

Potatoes Cold Climate Adaptation

Sina Mokhtar

2023-03-26

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(car)
```

```
## Loading required package: carData
```

```
##  
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   recode
```

```
library(erikmisc)
```

```
## — Attaching packages ————— erikmisc 0.1.18 —
```

```
## ✓ tibble 3.1.8
```

```
## — Conflicts ————— erikmisc_conflicts() —  
## X dplyr::filter() masks stats::filter()  
## X dplyr::lag()    masks stats::lag()  
## X car::recode()   masks dplyr::recode()
```

```
## erikmisc, solving common complex data analysis workflows  
##   by Dr. Erik Barry Erhardt <erik@StatAcumen.com>
```

```
library(MASS)
```

```
##  
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      select
```

Read Potato Data

```
potato.file=read.table("rawData/potato.txt",header=TRUE)
```

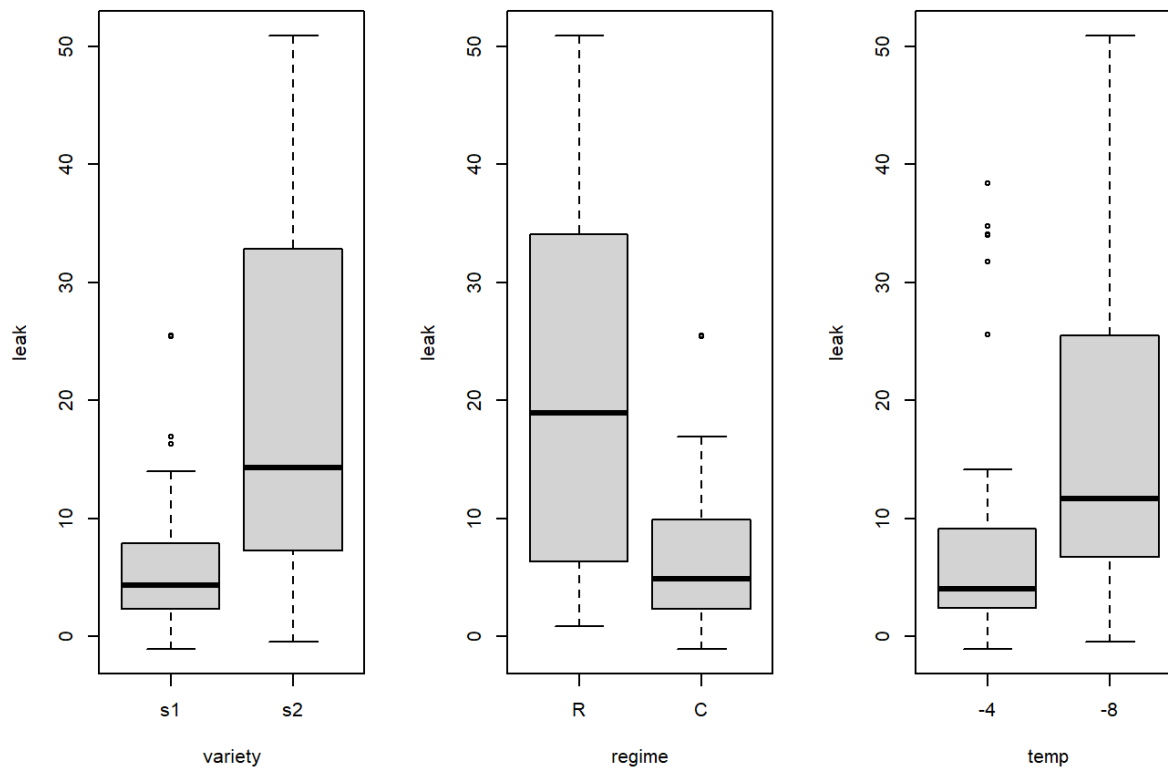
Clean Data

```
potato = potato.file %>%  
  dplyr::mutate(regime = factor(regime,label=c("R","C")),  
               variety = factor(variety,label=c("s1","s2")),  
               temp = factor(temp, label=c("-4","-8")),  
               id = 1:n())  
  
attach(potato)  
head(potato)
```

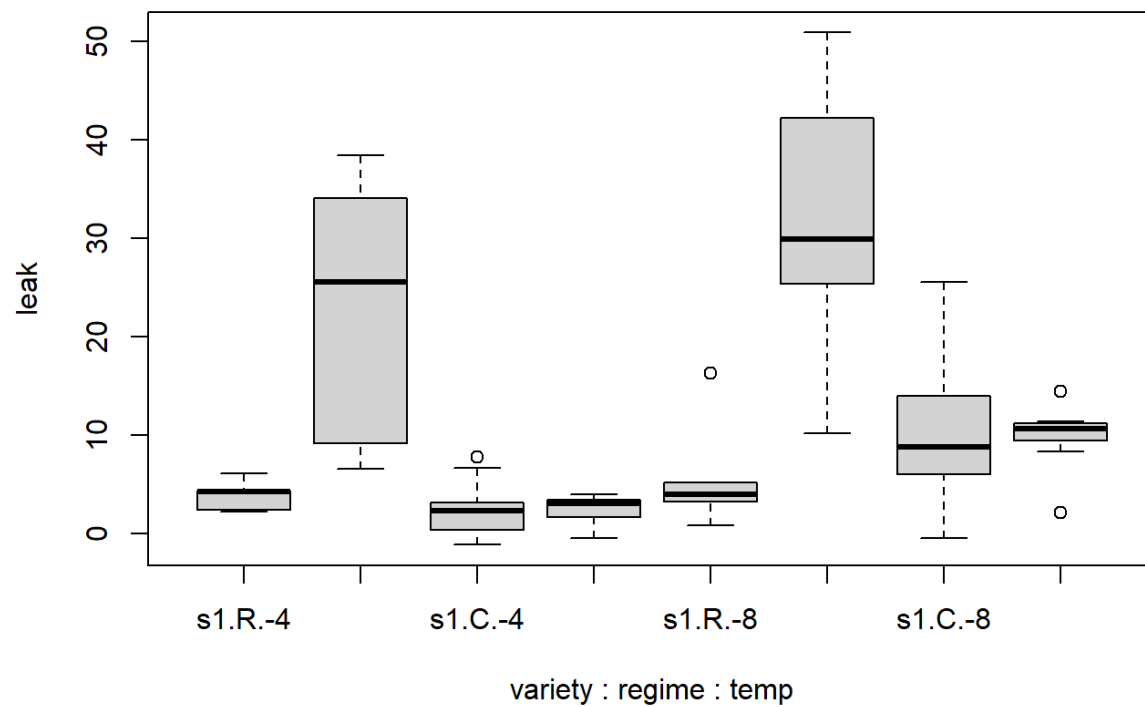
	variety <fct>	regime <fct>	temp <fct>	photo <dbl>	leak <dbl>	id <int>
1	s1	R	-4	4.6386	2.25	1
2	s1	R	-4	5.7914	4.34	2
3	s1	R	-4	29.3515	4.25	3
4	s1	R	-4	18.4173	6.14	4
5	s1	R	-4	2.2556	2.38	5
6	s1	R	-8	8.2358	16.30	6

6 rows

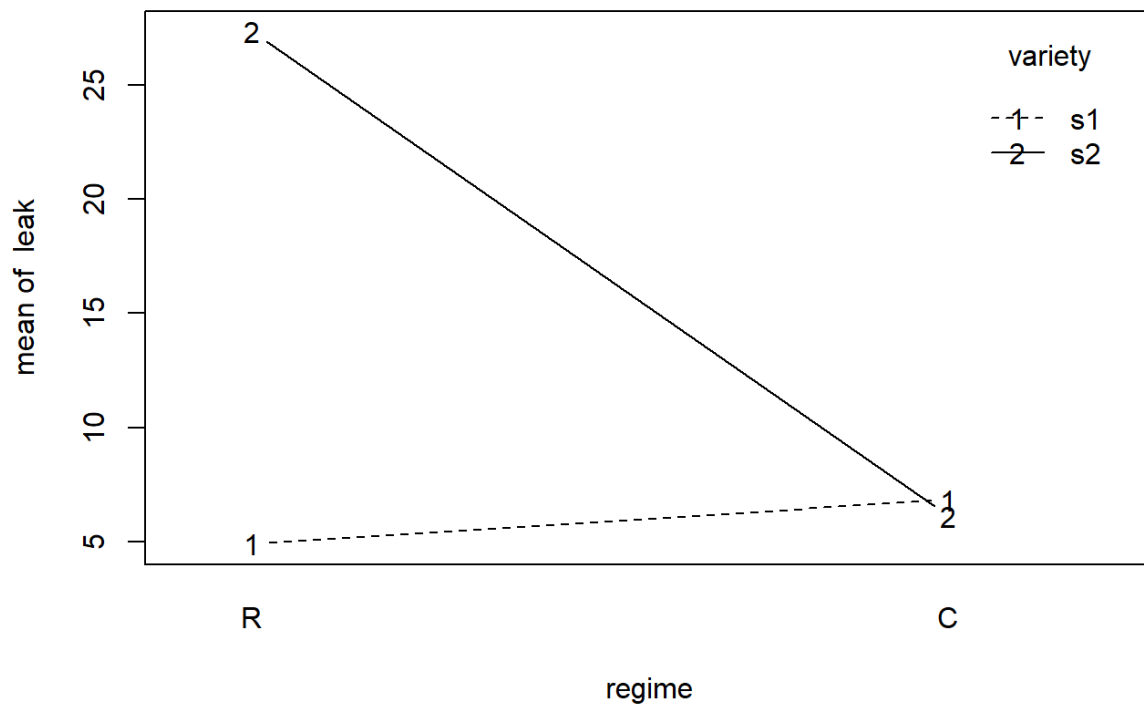
```
par(mfrow=c(1,3))  
boxplot(leak ~ variety,  
data = potato)  
boxplot(leak ~ regime,  
data = potato)  
boxplot(leak ~ temp,  
data = potato)
```



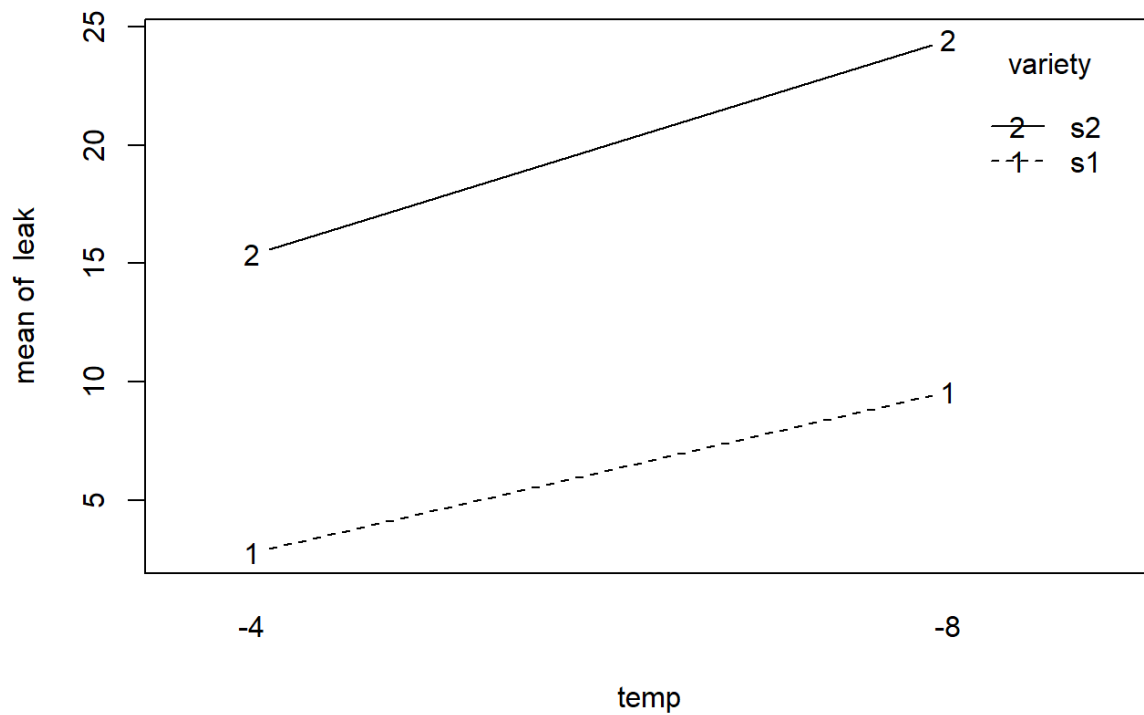
```
boxplot(leak ~ variety:regime:temp,
data = potato)
```



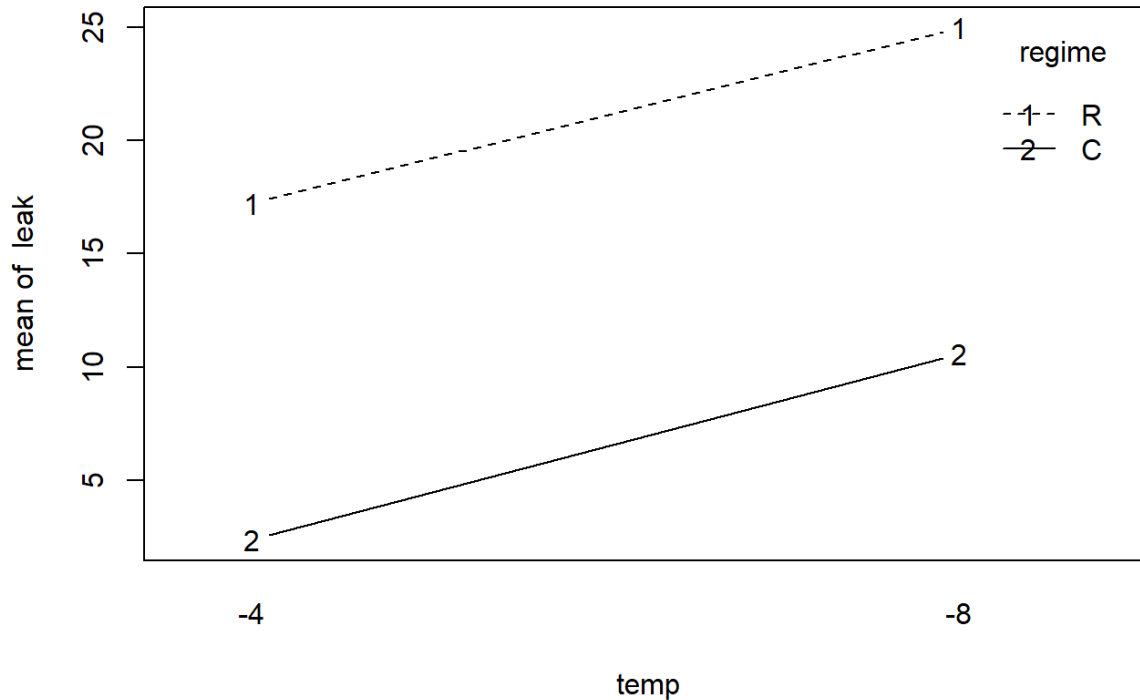
```
interaction.plot(regime, variety, leak, type="b")
```



```
interaction.plot(temp, variety, leak, type="b")
```



```
interaction.plot(temp, regime, leak, type="b")
```

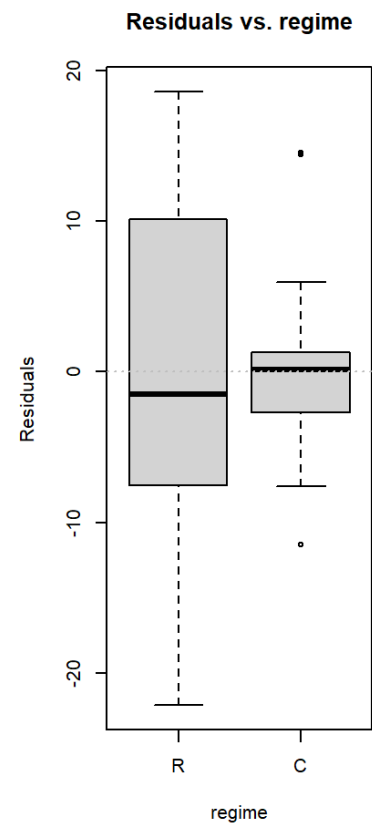
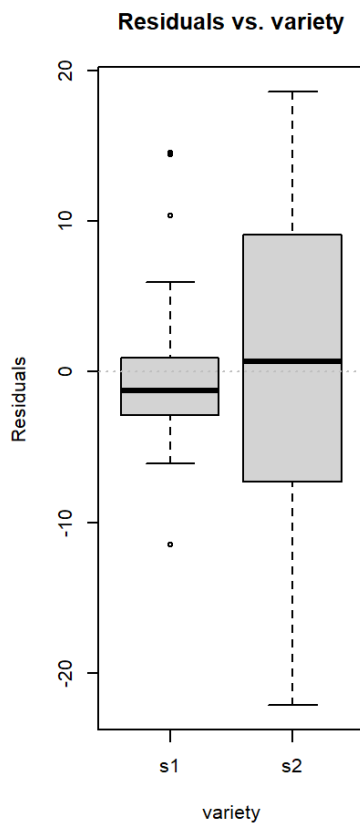
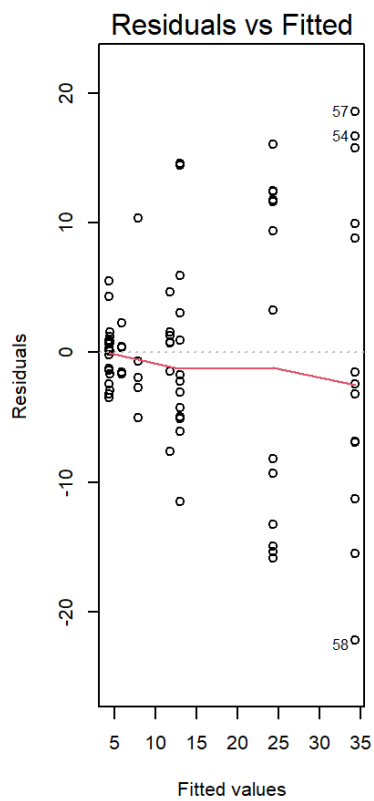
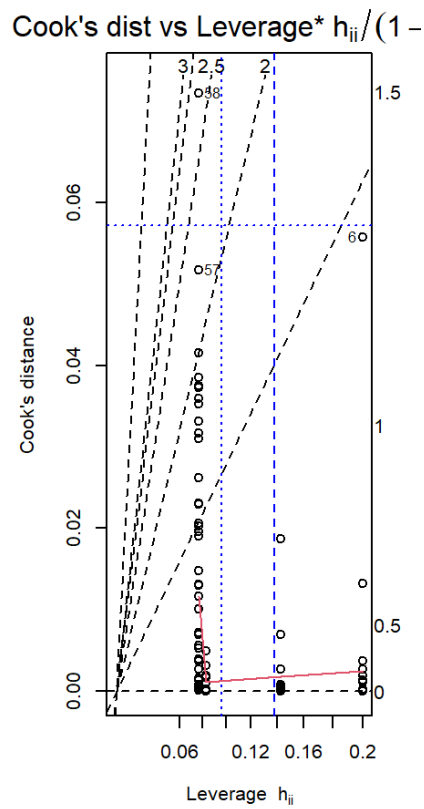
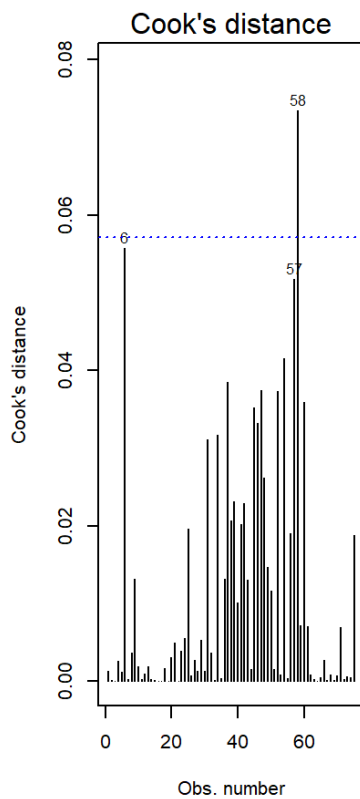
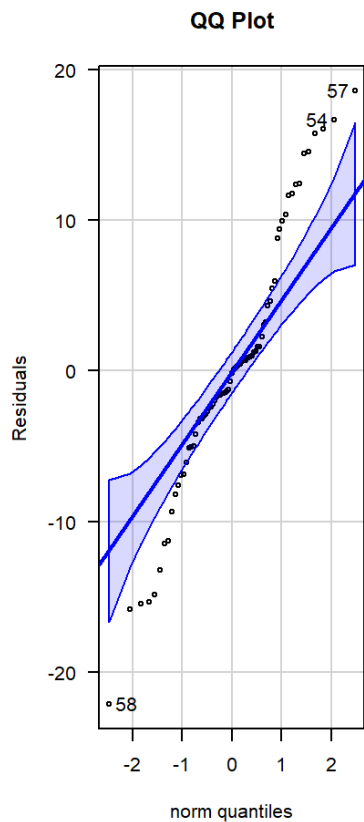


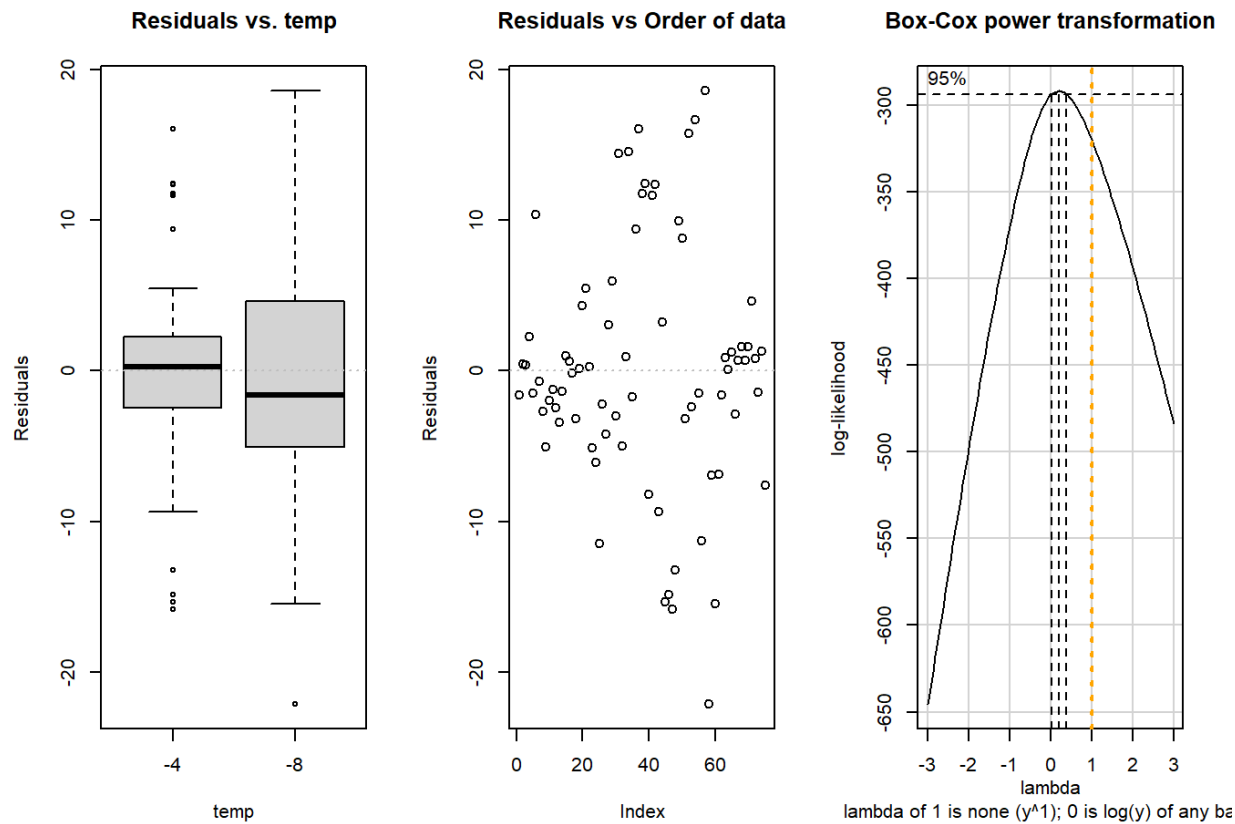
Fit the model

```
potato = potato %>%  
  dplyr::mutate(leak_positive = leak + 2)  
attach(potato)
```

```
## The following objects are masked from potato (pos = 3):  
##  
## id, leak, photo, regime, temp, variety
```

```
myfit<-lm(leak_positive~variety*regime*temp,  
  contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))  
e_plot_lm_diagnostics(myfit)
```



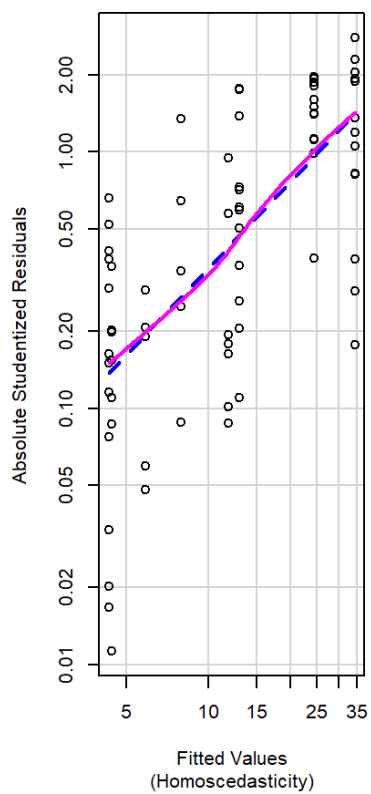


```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 31.92935, Df = 1, p = 1.5988e-08
```

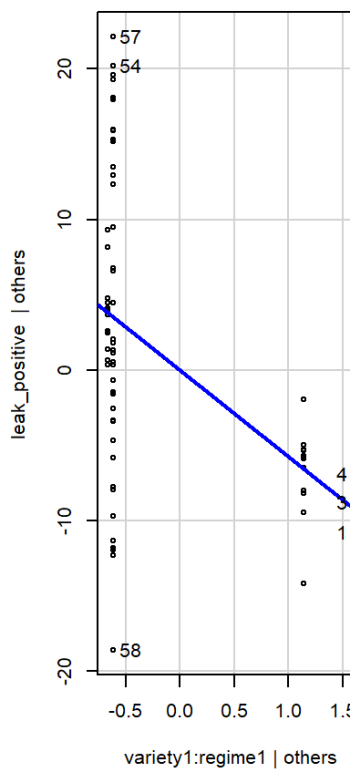
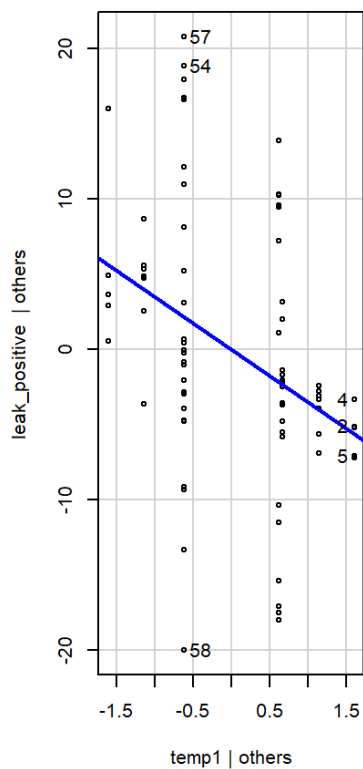
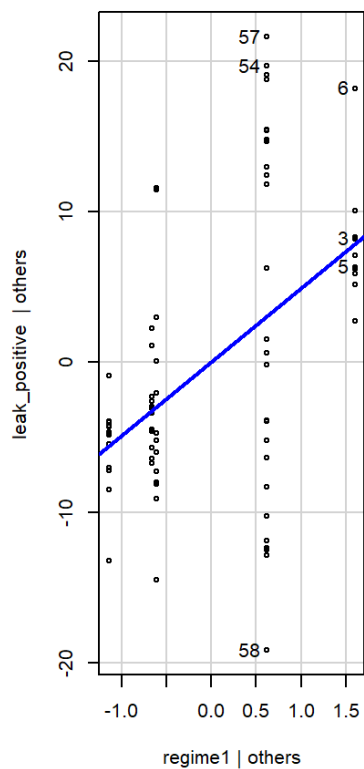
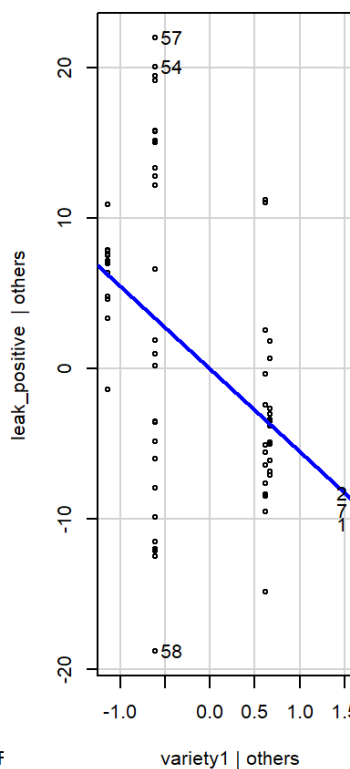
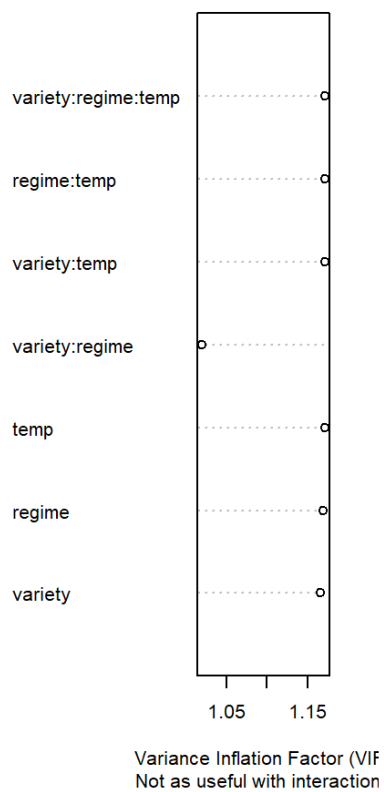
```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

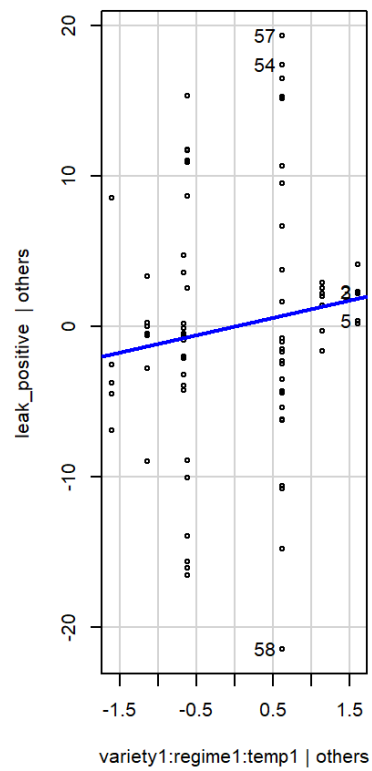
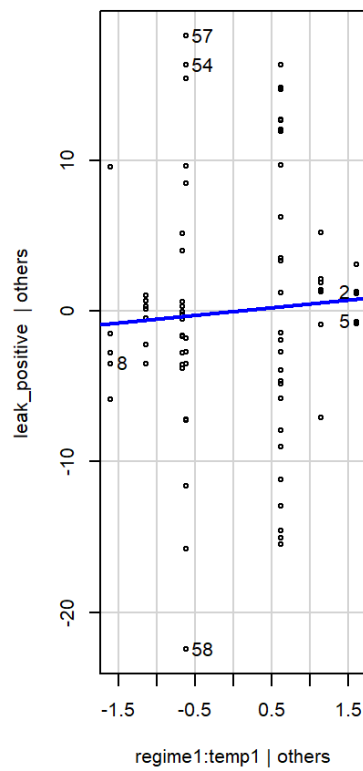
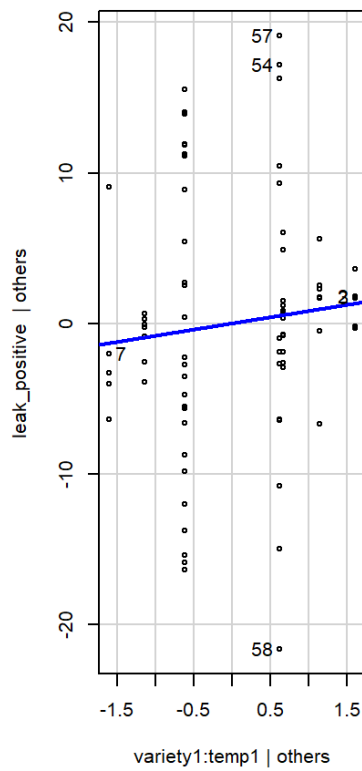
```
## Warning in e_plot_lm_diagnostics(myfit): Note: Collinearity plot unreliable for
## predictors that also have interactions in the model.
```

Spread-Level Plot for fit

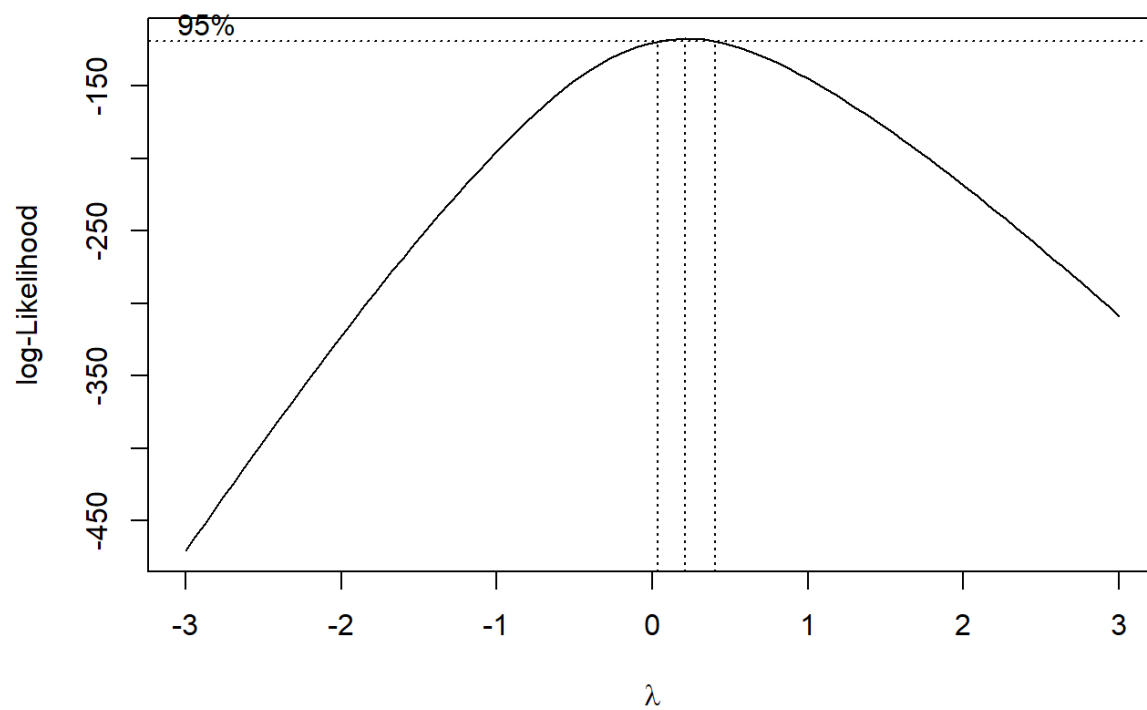


Collinearity





```
boxcox = boxcox(myfit, lambda = seq(-3, 3, length = 10))
```



```
# find optimize Lambda
```

```
optlambda = boxcox$x[which.max(boxcox$y)]  
optlambda
```

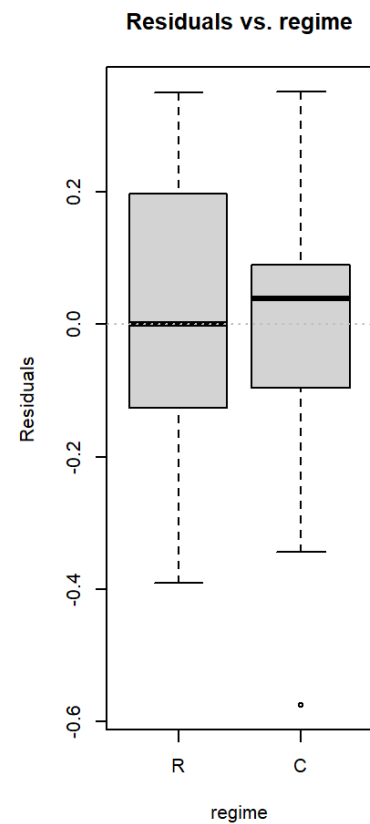
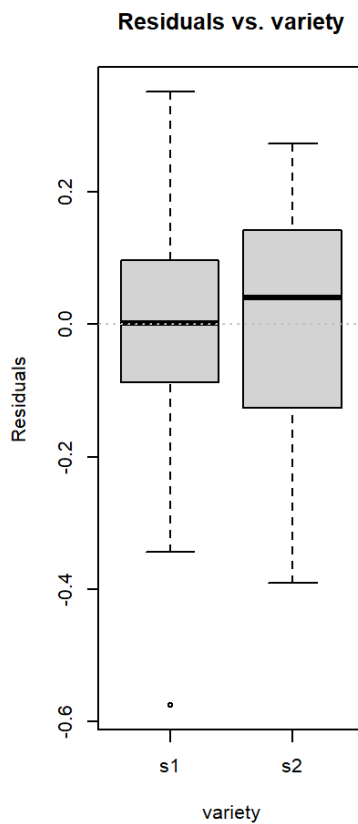
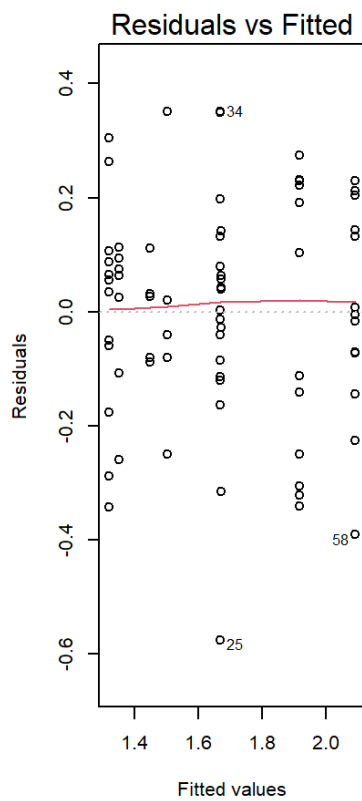
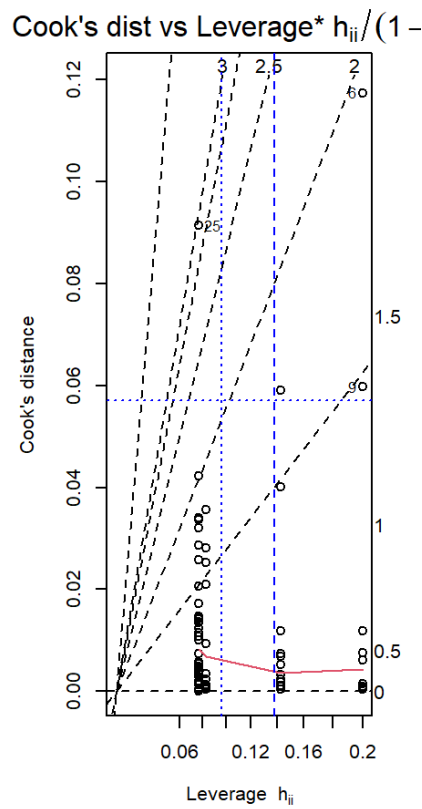
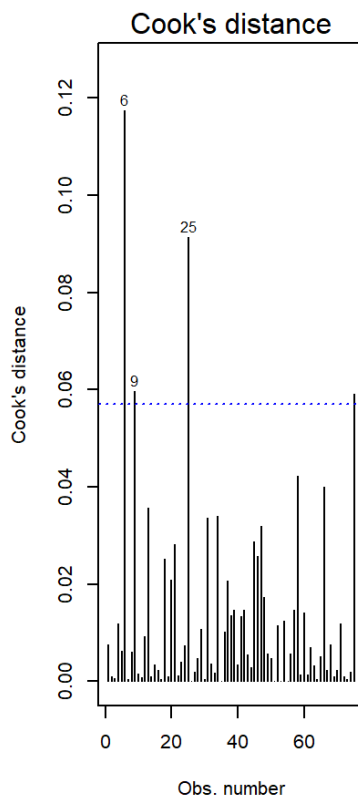
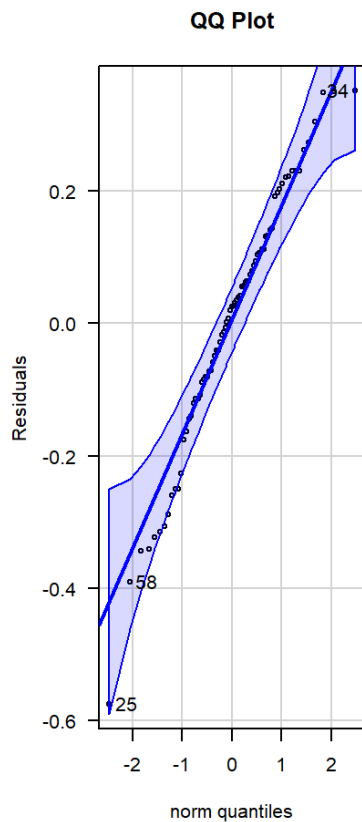
```
## [1] 0.2121212
```

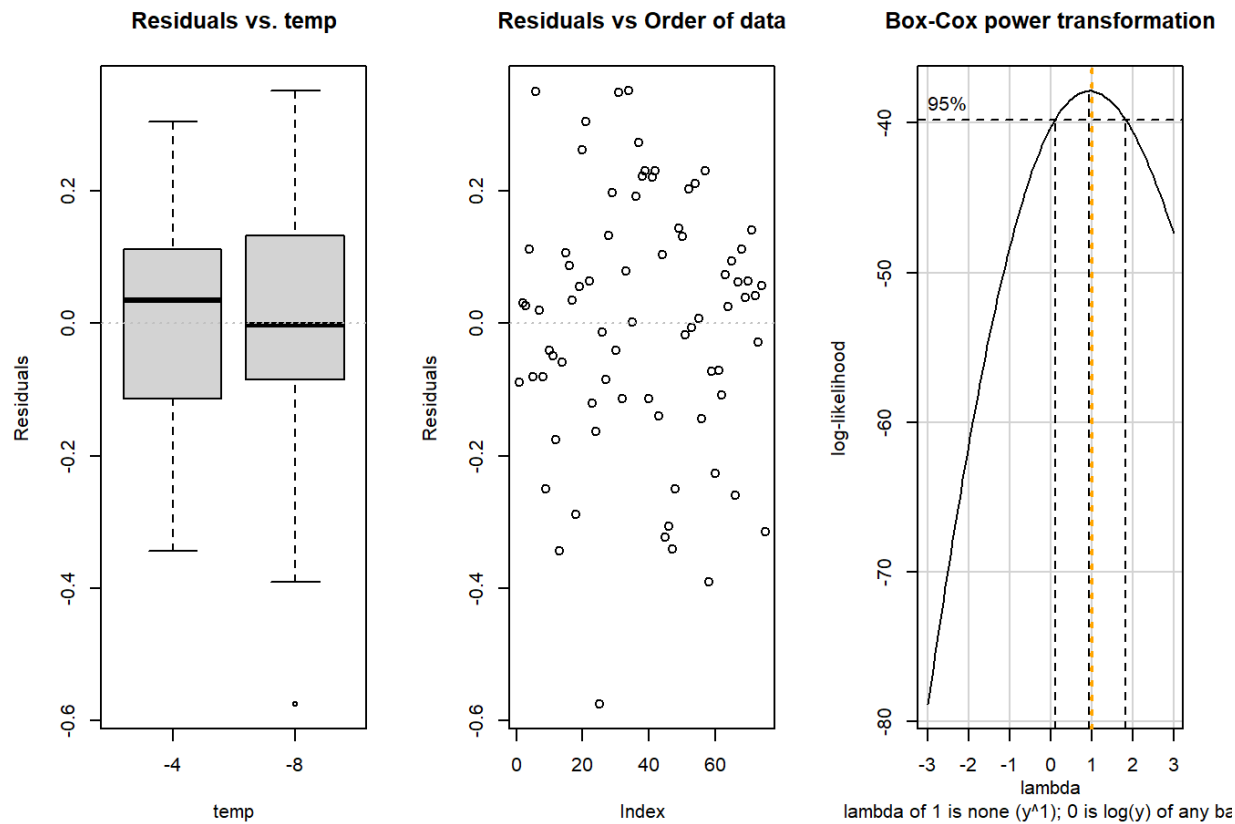
```
potato = potato %>%  
  dplyr::mutate(leak_trans = leak_positive^optlambda)  
attach(potato)
```

```
## The following objects are masked from potato (pos = 3):  
##  
##    id, leak, leak_positive, photo, regime, temp, variety
```

```
## The following objects are masked from potato (pos = 4):  
##  
##    id, leak, photo, regime, temp, variety
```

```
myfit_trans<-lm(leak_trans~variety*regime*temp,  
  contrasts = c(variety=contr.sum, regime=contr.sum,temp=contr.sum))  
e_plot_lm_diagnostics(myfit_trans)
```

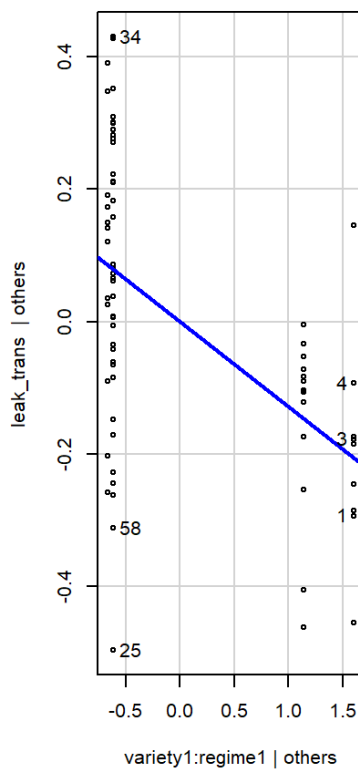
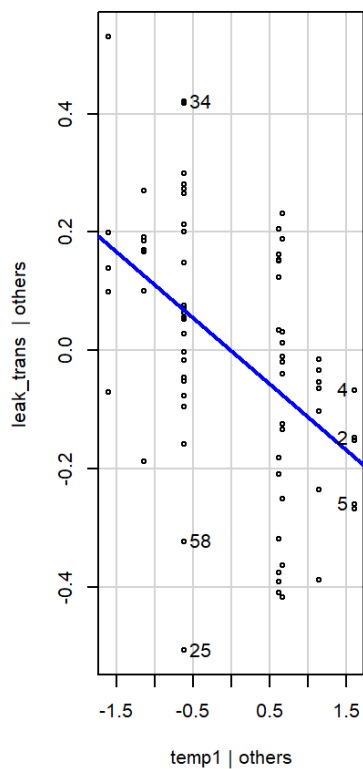
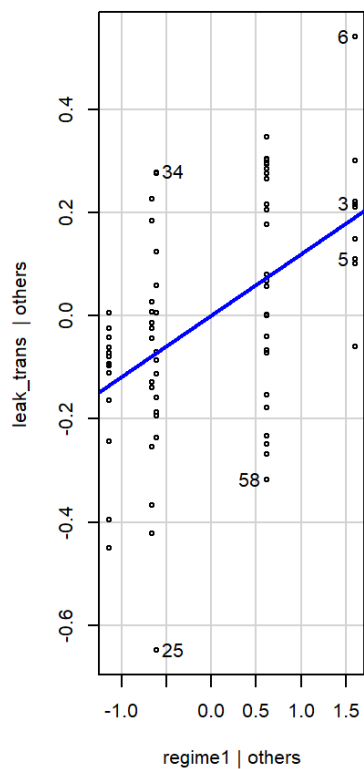
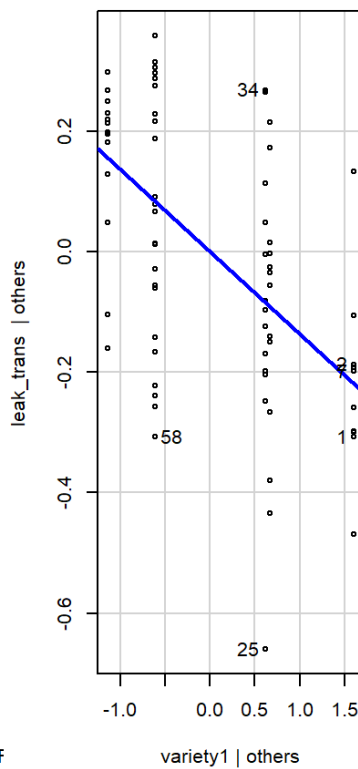
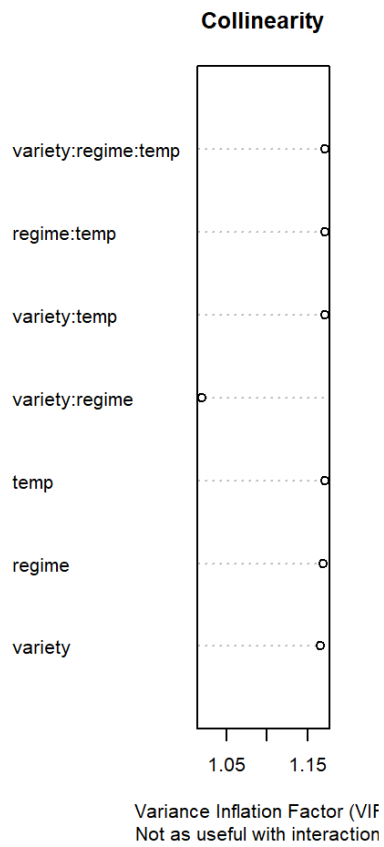
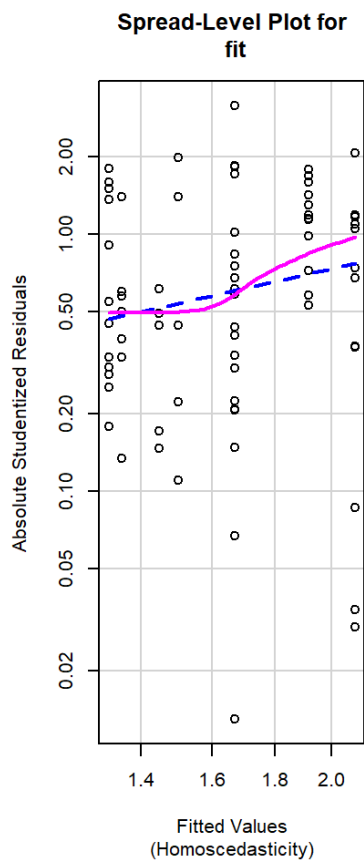


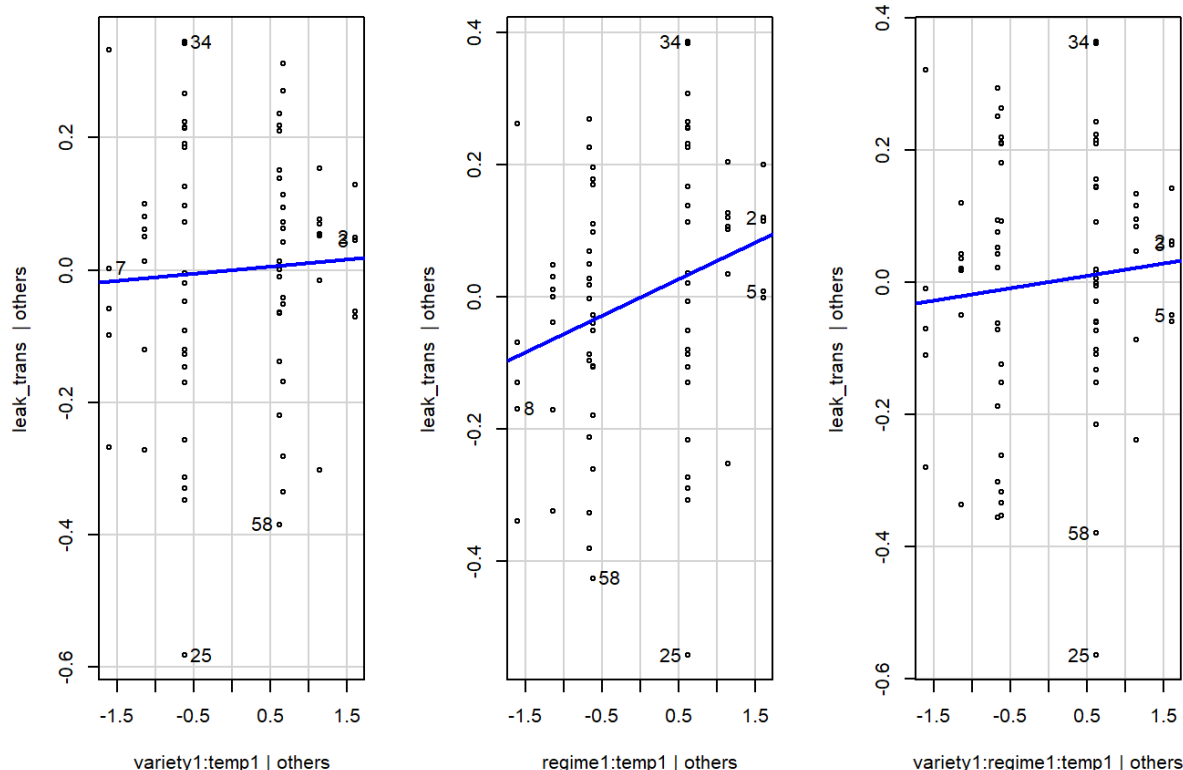


```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.9601756, Df = 1, p = 0.32714
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
## Warning in e_plot_lm_diagnostics(myfit_trans): Note: Collinearity plot
## unreliable for predictors that also have interactions in the model.
```





```
#outlier test
outlierTest(myfit_trans)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferroni p
## 25 -3.152623      0.0024345      0.18259
```

observation 25 has a standardized residual (rstudent) of -3.152623, and the Bonferroni p-value for observation 25 is 0.18259, which is greater than the conventional threshold of 0.05. This suggests that observation 25 is not statistically significant after correcting for multiple testing. Overall, observation 25 may be a potential outlier or influential observation but we keep it in model.

```
# violation of constant variance observed, not serious
leveneTest(leak_trans~variety*temp*regime, data=potato)
```

	Df <int>	F value <dbl>	Pr(>F) <dbl>
group	7	1.281869	0.2727386
	67	NA	NA
2 rows			

Since this p-value is greater than the typical significance level of 0.05, we fail to reject the null hypothesis of equal variances across all groups. We can conclude that there is not enough evidence to suggest that the variances of these three groups are different.

```
# shapiro test residuals normality
shapiro.test(myfit_trans$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data:  myfit_trans$residuals
## W = 0.97926, p-value = 0.2559
```

Since the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis that the residuals is normally distributed. Therefore, we can conclude that the residuals is normally distributed.

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

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
(Intercept)	1.681818e+02	1	4104.9696312	7.634365e-62
variety	1.191608e+00	1	29.0846807	9.724994e-07
regime	9.017248e-01	1	22.0092426	1.382549e-05
temp	8.003673e-01	1	19.5353144	3.702920e-05
variety:regime	1.049885e+00	1	25.6255205	3.460793e-06
variety:temp	7.774686e-03	1	0.1897640	6.645133e-01
regime:temp	1.960192e-01	1	4.7844249	3.221236e-02
variety:regime:temp	2.254935e-02	1	0.5503831	4.607540e-01
Residuals	2.745009e+00	67	NA	NA
9 rows				

The results suggest that all predictors except variety:temp:regime are significant at the conventional alpha level of 0.05. Specifically, the intercept, variety, regime, temp, and their two-way interactions (variety:regime, variety:temp, and regime:temp) have significant effects on leak_trans. However, the three-way interaction variety:temp:regime is not significant. therefore we reduce the model and remove three way interaction.

```
fit_reduced = update(myfit_trans, . ~ . -variety:regime:temp)
Anova(fit_reduced, type=3)
```

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
(Intercept)	1.682136e+02	1	4133.0758793	1.254803e-62
variety	1.189555e+00	1	29.2278479	8.955361e-07
regime	8.999343e-01	1	22.1117420	1.301744e-05
temp	1.048143e+00	1	25.7533007	3.216071e-06
variety:regime	1.051902e+00	1	25.8456466	3.106525e-06
variety:temp	5.241004e-03	1	0.1287735	7.208189e-01

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
regime:temp	1.829625e-01	1	4.4954614	3.763625e-02
Residuals	2.767558e+00	68	NA	NA
8 rows				

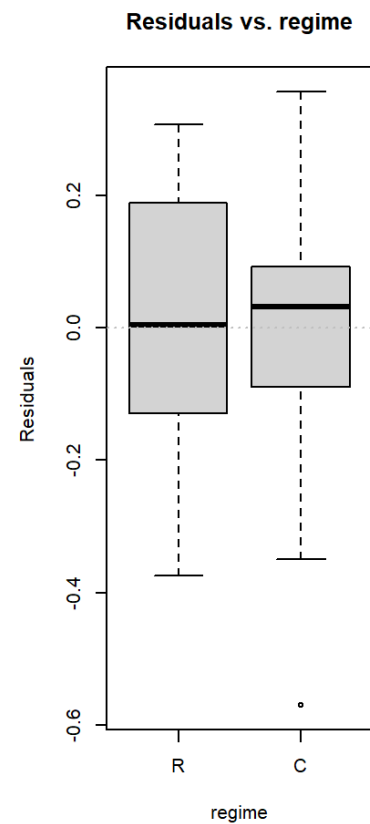
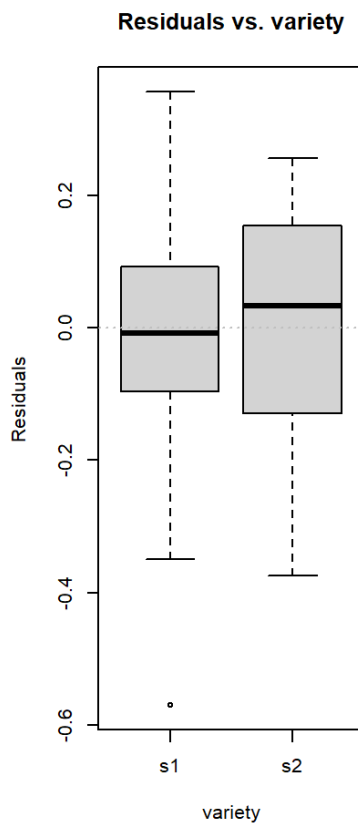
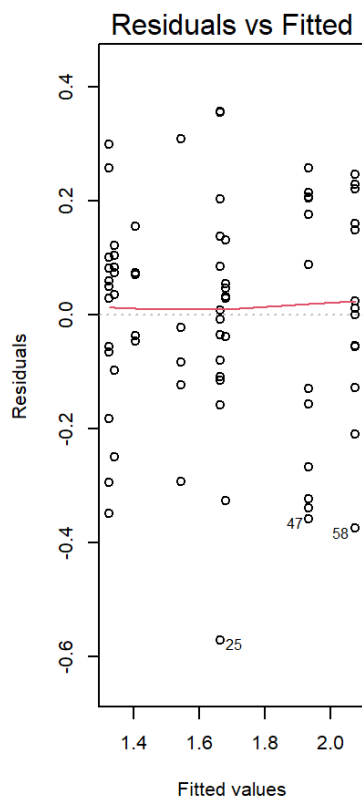
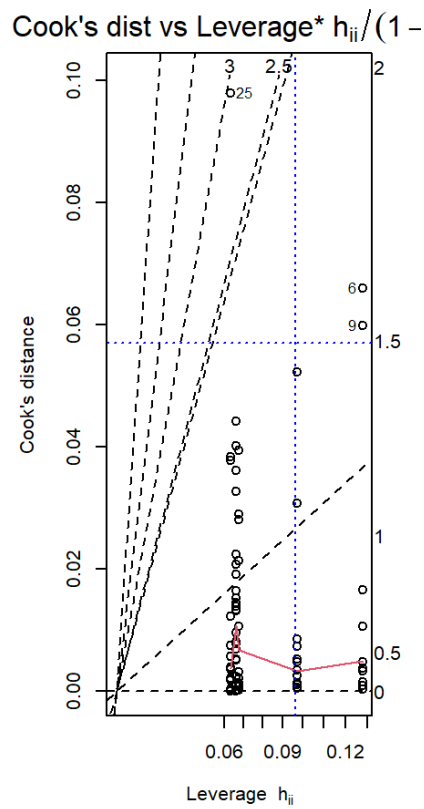
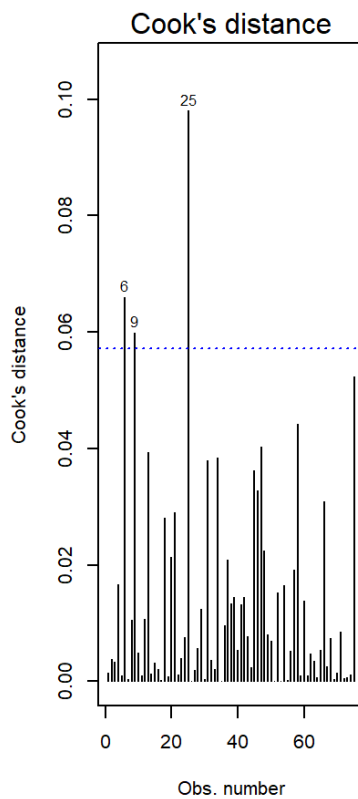
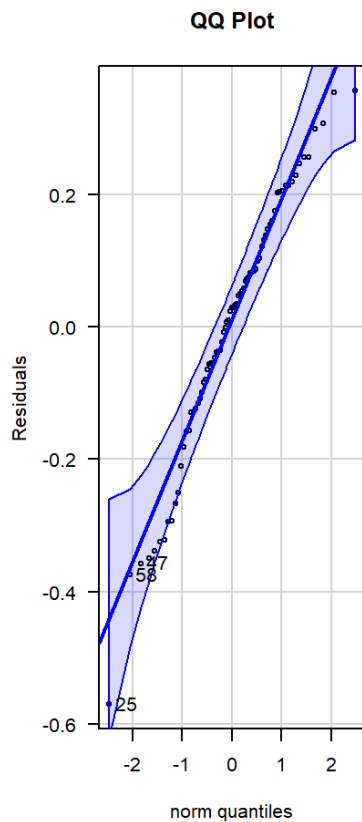
The interaction term variety:temp in the ANOVA table has a p-value of 0.72082, which is greater than the conventional significance level of 0.05. Therefore, we do not have evidence to reject the null hypothesis that the interaction term is not significant. So, variety:temp is not significant

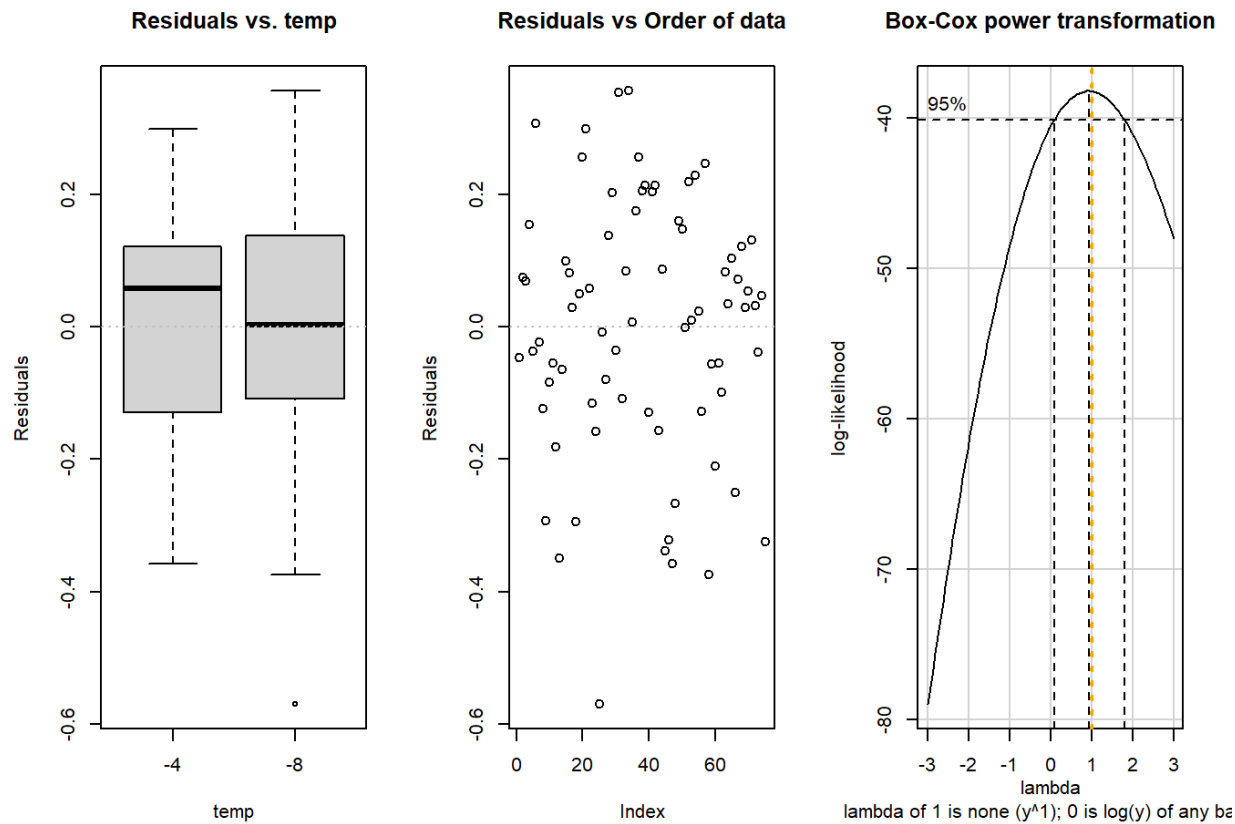
```
fit_reduced = update(fit_reduced, . ~ . -variety:temp)
Anova(fit_reduced, type=3)
```

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
(Intercept)	168.2088127	1	4185.809202	1.691695e-63
variety	1.1907448	1	29.631209	7.509585e-07
regime	0.9009619	1	22.420079	1.130676e-05
temp	1.0692892	1	26.608835	2.275807e-06
variety:regime	1.0508948	1	26.151097	2.701029e-06
regime:temp	0.1856692	1	4.620304	3.510467e-02
Residuals	2.7727991	69	NA	NA
7 rows				

```
finalfit = fit_reduced
```

```
e_plot_lm_diagnostics(finalfit)
```

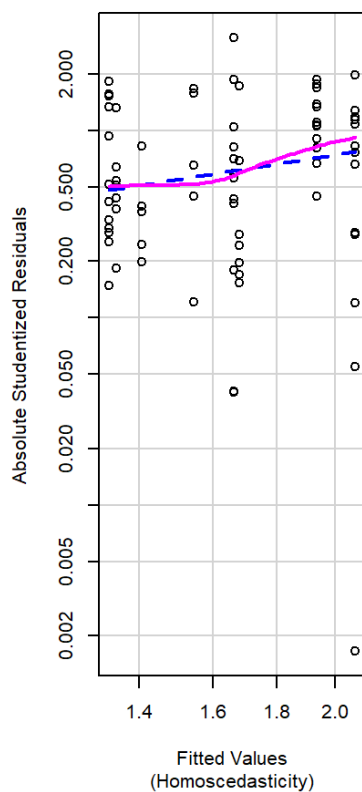


```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 1.103575, Df = 1, p = 0.29348
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
## Warning in e_plot_lm_diagnostics(finalfit): Note: Collinearity plot unreliable
## for predictors that also have interactions in the model.
```

Spread-Level Plot for fit



Collinearity

