

# Midterm Project 3

[Midterm Project: Stat 445/545: Analysis of Variance], Stat 445/545, Spring 2023, Behzad FallahiFard, UNM

AUTHOR  
Behzad FallahiFard

PUBLISHED  
March 29, 2023

```
library(lmtest)
```

Loading required package: zoo

Warning: package 'zoo' was built under R version 4.2.2

Attaching package: 'zoo'

The following objects are masked from 'package:base':

as.Date, as.Date.numeric

```
potato<-read.table("potato.txt",header=TRUE)
n<-nrow(potato)
n
```

[1] 75

```
potato$regime<-factor(potato$regime,label=c("R","C"))
potato$variety<-factor(potato$variety,label=c("s1","s2"))
potato$temp<-factor(potato$temp,label=c("-4","-8"))
attach(potato)
potato[1:10,]
```

	variety	regime	temp	photo	leak
1	s1	R	-4	4.6386	2.25
2	s1	R	-4	5.7914	4.34
3	s1	R	-4	29.3515	4.25
4	s1	R	-4	18.4173	6.14
5	s1	R	-4	2.2556	2.38
6	s1	R	-8	8.2358	16.30

7	s1	R	-8	8.1972	5.24
8	s1	R	-8	-2.4482	3.25
9	s1	R	-8	5.2820	0.88
10	s1	R	-8	-5.1283	3.98

```
tapply(leak,regime,mean)
```

	R	C
	21.114444	6.575128

```
tapply(leak,variety,mean)
```

	s1	s2
	6.281429	19.917500

```
tapply(leak,temp,mean)
```

	-4	-8
	9.602703	17.401316

```
aggregate(leak~variety+regime+temp, data=potato, mean)
```

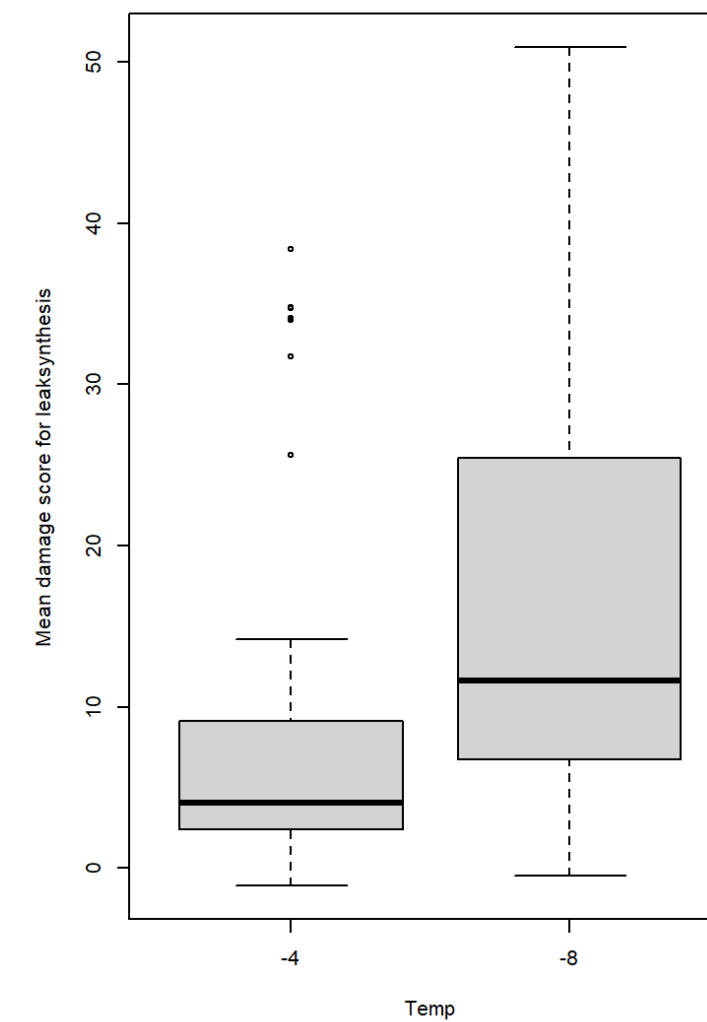
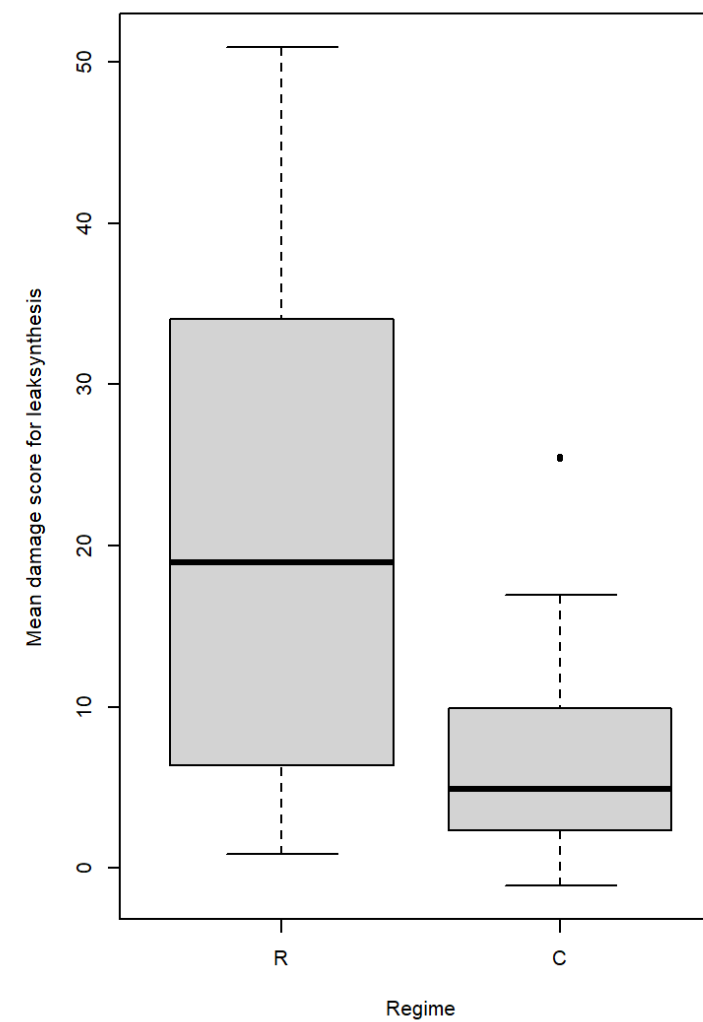
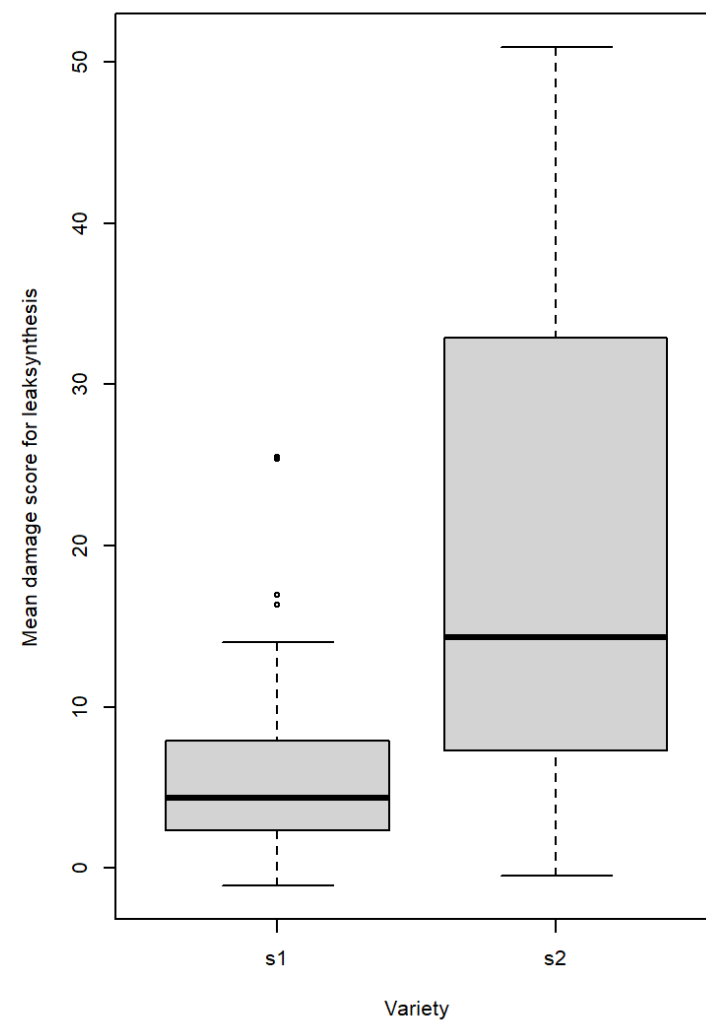
	variety	regime	temp	leak
1	s1	R	-4	3.872000
2	s2	R	-4	22.380000
3	s1	C	-4	2.339167
4	s2	C	-4	2.418571
5	s1	R	-8	5.930000
6	s2	R	-8	32.320769
7	s1	C	-8	10.982308
8	s2	C	-8	9.808571

```
#boxplot
par(mfrow=c(1,3))
boxplot(leak ~ variety,
        data = potato,
        xlab = "Variety",
        ylab = "Mean damage score for leaksynthesis")
```

```

boxplot(leak ~ regime,
        data = potato,
        xlab = "Regime",
        ylab = "Mean damage score for leaksynthesis")
boxplot(leak ~ temp,
        data = potato,
        xlab = "Temp",
        ylab = "Mean damage score for leaksynthesis")

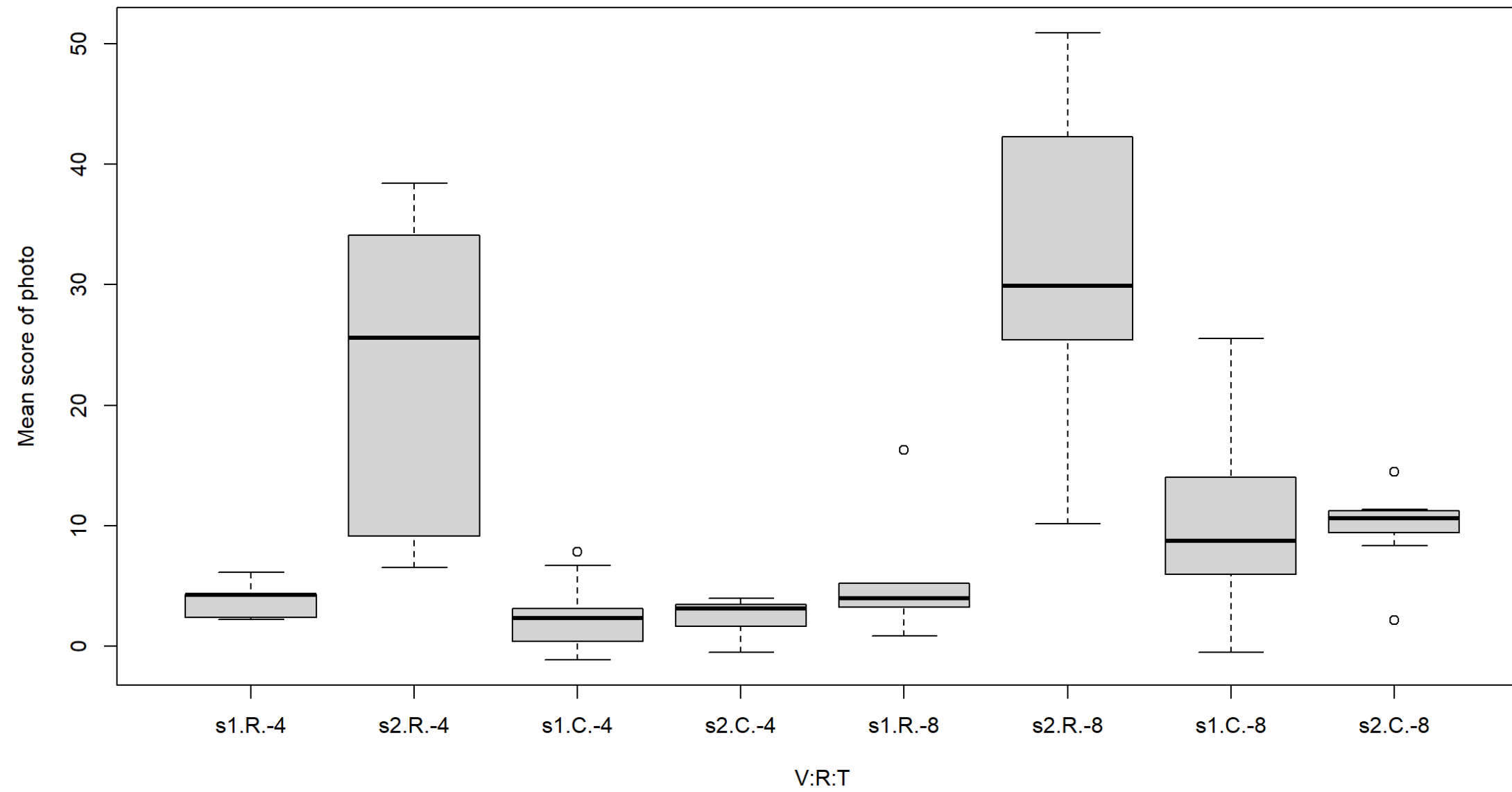
```



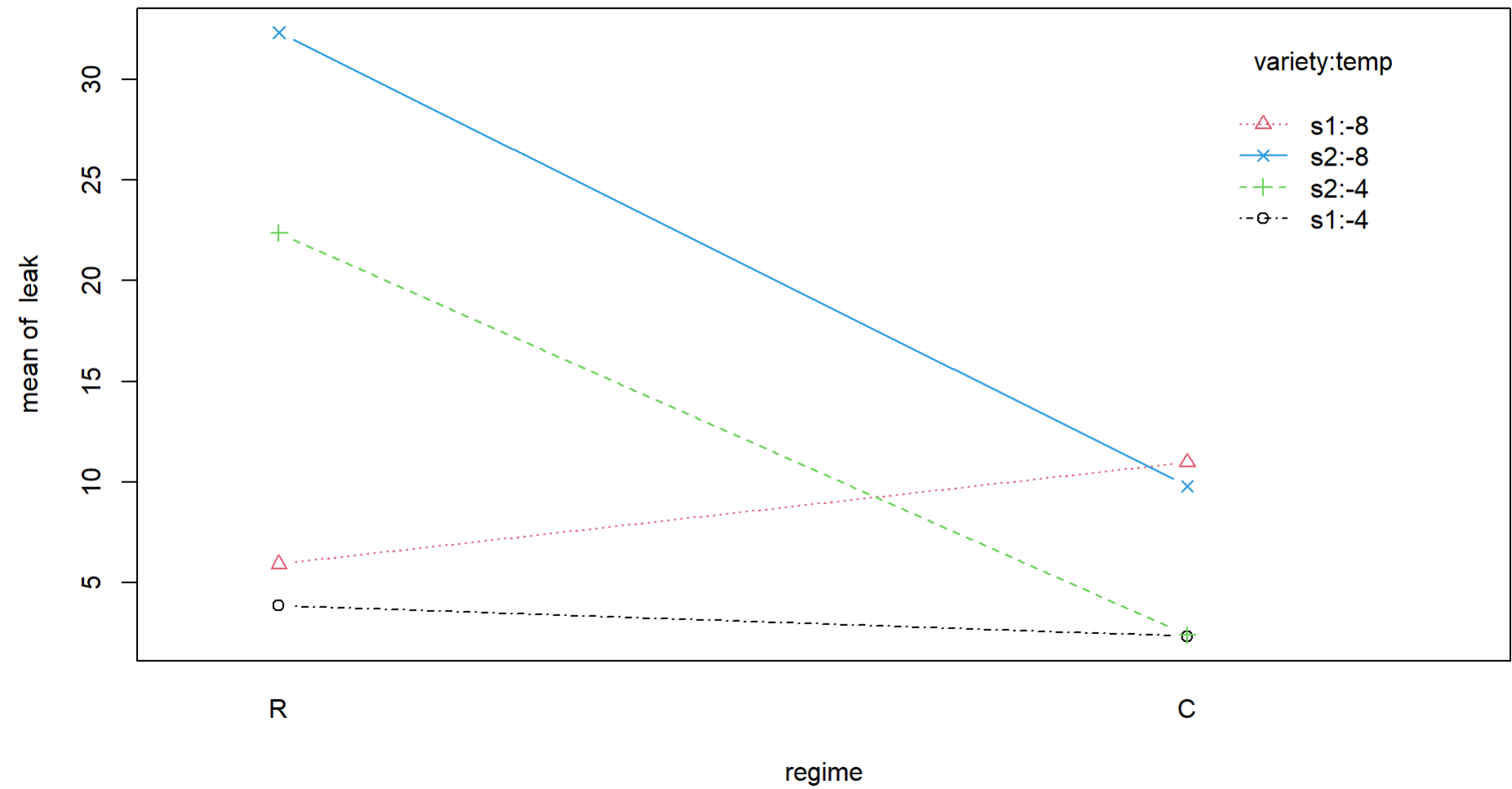
```

par(mfrow=c(1,1))
boxplot(leak ~ variety:regime:temp,
        data = potato,
        xlab = "V:R:T",
        ylab = "Mean score of photo")

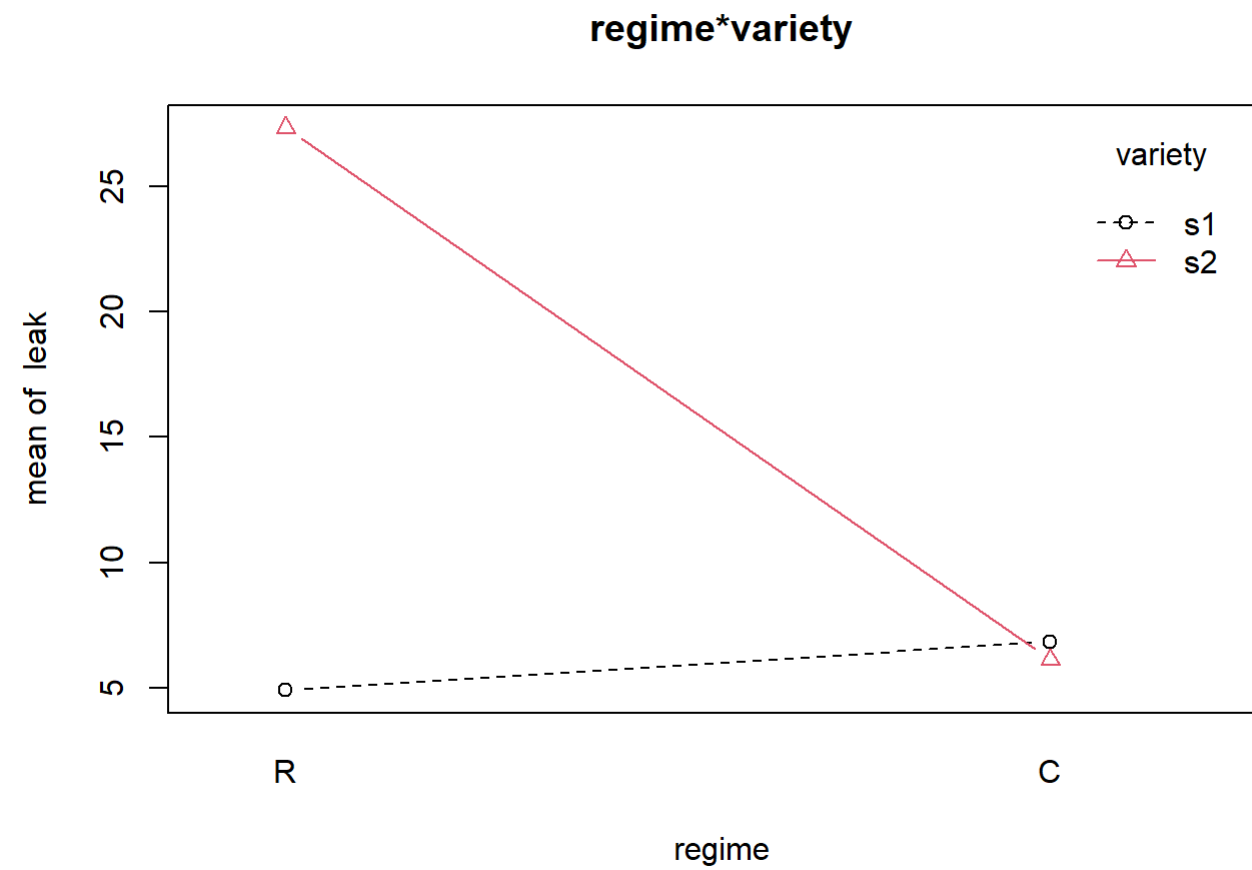
```



```
interaction.plot(regime,variety:temp,leak,type='b',  
col=1:4, pch=1:4)
```



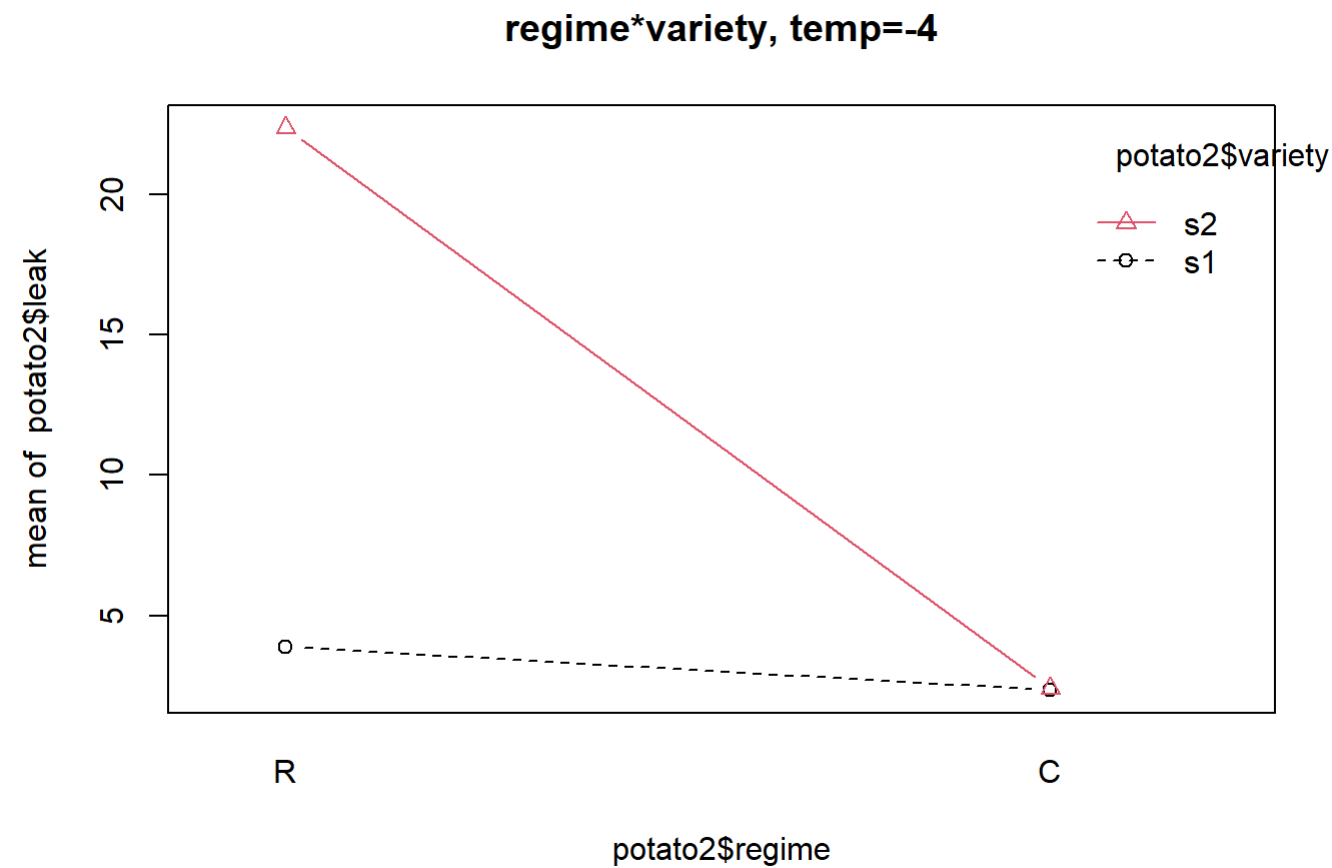
```
interaction.plot(regime,variety,leak,type='b',
col=1:2, pch=1:2,main="regime*variety")
```



```
potato2<-potato[temp== -4,]  
nrow(potato2)
```

```
[1] 37
```

```
interaction.plot(potato2$regime,potato2$variety,potato2$leak,type='b',  
col=1:2, pch=1:2,main="regime*variety, temp=-4")
```



#----- full

```
##fit full model
myfit<-lm(leak~variety*regime*temp,
contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))
library(car)
```

Loading required package: carData

```
Anova(myfit,type=3)
```

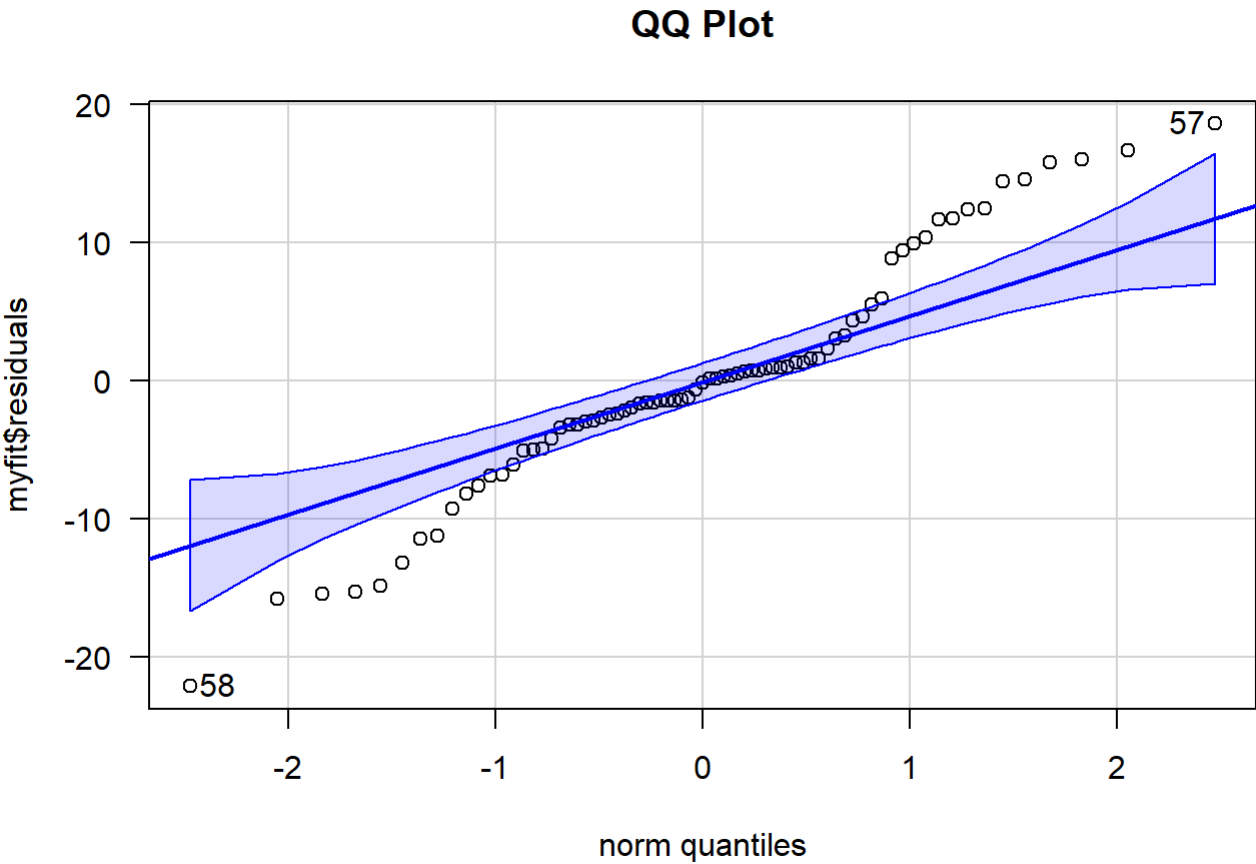
Anova Table (Type III tests)

Response: leak

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	8110.7	1	107.5408	1.441e-15	***
variety	1919.2	1	25.4465	3.701e-06	***
regime	1517.7	1	20.1233	2.921e-05	***
temp	785.9	1	10.4207	0.00193	**
variety:regime	2115.8	1	28.0529	1.413e-06	***

```
variety:temp      44.0  1  0.5829  0.44787
regime:temp       16.3  1  0.2158  0.64373
variety:regime:temp 83.5  1  1.1069  0.29654
Residuals        5053.1 67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
qqPlot(myfit$residuals, las = 1, main="QQ Plot")
```



```
[1] 58 57
```

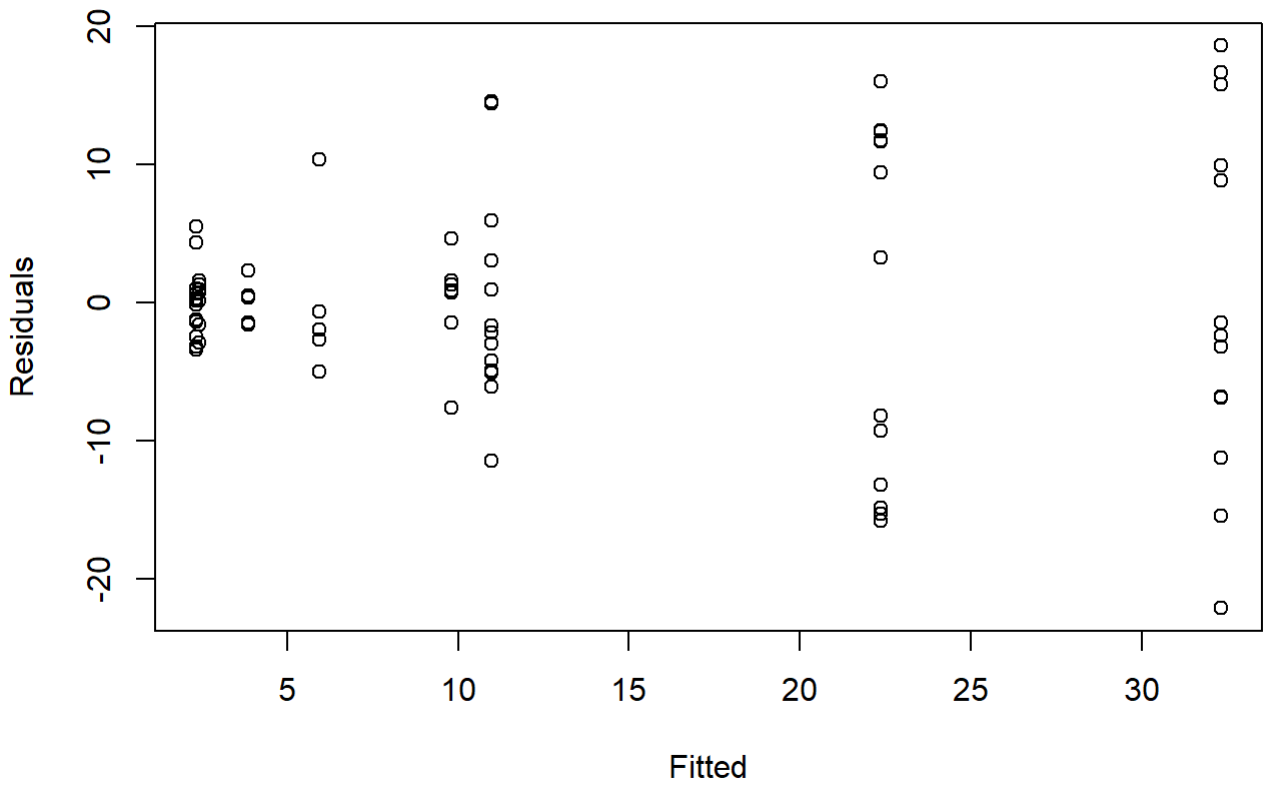
```
shapiro.test(myfit$resid)
```

Shapiro-Wilk normality test

data: myfit\$resid  
W = 0.96425, p-value = 0.03256



```
plot(myfit$fitted,myfit$res,xlab="Fitted",ylab="Residuals")
```



```
leveneTest(myfit)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  7  7.1323 2.307e-06 ***
      67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

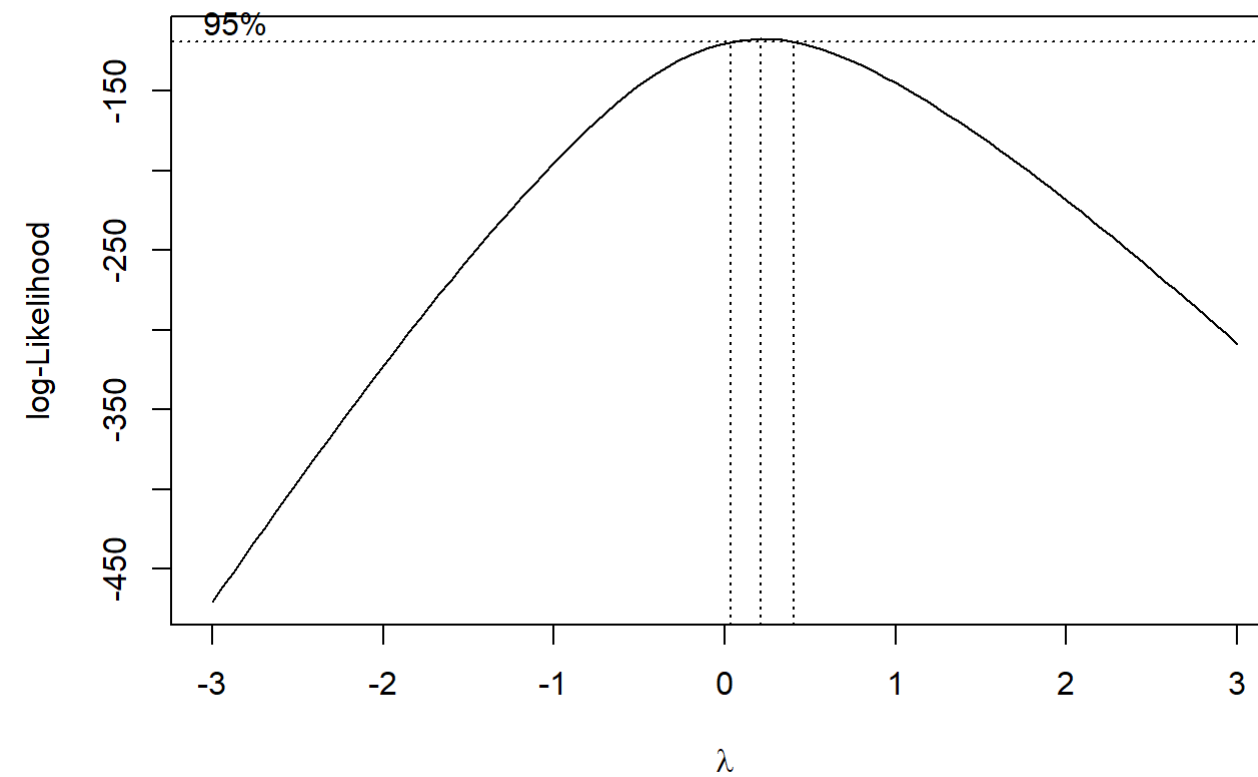
```
outlierTest(myfit)
```

```
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferroni p
58 -2.785452      0.0069699      0.52274
```

```
min(leak) # -1.11
```

```
[1] -1.11
```

```
leak2<- leak+2 #make response positive
myfit2<-lm(leak2~variety*regime*temp, contrasts = c(variety=contr.sum,
regime=contr.sum,temp=contr.sum))
library(MASS)
par(mfrow=c(1,1))
BC <- boxcox(myfit2, lambda = seq(-3, 3, length = 10))
```



```
#----- Full, log10
```

```
log10_myfit<-lm(log10(leak2)~variety*regime*temp,
contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))
library(car)
Anova(log10_myfit,type=3)
```

Anova Table (Type III tests)

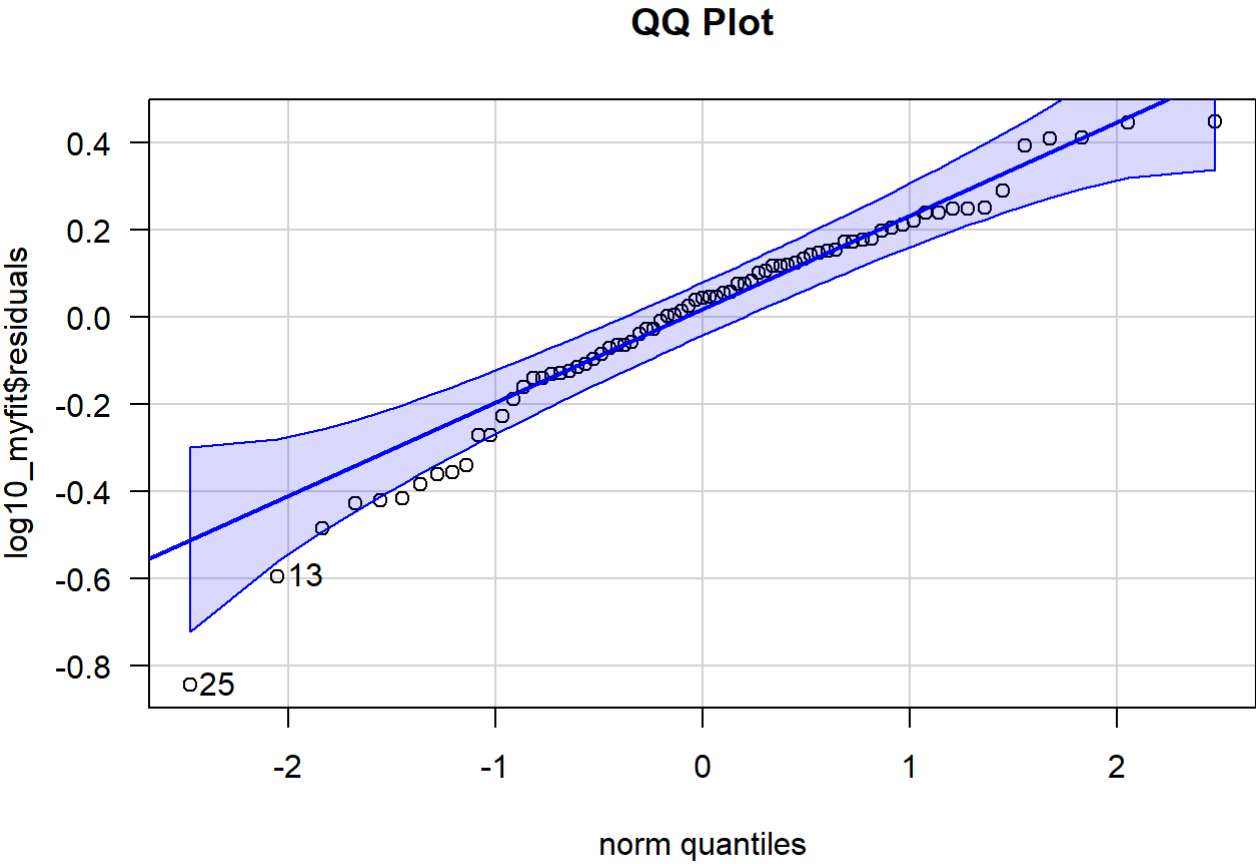
Response: log10(leak2)

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	57.929	1	850.0605	< 2.2e-16	***
variety	1.757	1	25.7774	3.270e-06	***
regime	1.369	1	20.0921	2.958e-05	***
temp	1.351	1	19.8309	3.286e-05	***
variety:regime	1.368	1	20.0795	2.973e-05	***
variety:temp	0.006	1	0.0915	0.76324	
regime:temp	0.452	1	6.6396	0.01218	*
variety:regime:temp	0.029	1	0.4285	0.51498	
Residuals	4.566	67			

---

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

```
qqPlot(log10_myfit$residuals, las = 1, main="QQ Plot")
```



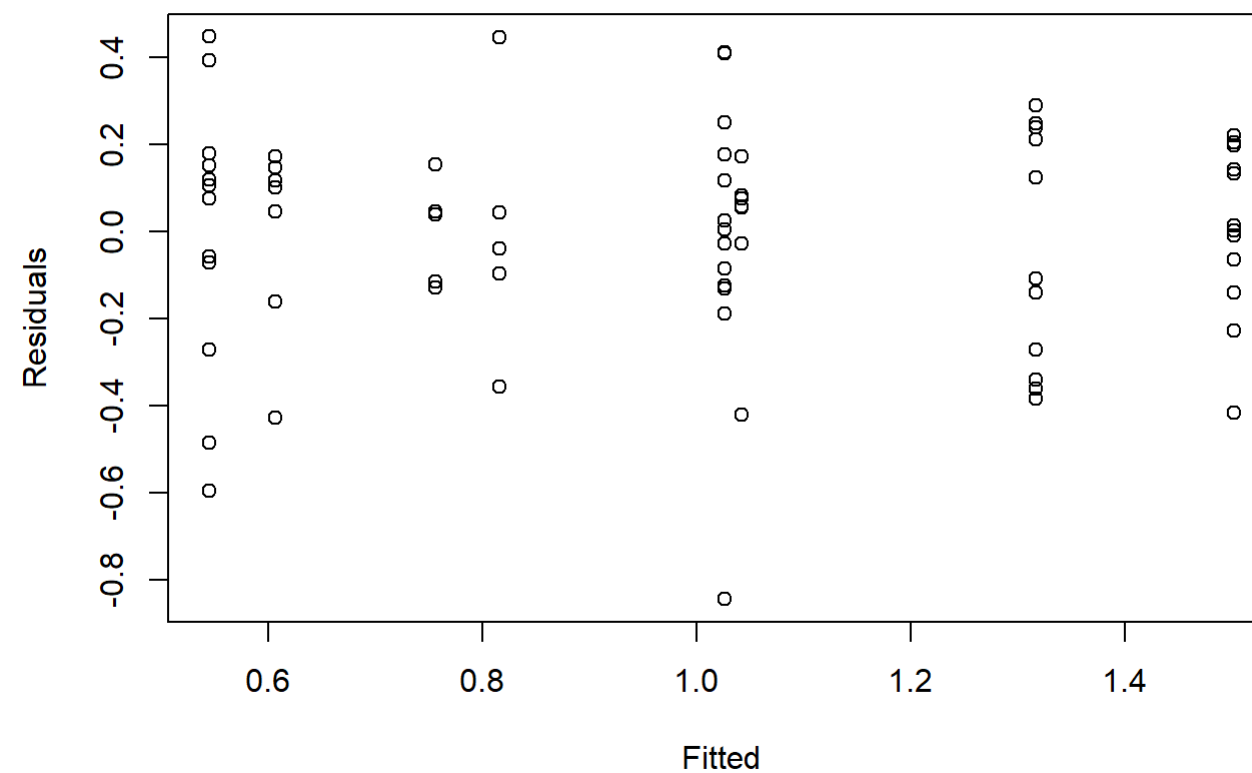
[1] 25 13

```
shapiro.test(log10_myfit$resid)
```

Shapiro-Wilk normality test

```
data: log10_myfit$resid  
W = 0.95957, p-value = 0.01729
```

```
plot(log10_myfit$fitted,log10_myfit$res,xlab="Fitted",ylab="Residuals")
```



```
leveneTest(log10_myfit) #can not be apply for reduced models
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	7	0.8579	0.5442
	67		

```
#perform Breusch-Pagan test
#bptest(log10_myfit)
outlierTest(log10_myfit)
```

```
      rstudent unadjusted p-value Bonferroni p
25 -3.667084      0.00049135      0.036851
#----- Full, (log10(leak2))^2
```

```
## the second transformation
#sq_log10_leak <- (log10(leak2)+1)^0.5 #+ (log10(leak2)+2)^0.5
sq_log10_leak <- (log10(leak2))^(1.5)
sq_log10_myfit<-lm(sq_log10_leak~variety*regime*temp,
  contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))
library(car)
Anova(sq_log10_myfit,type=3)
```

Anova Table (Type III tests)

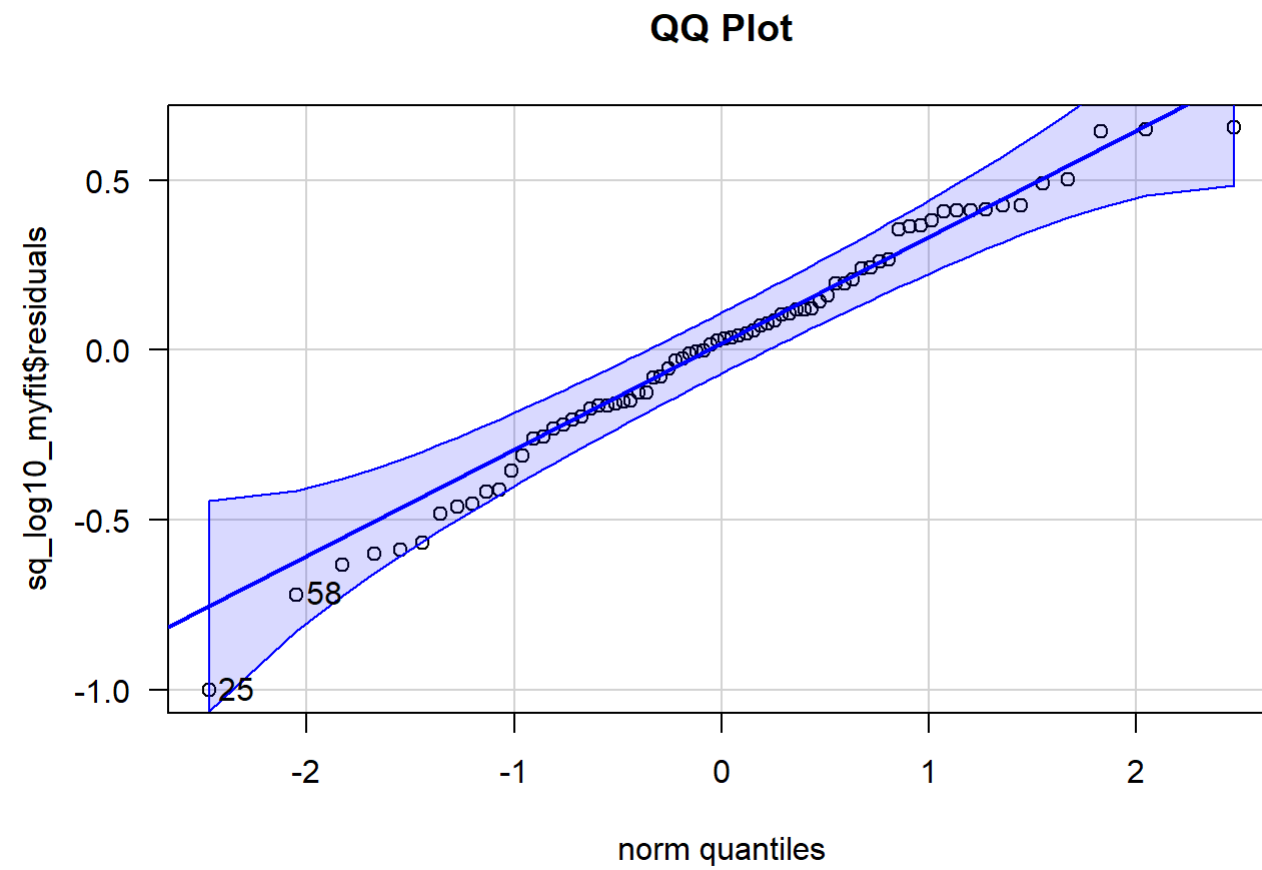
Response: sq\_log10\_leak

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	62.884	1	486.2824	< 2.2e-16 ***
variety	3.782	1	29.2463	9.472e-07 ***
regime	2.743	1	21.2151	1.927e-05 ***
temp	2.487	1	19.2345	4.253e-05 ***
variety:regime	3.830	1	29.6161	8.302e-07 ***
variety:temp	0.046	1	0.3586	0.55134
regime:temp	0.549	1	4.2472	0.04326 *
variety:regime:temp	0.045	1	0.3458	0.55851
Residuals	8.535	66		

---

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

```
qqPlot(sq_log10_myfit$residuals, las = 1, main="QQ Plot")
```



25 58

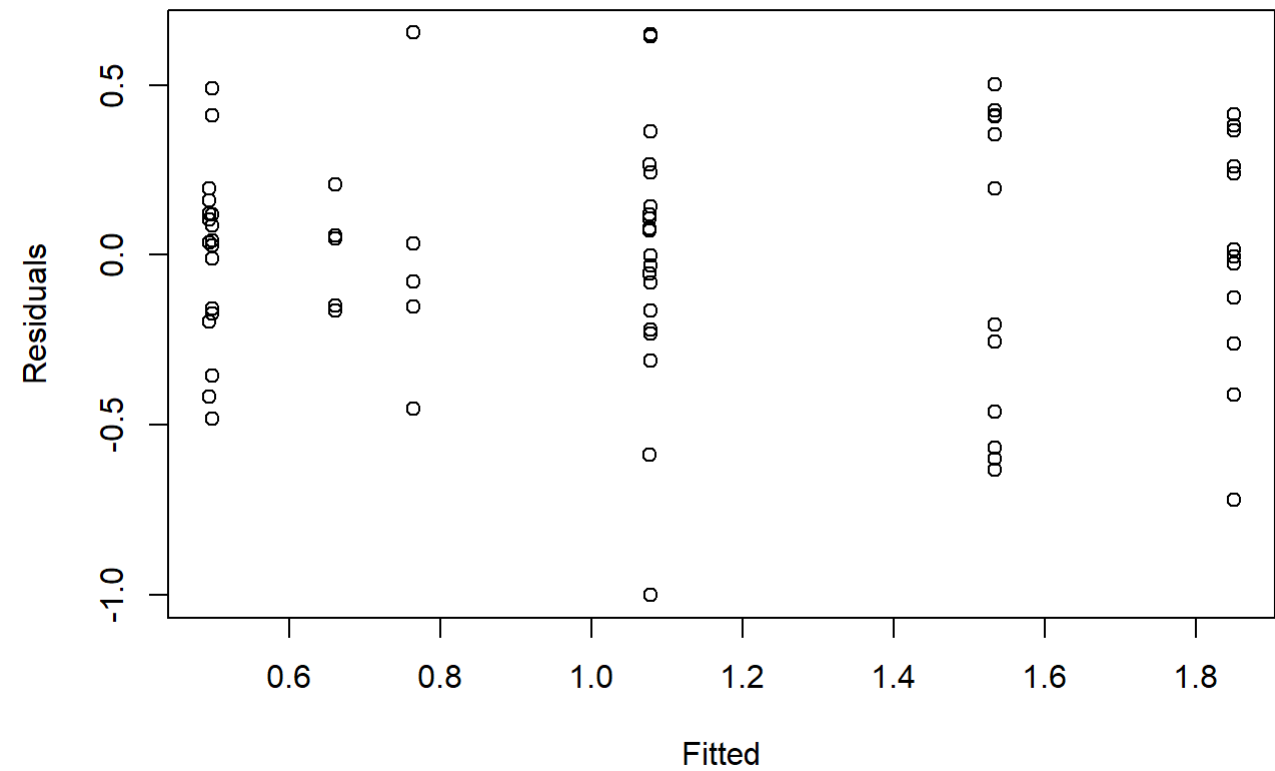
24 57

```
shapiro.test(sq_log10_myfit$resid)
```

Shapiro-Wilk normality test

data: sq\_log10\_myfit\$resid  
W = 0.98333, p-value = 0.4382

```
plot(sq_log10_myfit$fitted,sq_log10_myfit$res,xlab="Fitted",ylab="Residuals")
```



```
leveneTest(sq_log10_myfit)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	7	1.5066	0.1805
	66		

```
#bptest(sq_log10_myfit)
outlierTest(sq_log10_myfit)
```

No Studentized residuals with Bonferroni  $p < 0.05$

Largest  $|rstudent|$ :

	rstudent	unadjusted p-value	Bonferroni p
25	-3.079189	0.0030397	0.22494

#-----

Comparing log10, ^2: (full) vs (full-variety:regime:term)

```
sq_log10_myfit_2 <- lm(sq_log10_leak ~ variety+regime+temp+variety:regime+variety:temp +temp:regime,
  contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))
anova(sq_log10_myfit,sq_log10_myfit_2)
```

### Analysis of Variance Table

Model 1: sq\_log10\_leak ~ variety \* regime \* temp

Model 2: sq\_log10\_leak ~ variety + regime + temp + variety:regime + variety:temp +  
temp:regime

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	66	8.5349				
2	67	8.5796	-1	-0.044716	0.3458	0.5585

According to the p-value log10\_myfit\_2 does not make difference with log10\_myfit

```
Anova(sq_log10_myfit_2,type=3)
```

### Anova Table (Type III tests)

Response: sq\_log10\_leak

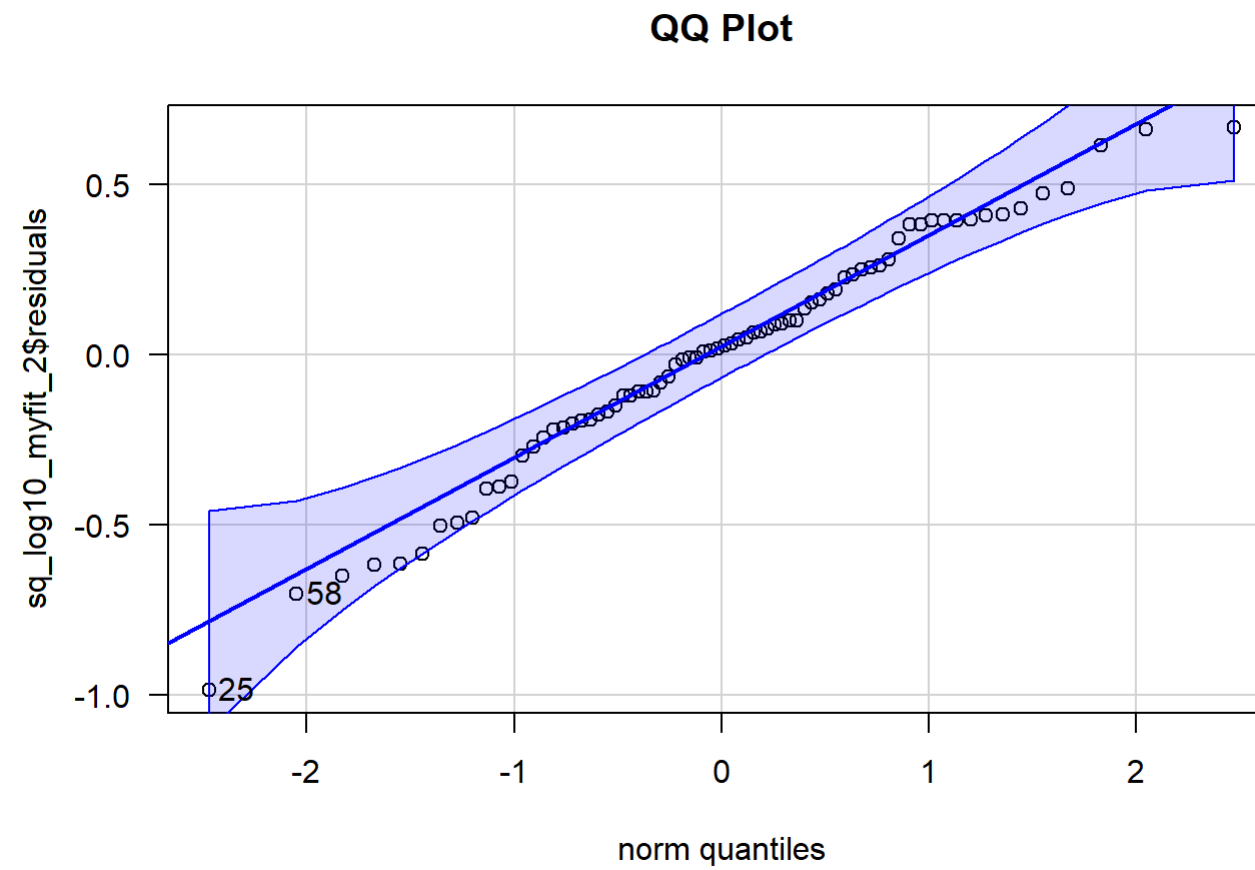
	Sum Sq	Df	F value	Pr(>F)
(Intercept)	62.943	1	491.5359	< 2.2e-16 ***
variety	3.771	1	29.4513	8.526e-07 ***
regime	2.734	1	21.3525	1.790e-05 ***
temp	3.145	1	24.5591	5.173e-06 ***
variety:regime	3.842	1	30.0037	7.000e-07 ***
variety:temp	0.038	1	0.2977	0.58713
regime:temp	0.518	1	4.0487	0.04823 *
Residuals	8.580	67		

---

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

```
qqPlot(sq_log10_myfit_2$residuals, las = 1, main="QQ Plot")
```





25 58

24 57

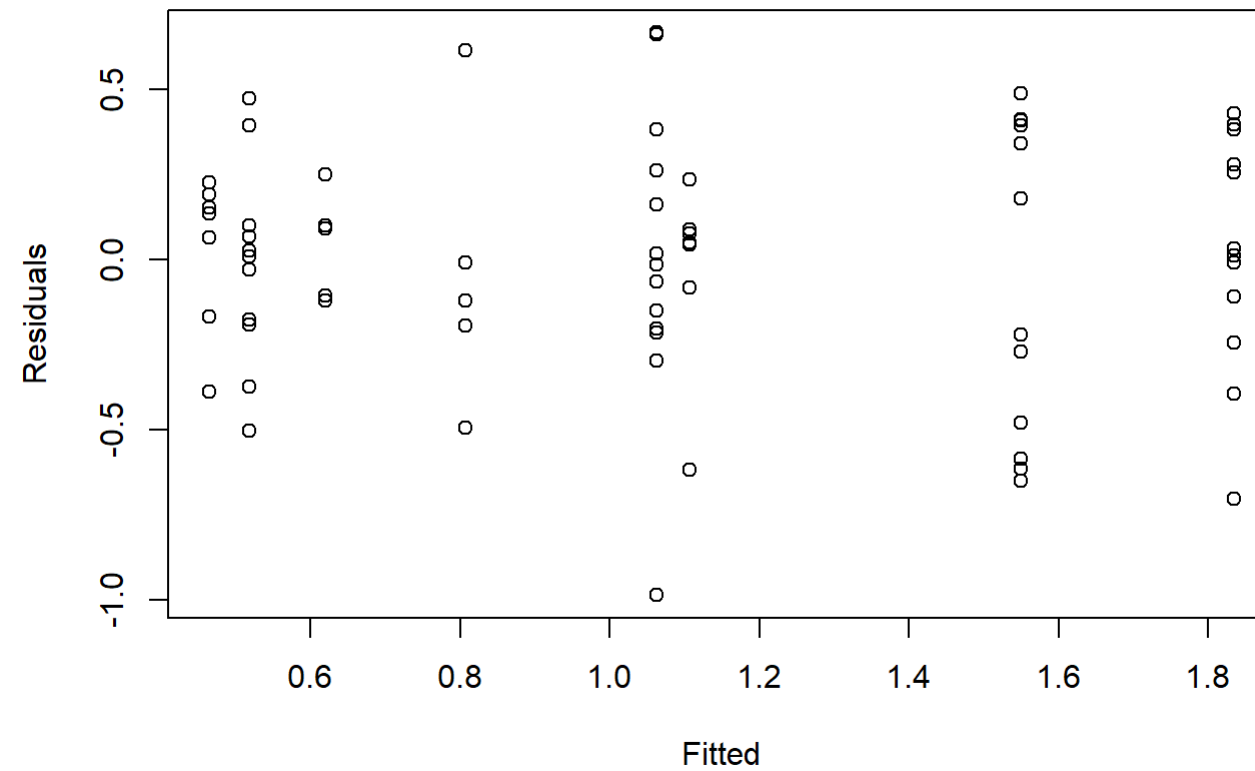
```
shapiro.test(sq_log10_myfit_2$resid)
```

Shapiro-Wilk normality test

data: sq\_log10\_myfit\_2\$resid

W = 0.98187, p-value = 0.3675

```
plot(sq_log10_myfit_2$fitted,sq_log10_myfit_2$res,xlab="Fitted",ylab="Residuals")
```



```
#LeveneTest(log10_myfit) can not be apply for reduced models
#perform Breusch-Pagan test
bptest(sq_log10_myfit_2)
```

studentized Breusch-Pagan test

```
data: sq_log10_myfit_2
BP = 8.6732, df = 6, p-value = 0.1928
```

```
outlierTest(sq_log10_myfit_2)
```

No Studentized residuals with Bonferroni  $p < 0.05$

Largest  $|rstudent|$ :

	rstudent	unadjusted p-value	Bonferroni p
25	-3.02544	0.0035363	0.26169

```
#----- Comparing log10, ^2: (full-variety:regime:term) VS (full- variety:temp -
variety:regime:term)
```

```
sq_log10_myfit_3<- lm(sq_log10_leak~variety+regime+temp+variety:regime +temp:regime,
  contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))
anova(sq_log10_myfit_3,sq_log10_myfit_2)
```

### Analysis of Variance Table

Model 1: sq\_log10\_leak ~ variety + regime + temp + variety:regime + temp:regime

Model 2: sq\_log10\_leak ~ variety + regime + temp + variety:regime + variety:temp +  
temp:regime

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	68	8.6177				
2	67	8.5796	1	0.038124	0.2977	0.5871

According to the p-value log10\_myfit\_3 does not make difference with log10\_myfit\_2

```
Anova(sq_log10_myfit_3,type=3)
```

### Anova Table (Type III tests)

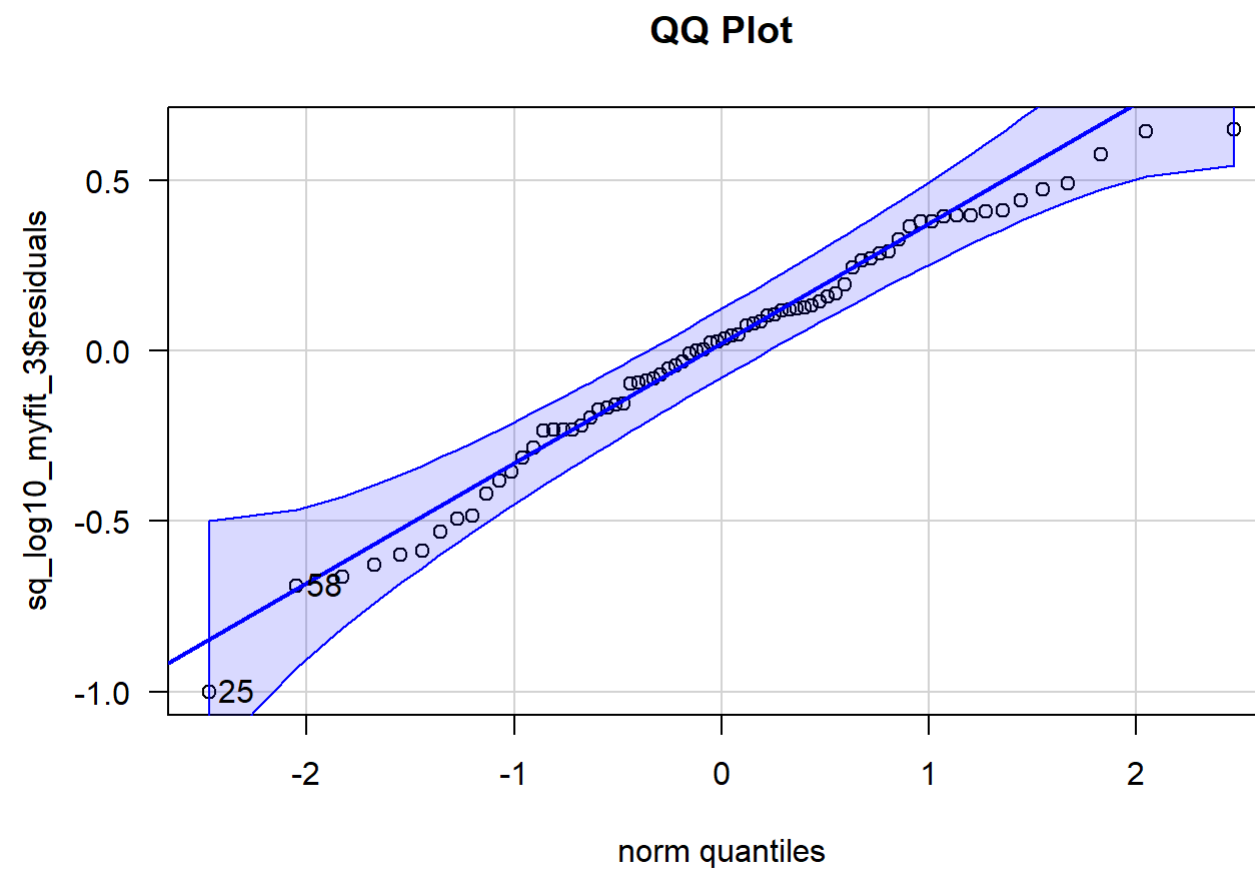
Response: sq\_log10\_leak

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	62.910	1	496.4071	< 2.2e-16 ***
variety	3.784	1	29.8575	7.142e-07 ***
regime	2.745	1	21.6587	1.556e-05 ***
temp	3.245	1	25.6045	3.401e-06 ***
variety:regime	3.831	1	30.2314	6.249e-07 ***
regime:temp	0.484	1	3.8217	0.05471 .
Residuals	8.618	68		

---

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

```
qqPlot(sq_log10_myfit_3$residuals, las = 1, main="QQ Plot")
```



25 58

24 57

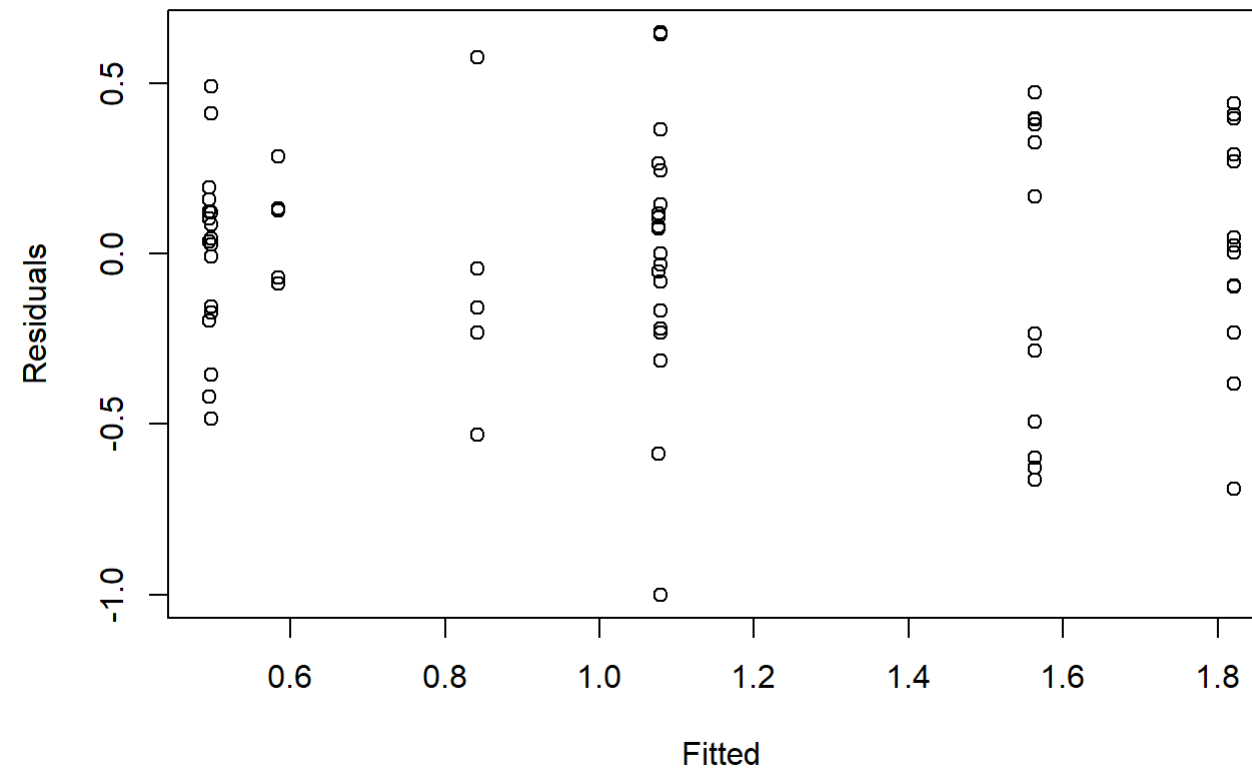
```
shapiro.test(sq_log10_myfit_3$resid)
```

Shapiro-Wilk normality test

data: sq\_log10\_myfit\_3\$resid

W = 0.98, p-value = 0.2905

```
plot(sq_log10_myfit_3$fitted,sq_log10_myfit_3$res,xlab="Fitted",ylab="Residuals")
```



```
#leveneTest(sq_log10_myfit_2) can not be apply for reduced models
#perform Breusch-Pagan test
bptest(sq_log10_myfit_3)
```

studentized Breusch-Pagan test

```
data: sq_log10_myfit_3
BP = 5.9081, df = 5, p-value = 0.3153
```

```
outlierTest(sq_log10_myfit_3)
```

No Studentized residuals with Bonferroni  $p < 0.05$

Largest  $|rstudent|$ :

	rstudent	unadjusted p-value	Bonferroni p
25	-3.085282	0.0029547	0.21865

```
#----- Comparing log10, ^2: (full-variety:regime:term) VS (full- variety:temp -
variety:regime:term-regime:temp )
```

```
sq_log10_myfit_4<- lm(sq_log10_leak~variety+regime+temp+variety:regime,
  contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))
anova(sq_log10_myfit_4,sq_log10_myfit_3)
```

### Analysis of Variance Table

Model 1: sq\_log10\_leak ~ variety + regime + temp + variety:regime

Model 2: sq\_log10\_leak ~ variety + regime + temp + variety:regime + temp:regime

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	69	9.1021				
2	68	8.6177	1	0.48433	3.8217	0.05471 .
---						

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

According to the p-value log10\_myfit\_3 does not make difference with log10\_myfit\_2

```
Anova(sq_log10_myfit_4,type=3)
```

### Anova Table (Type III tests)

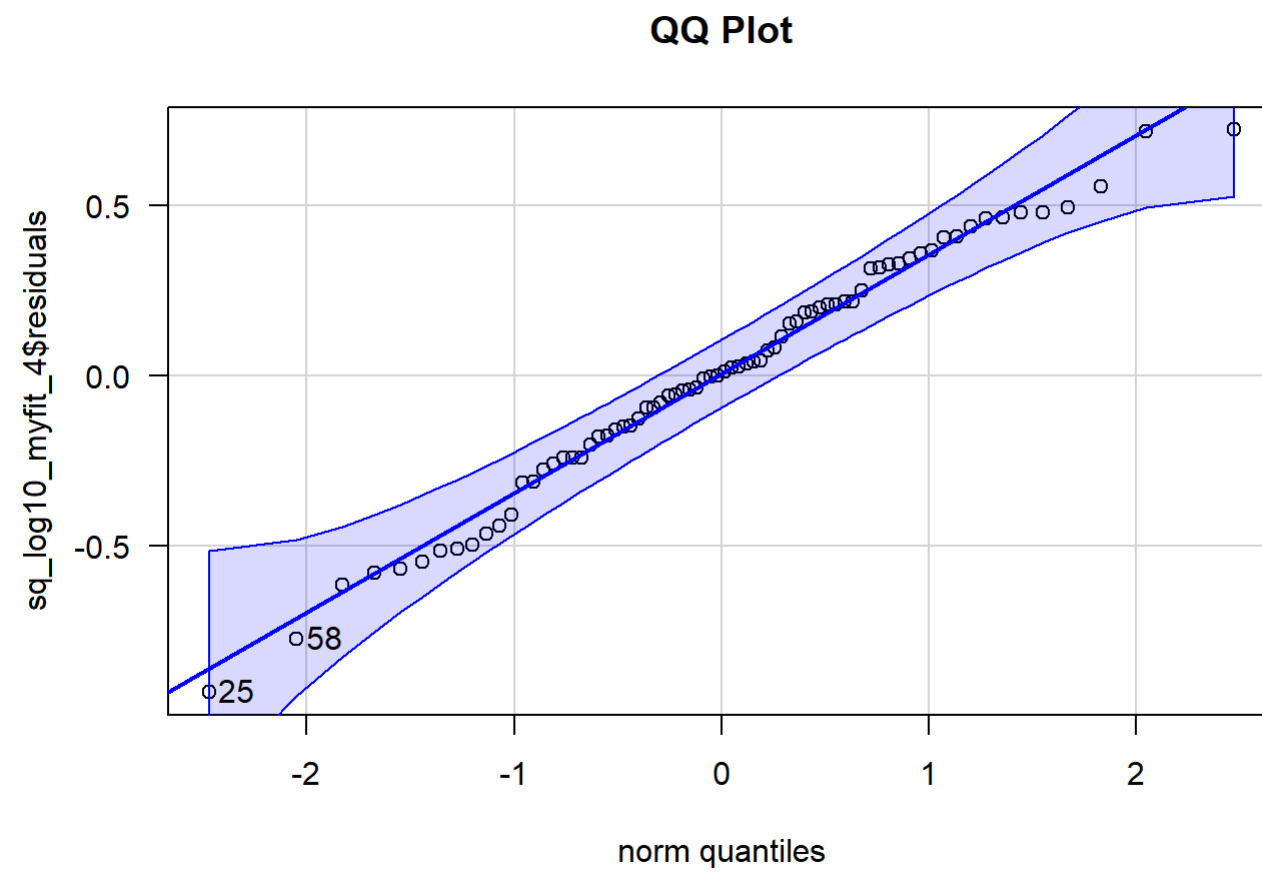
Response: sq\_log10\_leak

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	63.141	1	478.654	< 2.2e-16 ***
variety	3.734	1	28.309	1.214e-06 ***
regime	2.702	1	20.487	2.438e-05 ***
temp	3.309	1	25.088	4.035e-06 ***
variety:regime	3.884	1	29.445	8.032e-07 ***
Residuals	9.102	69		

---

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

```
qqPlot(sq_log10_myfit_4$residuals, las = 1, main="QQ Plot")
```



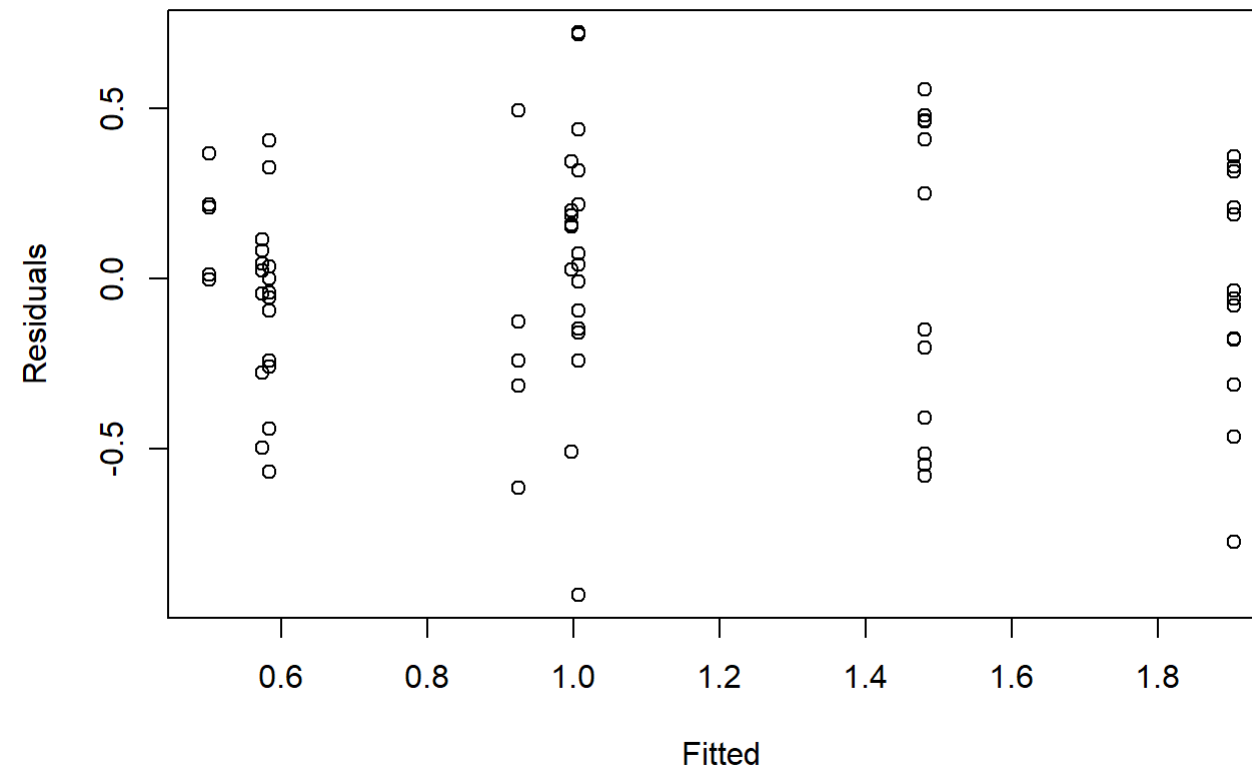
25 58  
24 57

```
shapiro.test(sq_log10_myfit_4$resid)
```

Shapiro-Wilk normality test

data: sq\_log10\_myfit\_4\$resid  
W = 0.98713, p-value = 0.6591

```
plot(sq_log10_myfit_4$fitted,sq_log10_myfit_4$res,xlab="Fitted",ylab="Residuals")
```



```
#leveneTest(sq_log10_myfit_2) can not be apply for reduced models
#perform Breusch-Pagan test
bptest(sq_log10_myfit_4)
```

studentized Breusch-Pagan test

```
data: sq_log10_myfit_4
BP = 3.6537, df = 4, p-value = 0.4549
```

```
outlierTest(sq_log10_myfit_4)
```

No Studentized residuals with Bonferroni  $p < 0.05$

Largest  $|rstudent|$ :

	$rstudent$	unadjusted p-value	Bonferroni p
25	-2.751907	0.0075886	0.56156



# Final Model:

sq\_log10\_leak~variety+regime+temp+variety:regime

```
summary(sq_log10_myfit_4)
```

Call:  
lm(formula = sq\_log10\_leak ~ variety + regime + temp + variety:regime,  
 contrasts = c(variety = contr.sum, regime = contr.sum, temp = contr.sum))

Residuals:

Min	1Q	Median	3Q	Max
-0.92936	-0.23094	0.00605	0.24127	0.72075

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.99654	0.04555	21.878	< 2e-16 ***
variety1	-0.24235	0.04555	-5.321	1.21e-06 ***
regime1	0.20617	0.04555	4.526	2.44e-05 ***
temp1	-0.21172	0.04227	-5.009	4.03e-06 ***
variety1:regime1	-0.24717	0.04555	-5.426	8.03e-07 ***

---

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

Residual standard error: 0.3632 on 69 degrees of freedom  
(1 observation deleted due to missingness)  
Multiple R-squared: 0.6551, Adjusted R-squared: 0.6351  
F-statistic: 32.76 on 4 and 69 DF, p-value: 2.654e-15

# Multiple comparison

```
library(lsmeans)
```

Warning: package 'lsmeans' was built under R version 4.2.3

Loading required package: emmeans

Warning: package 'emmeans' was built under R version 4.2.2

The 'lsmeans' package is now basically a front end for 'emmeans'.  
Users are encouraged to switch the rest of the way.  
See help('transition') for more information, including how to  
convert old 'lsmeans' objects and scripts to work with 'emmeans'.

```
library(multcomp)
```

Loading required package: mvtnorm

Loading required package: survival

Loading required package: TH.data

Attaching package: 'TH.data'

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

geyser

```
comp1<-lsmeans(sq_log10_myfit_4, pairwise ~ variety:regime, adjust="tukey")  
plot(comp1$contrasts)
```

