet0 EnzymeE:DetStockDet0   
## -2.3775 0.4654 -1.1219

lm.minimal <- lm(formula = Response ~ EnzymeConc + Enzyme + DetStock + EnzymeConc:Enzyme +   
 Enzyme:DetStock, data = spr)  
boxCox(lm.minimal)



## Analyze minimal model

Anova(lm.minimal)

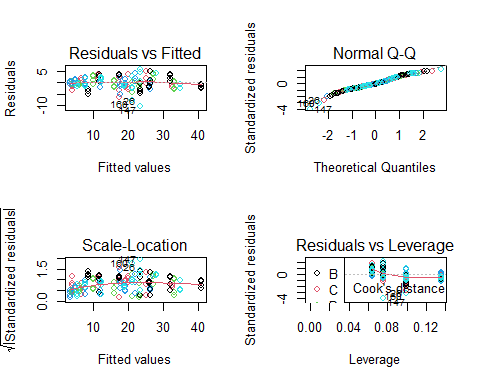
## Anova Table (Type II tests)  
##   
## Response: Response  
## Sum Sq Df F value Pr(>F)   
## EnzymeConc 4189.9 1 569.0079 < 2.2e-16 \*\*\*  
## Enzyme 1422.4 4 48.2938 < 2.2e-16 \*\*\*  
## DetStock 9081.5 1 1233.3290 < 2.2e-16 \*\*\*  
## EnzymeConc:Enzyme 205.2 4 6.9663 3.704e-05 \*\*\*  
## Enzyme:DetStock 64.8 4 2.1997 0.0719 .   
## Residuals 1067.7 145   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(lm.complete, lm.minimal)

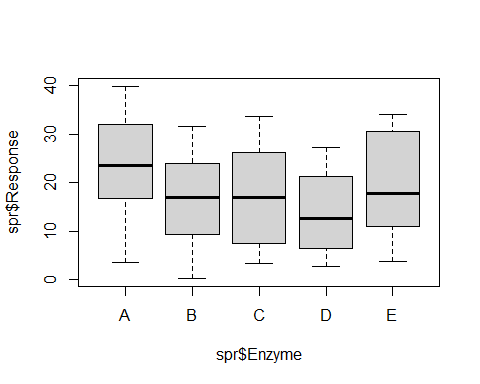
## Analysis of Variance Table  
##   
## Model 1: Response ~ (EnzymeConc + Enzyme + DetStock + CaStock)^2  
## Model 2: Response ~ EnzymeConc + Enzyme + DetStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 137 1023.1   
## 2 145 1067.7 -8 -44.58 0.7462 0.6506

## as

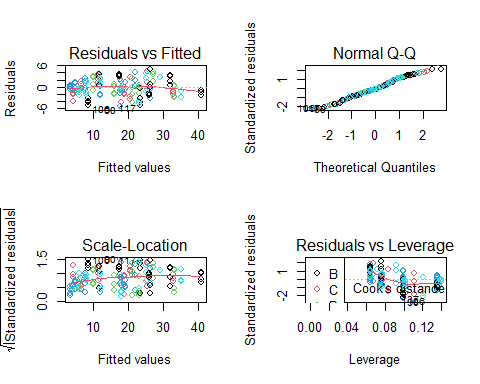
par(mfrow=c(2,2))  
plot(lm.minimal, col=spr$Enzyme)  
legend("topleft", legend=unique(spr$Enzyme), col=1:length(spr$Enzyme), pch=1)



par(mfrow=c(1,1))  
plot(spr$Response~spr$Enzyme)



spr.new <- spr[c(-26, -147, -160), ]  
lm.final <- update(lm.minimal, data=spr.new)  
par(mfrow=c(2,2))  
plot(lm.final, col=spr.new$Enzyme)  
legend("topleft", legend=unique(spr.new$Enzyme), col=1:length(spr.new$Enzyme), pch=1)

 # How does hardness and detergent influence the catalytic activity? Hardness (calcium) is not in our model - it does not influence the catalytic activity

summary(lm.final)

##   
## Call:  
## lm(formula = Response ~ EnzymeConc + Enzyme + DetStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock, data = spr.new)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.2279 -1.5836 -0.1801 1.5562 4.8551   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 23.38134 0.75618 30.920 < 2e-16 \*\*\*  
## EnzymeConc 1.16144 0.07391 15.715 < 2e-16 \*\*\*  
## EnzymeB -5.31752 1.09860 -4.840 3.34e-06 \*\*\*  
## EnzymeC -2.23093 1.06940 -2.086 0.038753 \*   
## EnzymeD -5.70759 1.06940 -5.337 3.65e-07 \*\*\*  
## EnzymeE -0.11263 1.12361 -0.100 0.920297   
## DetStockDet0 -14.71380 0.84669 -17.378 < 2e-16 \*\*\*  
## EnzymeConc:EnzymeB -0.18299 0.10559 -1.733 0.085273 .   
## EnzymeConc:EnzymeC -0.35709 0.10452 -3.417 0.000827 \*\*\*  
## EnzymeConc:EnzymeD -0.58262 0.10452 -5.574 1.21e-07 \*\*\*  
## EnzymeConc:EnzymeE -0.36441 0.10610 -3.435 0.000778 \*\*\*  
## EnzymeB:DetStockDet0 0.80262 1.20775 0.665 0.507411   
## EnzymeC:DetStockDet0 -2.37748 1.19740 -1.986 0.049012 \*   
## EnzymeD:DetStockDet0 0.46537 1.19740 0.389 0.698117   
## EnzymeE:DetStockDet0 -2.50165 1.21981 -2.051 0.042120 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.395 on 142 degrees of freedom  
## Multiple R-squared: 0.9489, Adjusted R-squared: 0.9438   
## F-statistic: 188.2 on 14 and 142 DF, p-value: < 2.2e-16

# Show that there is no significant difference between the minimal model and a model considering the hardness (meaning that hardness does not influence activity)  
lm.withca <- update(lm.final, ~ . + CaStock)  
anova(lm.final, lm.withca)

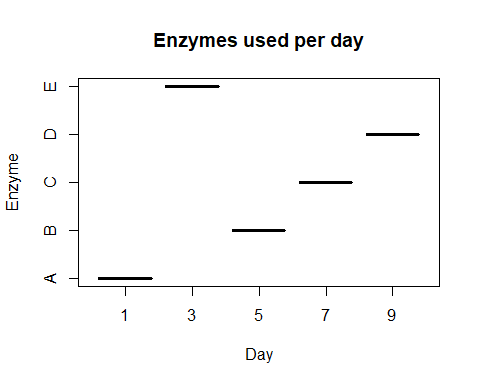
## Analysis of Variance Table  
##   
## Model 1: Response ~ EnzymeConc + Enzyme + DetStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock  
## Model 2: Response ~ EnzymeConc + Enzyme + DetStock + CaStock + EnzymeConc:Enzyme +   
## Enzyme:DetStock  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 142 814.39   
## 2 141 813.97 1 0.41217 0.0714 0.7897

# With detergence: intercept@23.43  
# without detergence: intercept@(23.43-14.81)=8.62

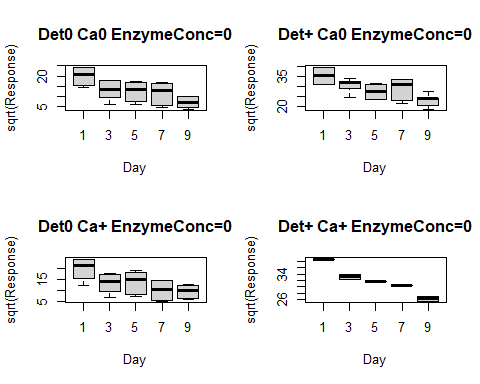
## Check for systematic errors

It seems that on a given day, experiments are only run with one specific enzyme. This could introduce errors to the data which may be hidden as an effect of the enzyme - but in reality it could be instrumental errors. By inspecting the subset of the data where the enzyme concentration is 0nM (that is, no enzyme is present) and where we keep the other variables constant, we can detect irregularities in the measured response. In an ideal experimental setting, the response would stay the same across all observations - but we detect the response differs on different run dates - meaning

plot((spr$RunDate), as.numeric(spr$Enzyme), yaxt = "n", xlab="Day", ylab="Enzyme", main="Enzymes used per day")  
axis(2, at=1:5, labels=LETTERS[1:5])



# Remove the enzyme from data  
spr.new <- spr[, -4]  
  
# For similar data on different days, is there different response?  
par(mfrow=c(2,2))  
plot(spr.new[spr.new$DetStock == "Det0" & spr.new$CaStock == "Ca0" & spr.new$EnzymeConc >= 1, ]$RunDate, spr.new[spr.new$DetStock == "Det0" & spr.new$CaStock == "Ca0" & spr.new$EnzymeConc >= 1, ]$Response, xlab="Day", ylab="sqrt(Response)", main="Det0 Ca0 EnzymeConc=0")  
  
plot(spr.new[spr.new$DetStock == "Det+" & spr.new$CaStock == "Ca0" & spr.new$EnzymeConc >= 1, ]$RunDate, spr.new[spr.new$DetStock == "Det+" & spr.new$CaStock == "Ca0" & spr.new$EnzymeConc >= 1, ]$Response, xlab="Day", ylab="sqrt(Response)", main="Det+ Ca0 EnzymeConc=0")  
  
plot(spr.new[spr.new$DetStock == "Det0" & spr.new$CaStock == "Ca+" & spr.new$EnzymeConc >= 1, ]$RunDate, spr.new[spr.new$DetStock == "Det0" & spr.new$CaStock == "Ca+" & spr.new$EnzymeConc >= 1, ]$Response, xlab="Day", ylab="sqrt(Response)", main="Det0 Ca+ EnzymeConc=0")  
  
plot(spr.new[spr.new$DetStock == "Det+" & spr.new$CaStock == "Ca+" & spr.new$EnzymeConc >= 12, ]$RunDate, spr.new[spr.new$DetStock == "Det+" & spr.new$CaStock == "Ca+" & spr.new$EnzymeConc >= 12, ]$Response, xlab="Day", ylab="sqrt(Response)", main="Det+ Ca+ EnzymeConc=0")



oneway.test(Response~RunDate, data = spr.new[spr.new$DetStock == "Det0" & spr.new$CaStock == "Ca0" & spr.new$EnzymeConc >= 12, ], var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Response and RunDate  
## F = 1095.1, num df = 4.0000, denom df = 2.4081, p-value = 0.0002617

oneway.test(Response~RunDate, data = spr.new[spr.new$DetStock == "Det+" & spr.new$CaStock == "Ca0" & spr.new$EnzymeConc >= 12, ], var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Response and RunDate  
## F = 117.01, num df = 4.0000, denom df = 2.3172, p-value = 0.004567

oneway.test(Response~RunDate, data = spr.new[spr.new$DetStock == "Det0" & spr.new$CaStock == "Ca+" & spr.new$EnzymeConc >= 12, ], var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Response and RunDate  
## F = 1396.5, num df = 4.0000, denom df = 2.1622, p-value = 0.0004257

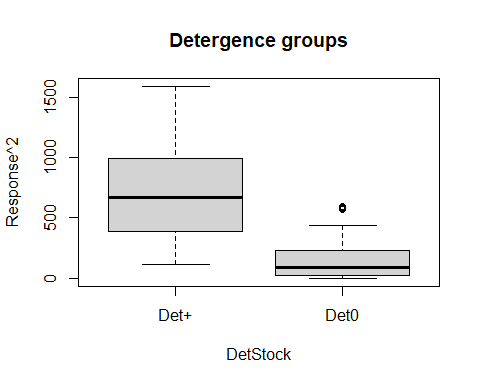
oneway.test(Response~RunDate, data = spr.new[spr.new$DetStock == "Det+" & spr.new$CaStock == "Ca+" & spr.new$EnzymeConc >= 12, ], var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Response and RunDate  
## F = NaN, num df = 4, denom df = NaN, p-value = NA

oneway.test(Response~DetStock, data = spr.new, var.equal = FALSE)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Response and DetStock  
## F = 206.46, num df = 1.00, denom df = 156.41, p-value < 2.2e-16

par(mfrow=c(1,1))  
plot(Response^2~DetStock, data = spr.new, main="Detergence groups")



confint(lm.minimal, 'EnzymeD', level=0.95)

## 2.5 % 97.5 %  
## EnzymeD -8.102532 -3.312638

6.291-1.811

## [1] 4.48