Effect of metal stress on growth dynamics of yeast

Javad Meghrazi

3/11/2022

Background

Evolution can provide more than one way to deal with an environmental stress. It has always been a matter of debate that how diverse these evolutionary strategies could possibly be. Recently in Dr. Otto's lab at UBC, growth dynamics of multiple strains of *S.cerevisiae* was investigated in semi-lethal concentration of multiple metals. The tested strains had been originally evolved to grow under copper stress (Gerstein et al. 2015). Genome of all the strains have been fully sequenced and their important mutations were found.

I'm trying to make mathematical models to provide a framework to explain the change in growth dynamics of yeasts. Based on the models, there are two main reasons that cells struggle in high concentration of metals: 1- they should spend energy to pump out the metal out of cytoplasm and 2- they have to deal with damages caused by Reactive Oxygen species produced under metal stress. I believe the relative importance of these two pathways are reflected in the relative change of two main parameters that could be extracted from the curves; growth rate (r) and carrying capacity (k). As the growth of the yeasts are mainly restricted by lack of sugar when they reach the stationary phase of growth curve, k is a measure of how efficient cells are converting the fuels into biomass. The relative importance of the first pathway is expected to be inversely proportional to k. On the other hand, both pathways could potentially affect r. Overall, the relative change in the k and r could tell us about the relative importance of each pathway. Also by looking at how different mutations are changing r and k we can have a rough idea of how they are helping the cells to deal with metal stress.

One limitation is that there are couple of assumptions that I don't if they are well supported or not: 1- growth is mainly limited by lack of nutrients when yeasts enter stationary phase. 2- These two pathways are the main pathways and the second pathway don't significantly change k

Analysis

In this report, I'm trying to find that how stress from different metals alters the k and what is the effect of different mutation of evolved strains on k. I have previously extracted r and k from the data of growth curves and here I'm going to make linear model with k as response variable and mutations and metals as explanatory variables.

```
# importing the data of coefficients extracted from the curves
curves_analysis <- read.csv("D:/YeastMetal/YeastMetal/analyzed_data/curves_analysis.csv") %>%
  filter (strain != "blank")

# importing the data of mutations that each strain has + making the name of
# strains consistent across two datasets
CBMmut <- read.csv ("D:/YeastMetal/YeastMetal/raw_data/Cupper_lines/CupLineMut.csv") %>%
  mutate (strain = case_when(Line == "wt" ~ "OLY008",
                             Line == "cbm44" ~ "CBM44h",
```

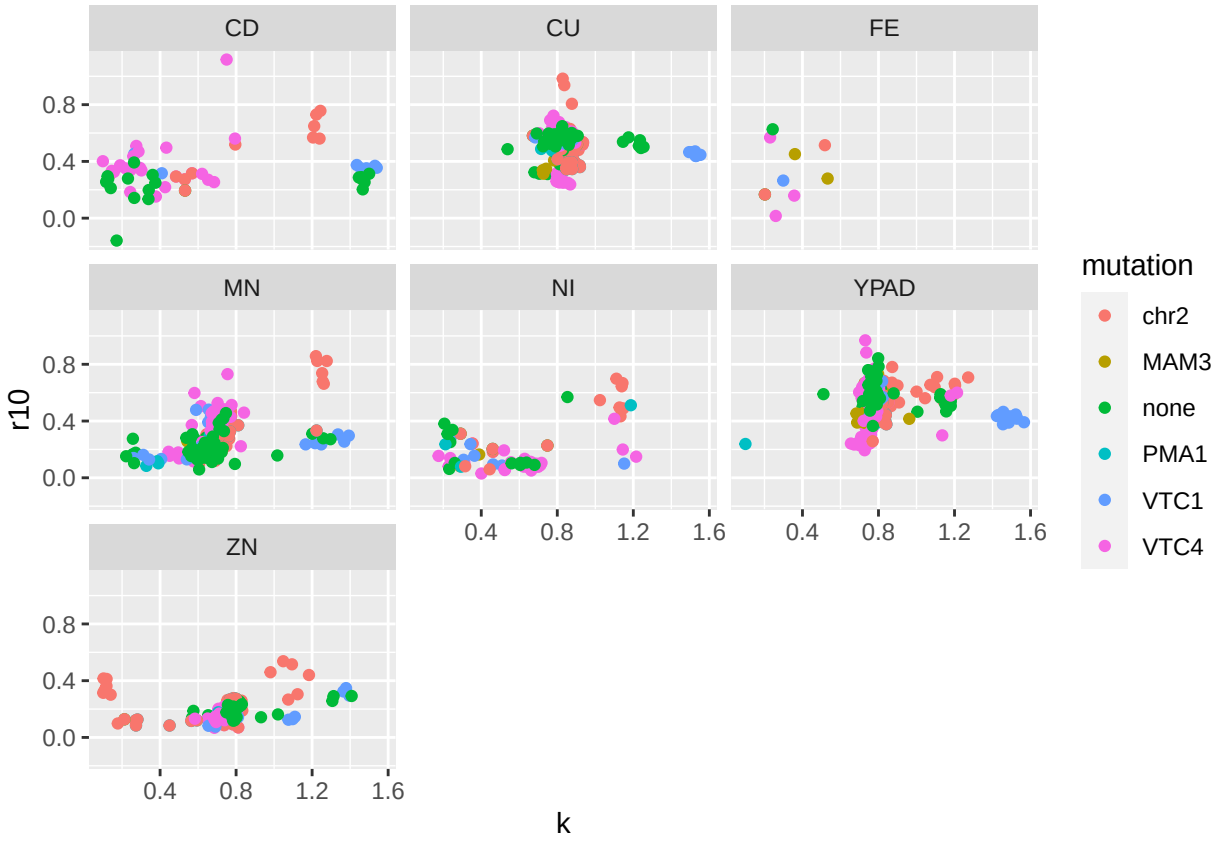
```
                            !(Line %in% c("wt", "cbm44")) ~ casefold(Line, upper = TRUE) )) %>%
  select (strain, name) %>%
  rename (mutation = name)

# some lines have more than one mutation and I'm turning those into multiple
# rows of data that each have 1 mutation. The analysis becomes simpler but
# the it artificially increases the sample size (but there are very few lines
# with multiple mutations so it shouldn't be a big problem)
# joining the two datasets
curve_mut <- CBMmut %>% right_join(curves_analysis)
```

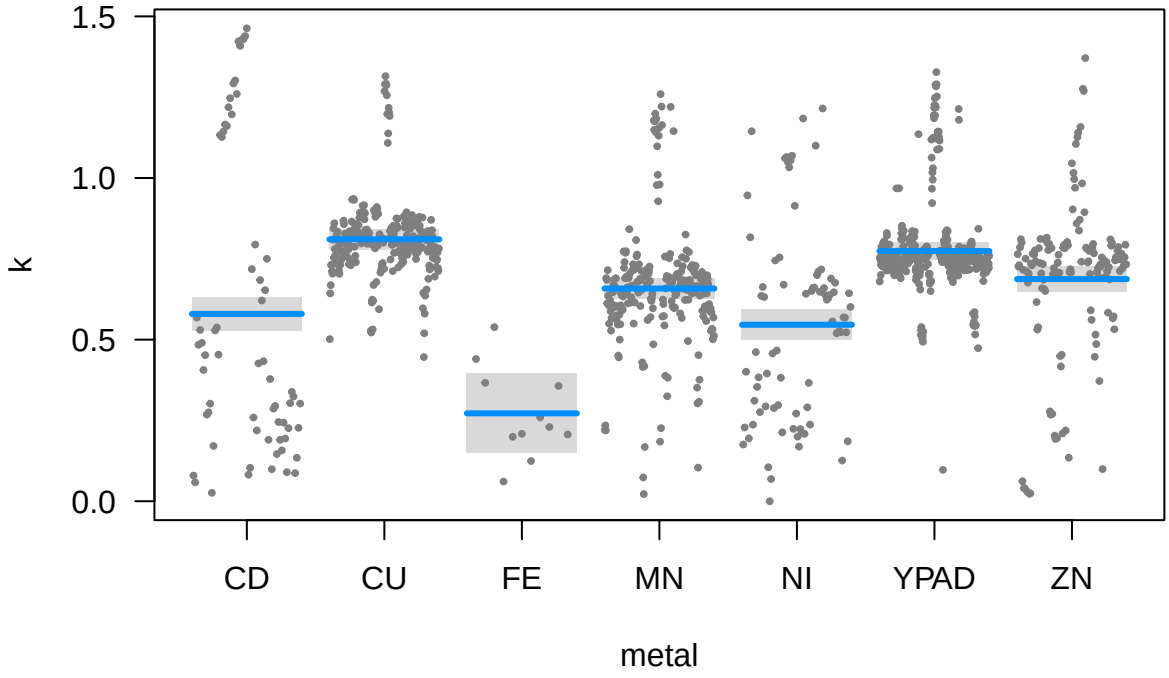
```
## Joining, by = "strain"
```

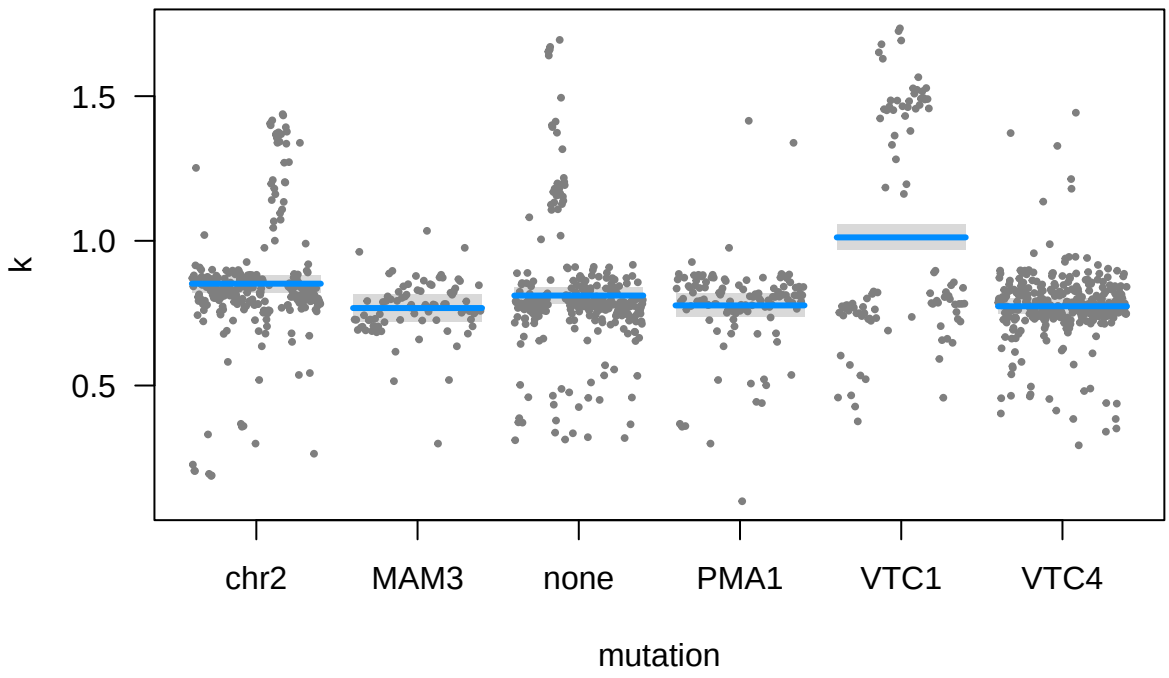
```
# some rows have very low k (k < 0.08) or don't have r and should be
# filtered because the parameters extracted from those rows are not reliable.
# These are usually the experiments where there was no significant population
# growth (OD_change < 0.08). This makes the result of analysis biased because
# for example in cadmium there are many data points where k is small because
# there is almost no growth, but those data points are eliminated. There is
# no clear way solvethis problem because if we want to keep those data, the
# estimates of k that we have from them is not reliable.
curve_mut <- curve_mut %>%
  filter (OD_change > 0.08, k > 0.08 ) %>%
# measures of r are not always reliable in this dataset. there are two
# measures of r (r15 and r10) and if their difference is high, our data is
# not reliable. here we are changing infinite values of r to NA and
# calculating rdif as a measure of robustness of r. and only keeping those
# whose rdiff is less than 0.2
  mutate (r10 = case_when(is.infinite(r10) == FALSE ~ r10 )) %>%
  mutate (r15 = case_when(is.infinite(r15) == FALSE ~ r15 )) %>%
  mutate (rdif = r10 - r15) %>%
  filter (rdif < 0.3) %>%
# creating a new column for concentration and removing "cont1" and "little1"
# values from conc column. Because we are not interested in that data
  filter (! conc %in% c("cont1", "little1", "cont2")) %>%
# the data for concentration of metals and the type of metal is in the same
# column so we have to separate them
  mutate (concentration = case_when( conc != "YPAD" ~ substr (conc, nchar (conc),nchar(conc)),
                                     conc == "YPAD" ~ "1")) %>%
  mutate (metal = case_when(conc != "YPAD" ~ substr (conc, 1 ,nchar(conc) -1),
                                     conc == "YPAD" ~ conc))
```

```
# this plot gives a general idea of variation of r (r10) and k in differnt
# metals and how different mutations have changed these values.
ggplot (data = curve_mut, aes (x = k, y = r10, color = mutation))+
  geom_point()+
  facet_wrap(~ metal)
```




```
# a linear model of k based with two explanatory variables: 1- metal
# 2- mutation. Because each variable has multiple categories and the data is
# limited, the interaction term is not added
k_lm <- lm (data = curve_mut, k ~ metal + mutation)
visreg(k_lm)
```


1.5
1.0
0.5
0.0
k
CD
CU
FE
MN
NI
YPAD
ZN
metal




```
summary (k_lm)
```

```
##
## Call:
## lm(formula = k ~ metal + mutation, data = curve_mut)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.67685 -0.05888 -0.01238  0.04692  0.88349
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.65754    0.02826  23.269  < 2e-16 ***
## metalCU       0.23082    0.02871   8.041 2.18e-15 ***
## metalFE      -0.30764    0.06699  -4.592 4.86e-06 ***
## metalMN       0.07867    0.02886   2.726 0.006514 **
## metalNI      -0.03346    0.03444  -0.972 0.331459
## metalYPAD     0.19437    0.02738   7.099 2.18e-12 ***
## metalZN       0.10776    0.03095   3.482 0.000517 ***
## mutationMAM3 -0.08421    0.02631  -3.200 0.001411 **
## mutationnone -0.04104    0.01786  -2.297 0.021781 *
## mutationPMA1 -0.07506    0.02298  -3.266 0.001122 **
## mutationVTC1  0.16000    0.02455   6.518 1.06e-10 ***
## mutationVTC4 -0.07807    0.01707  -4.573 5.31e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2045 on 1167 degrees of freedom
## Multiple R-squared:  0.2337, Adjusted R-squared:  0.2265
## F-statistic: 32.36 on 11 and 1167 DF,  p-value: < 2.2e-16
```

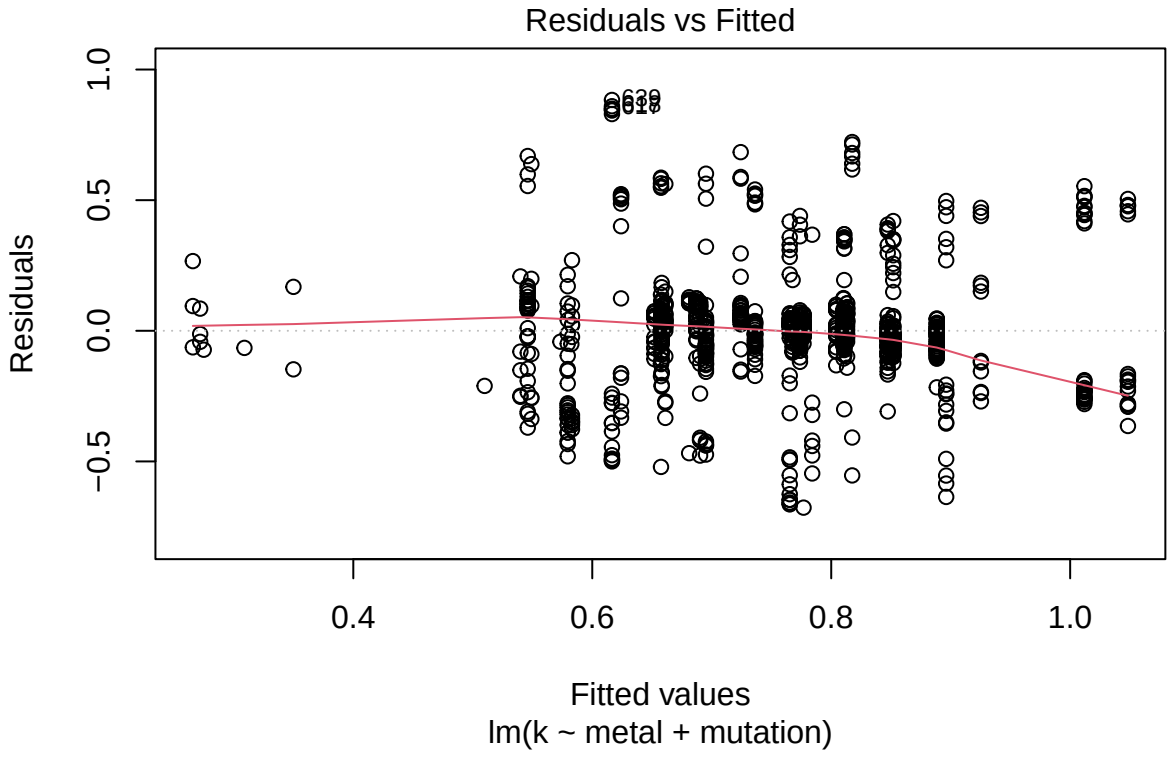
The diagrams show the mean and distribution of k for each category. Also summary table is showing the estimates of means of each category (in relation to intercept). In order to check the significance of the results, ANOVA should be done.

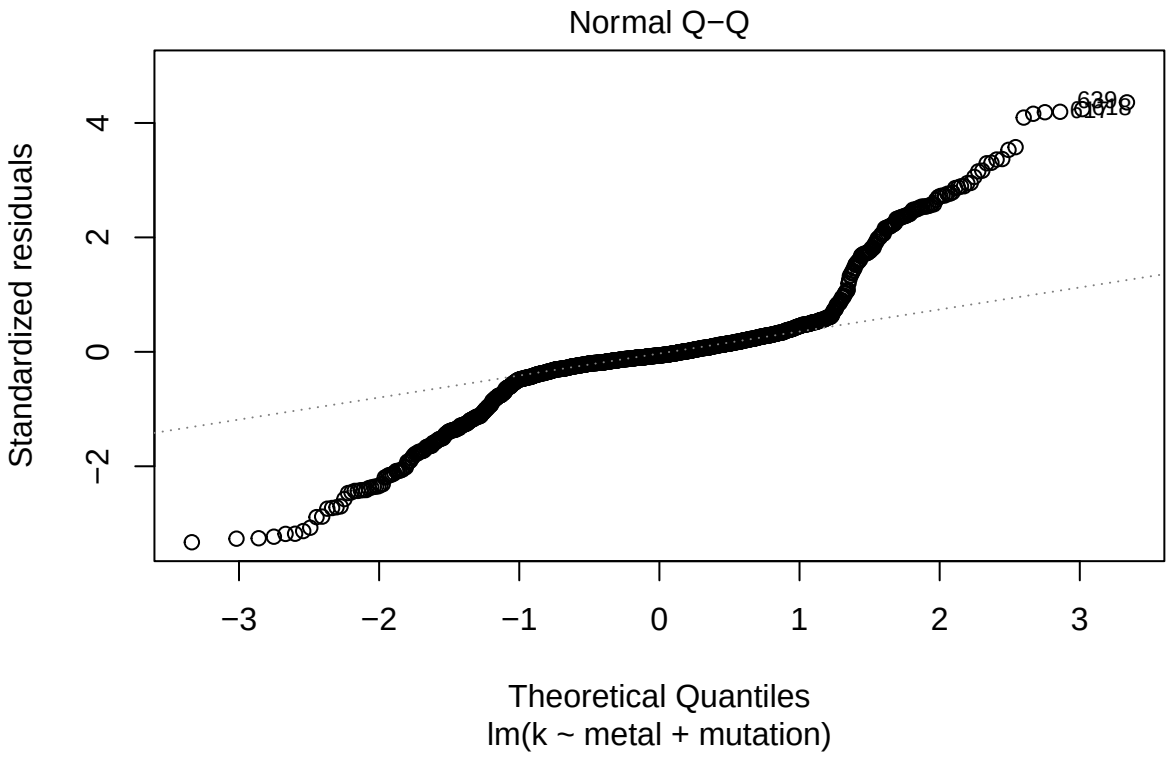
```
anova (k_lm)
```

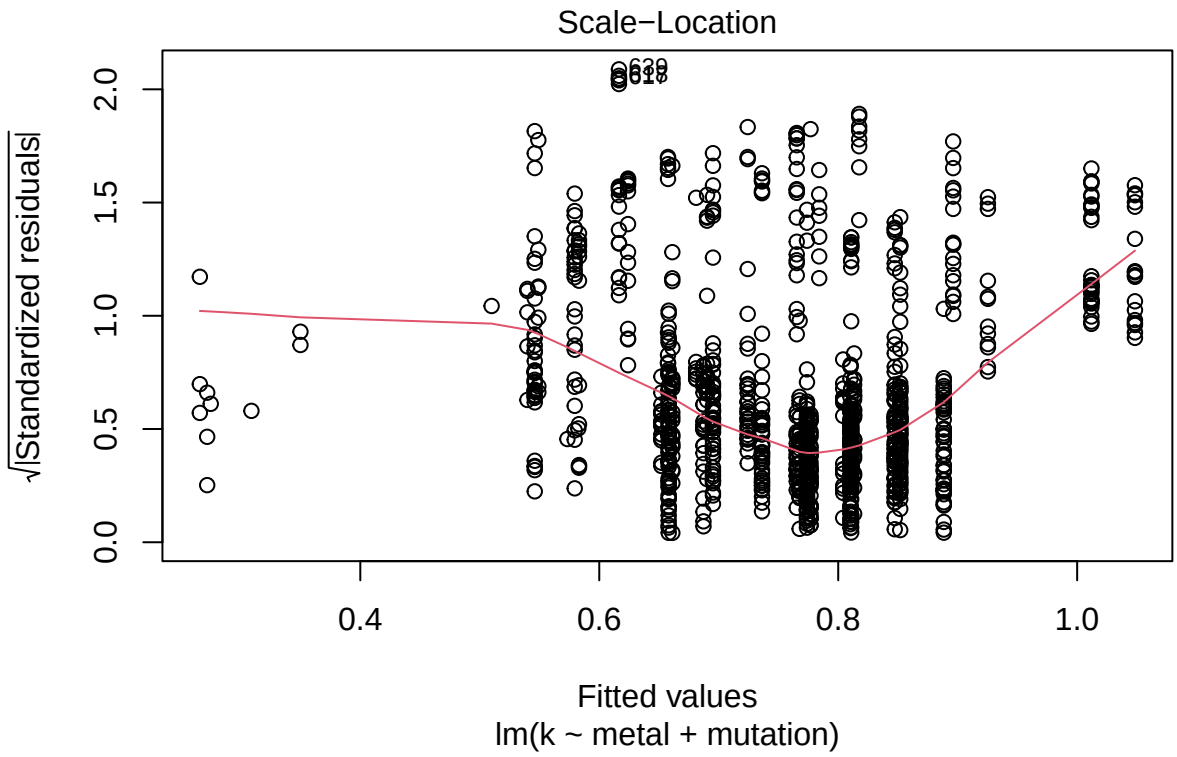
```
## Analysis of Variance Table
##
## Response: k
##             Df Sum Sq Mean Sq F value    Pr(>F)
## metal        6  9.859  1.6432  39.307 < 2.2e-16 ***
## mutation     5  5.023  1.0045  24.029 < 2.2e-16 ***
## Residuals 1167 48.785  0.0418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

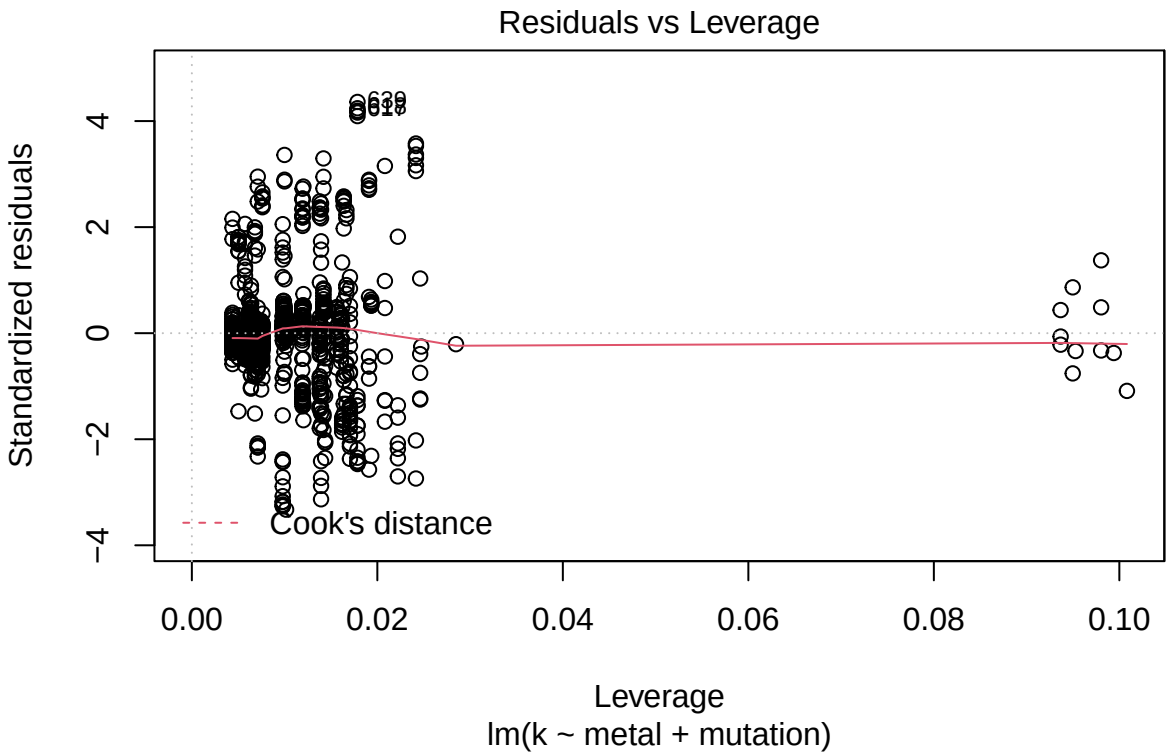
Results show that both variables have significant effects on k. Also we should check the assumptions:

```
plot (k_lm)
```



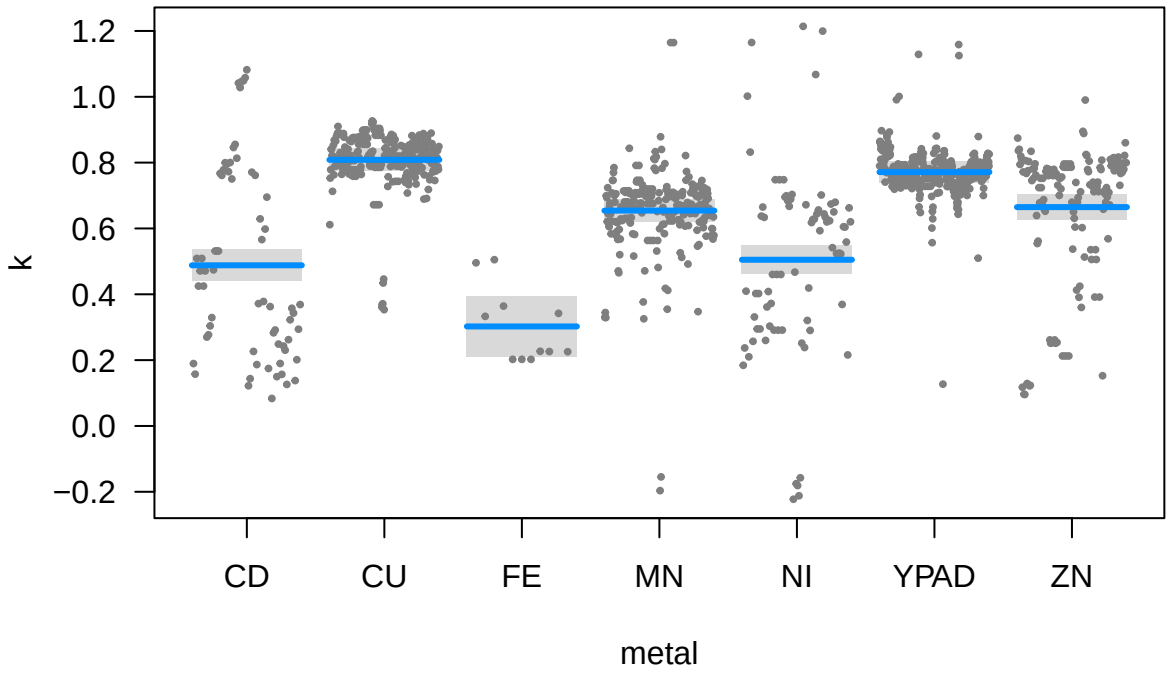

Normal Q−Q
Standardized residuals
−2
0
2
4
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(k ~ metal + mutation)

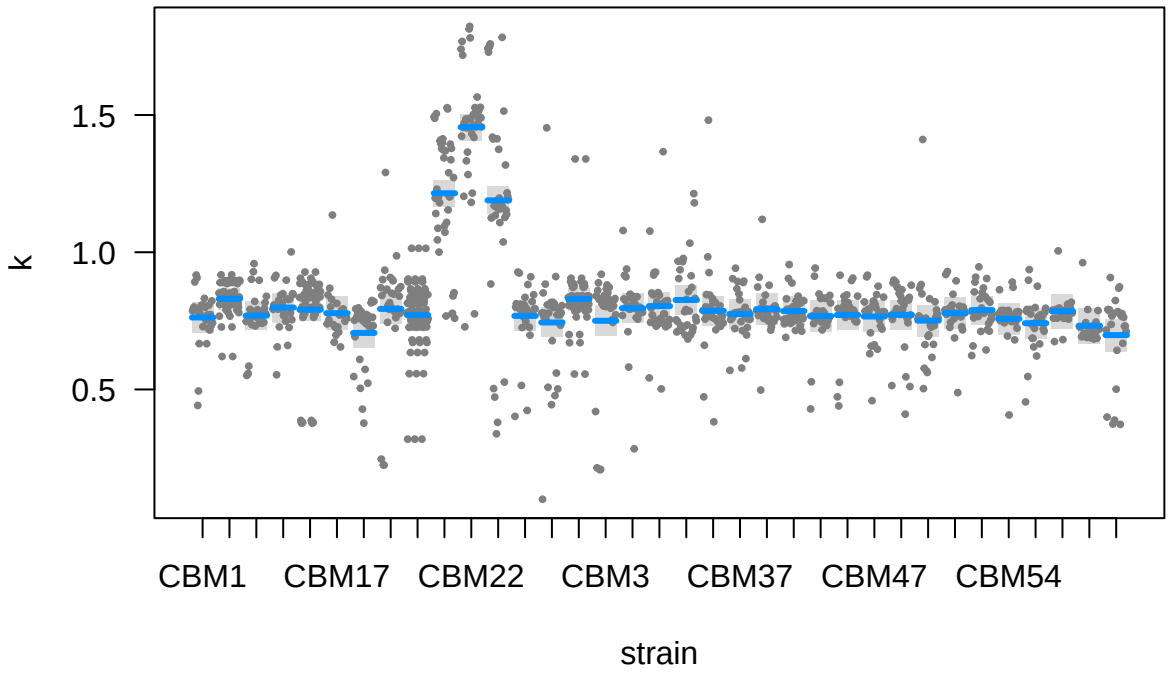

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0.4
0.6
0.8
1.0
Fitted values
lm(k ~ metal + mutation)



The QQ plot shows a significant deviation of residuals from normal distribution. The reason in this case is some strains have significantly higher K than others and this difference is not necessarily captured in the mutations that we monitor. So maybe if we chose strains instead of mutation as out explanatory variable, we get better results.

```
k_line_lm <- lm (data = curve_mut, k ~ metal + strain)
visreg(k_line_lm)
```


1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
k
CD
CU
FE
MN
NI
YPAD
ZN
metal




```
summary (k_line_lm)
```

```
##
## Call:
## lm(formula = k ~ metal + strain, data = curve_mut)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.85071 -0.03006  0.00849  0.05447  0.70907
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.479275   0.034039  14.080  < 2e-16 ***
## metalCU       0.320476   0.021226  15.098  < 2e-16 ***
## metalFE      -0.185780   0.049157  -3.779 0.000165 ***
## metalMN       0.166142   0.021302   7.799 1.40e-14 ***
## metalNI       0.016801   0.025188   0.667 0.504902
## metalYPAD     0.283071   0.020281  13.958  < 2e-16 ***
## metalZN       0.176769   0.022857   7.734 2.29e-14 ***
## strainCBM11   0.067883   0.033729   2.013 0.044396 *
## strainCBM13   0.006783   0.038312   0.177 0.859502
## strainCBM14   0.036141   0.038982   0.927 0.354056
## strainCBM16   0.028365   0.034304   0.827 0.408479
## strainCBM17   0.015448   0.041677   0.371 0.710954
## strainCBM18  -0.056375   0.038604  -1.460 0.144469
```

```
## strainCBM2    0.031149   0.038927   0.800 0.423761
## strainCBM20   0.008814   0.031920   0.276 0.782504
## strainCBM21   0.452589   0.036742  12.318  < 2e-16 ***
## strainCBM22   0.693445   0.036764  18.862  < 2e-16 ***
## strainCBM24   0.426749   0.037189  11.475  < 2e-16 ***
## strainCBM25   0.005662   0.039222   0.144 0.885234
## strainCBM26  -0.018150   0.038862  -0.467 0.640554
## strainCBM29   0.067125   0.034304   1.957 0.050620 .
## strainCBM3   -0.012086   0.039249  -0.308 0.758190
## strainCBM30   0.033852   0.038872   0.871 0.384016
## strainCBM33   0.041393   0.038066   1.087 0.277091
## strainCBM34   0.063682   0.038459   1.656 0.098030 .
## strainCBM36   0.024217   0.038601   0.627 0.530544
## strainCBM37   0.012380   0.038934   0.318 0.750567
## strainCBM4    0.030371   0.039891   0.761 0.446604
## strainCBM44h  0.023496   0.038889   0.604 0.545837
## strainCBM45   0.005092   0.039936   0.128 0.898559
## strainCBM46   0.009078   0.038601   0.235 0.814123
## strainCBM47   0.003643   0.039907   0.091 0.927288
## strainCBM49   0.009639   0.038601   0.250 0.802854
## strainCBM5   -0.011482   0.039888  -0.288 0.773506
## strainCBM51   0.015343   0.040294   0.381 0.703437
## strainCBM53   0.026787   0.038344   0.699 0.484943
## strainCBM54  -0.005037   0.040324  -0.125 0.900619
## strainCBM55  -0.020979   0.039615  -0.530 0.596509
## strainCBM6    0.022770   0.042190   0.540 0.589515
## strainCBM7   -0.030128   0.043390  -0.694 0.487605
## strainOLY008 -0.064192   0.040760  -1.575 0.115563
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1479 on 1138 degrees of freedom
## Multiple R-squared:  0.6091, Adjusted R-squared:  0.5954
## F-statistic: 44.33 on 40 and 1138 DF,  p-value: < 2.2e-16
```

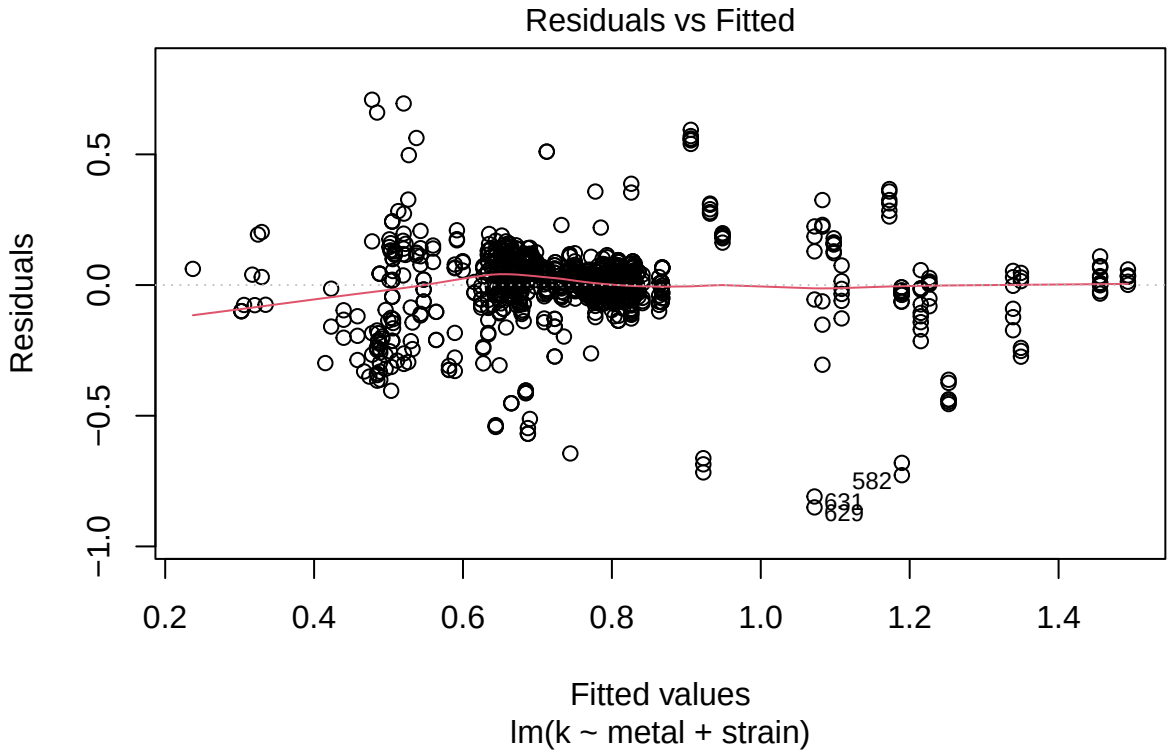
The diagrams show the mean and distribution of k for each category. Also summary table is showing the estimates of means of each category (in relation to intercept). In order to check the significance of the results, ANOVA should be done.

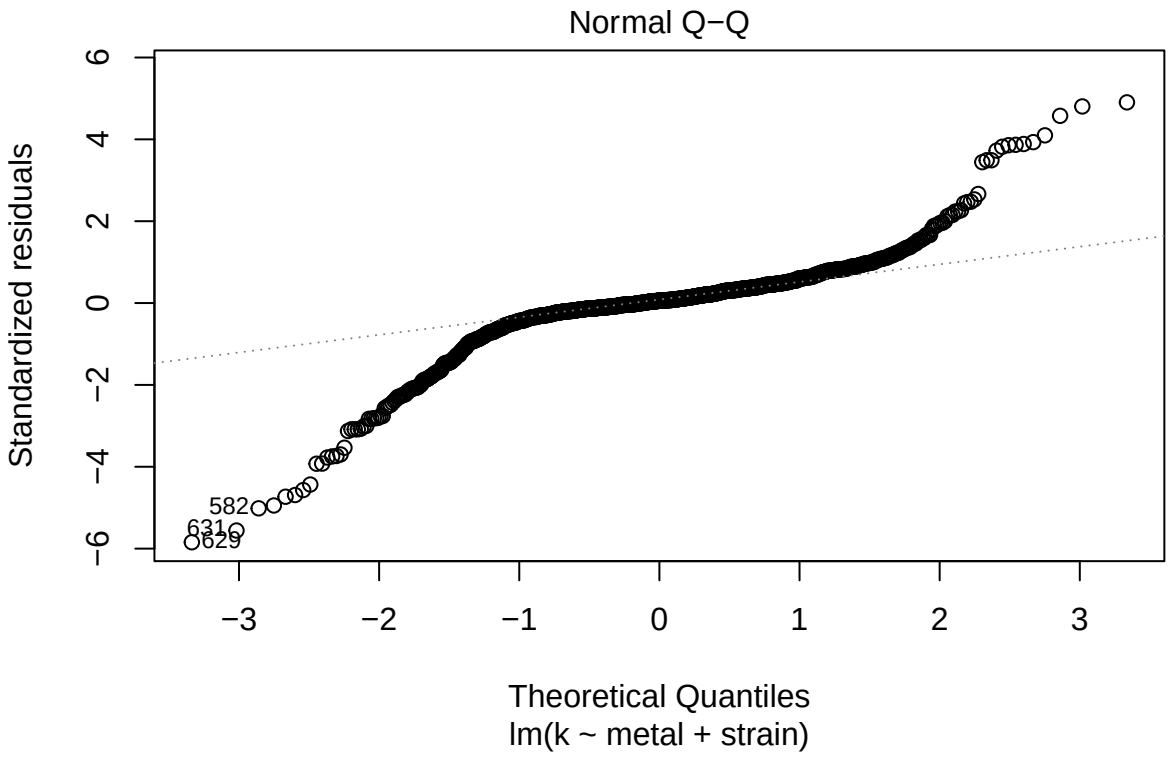
```
anova (k_line_lm)
```

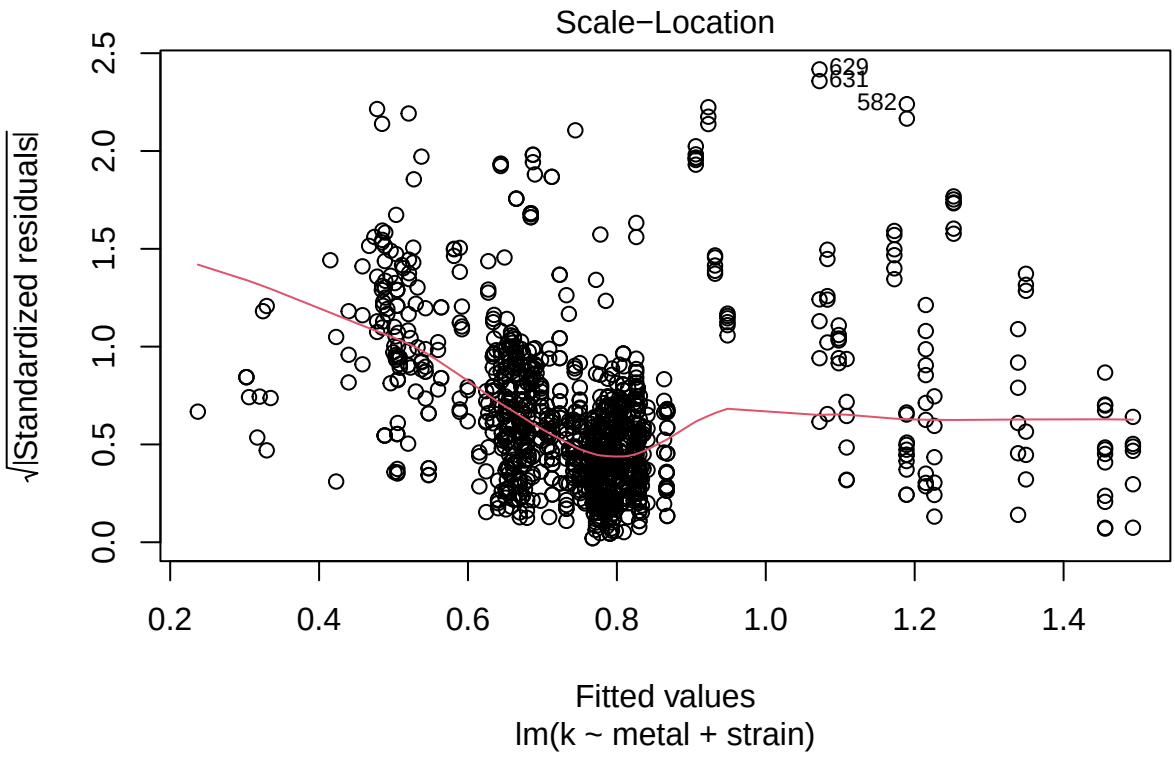
```
## Analysis of Variance Table
##
## Response: k
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## metal        6  9.8592 1.64319  75.138 < 2.2e-16 ***
## strain      34 28.9211 0.85062  38.896 < 2.2e-16 ***
## Residuals 1138 24.8868 0.02187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

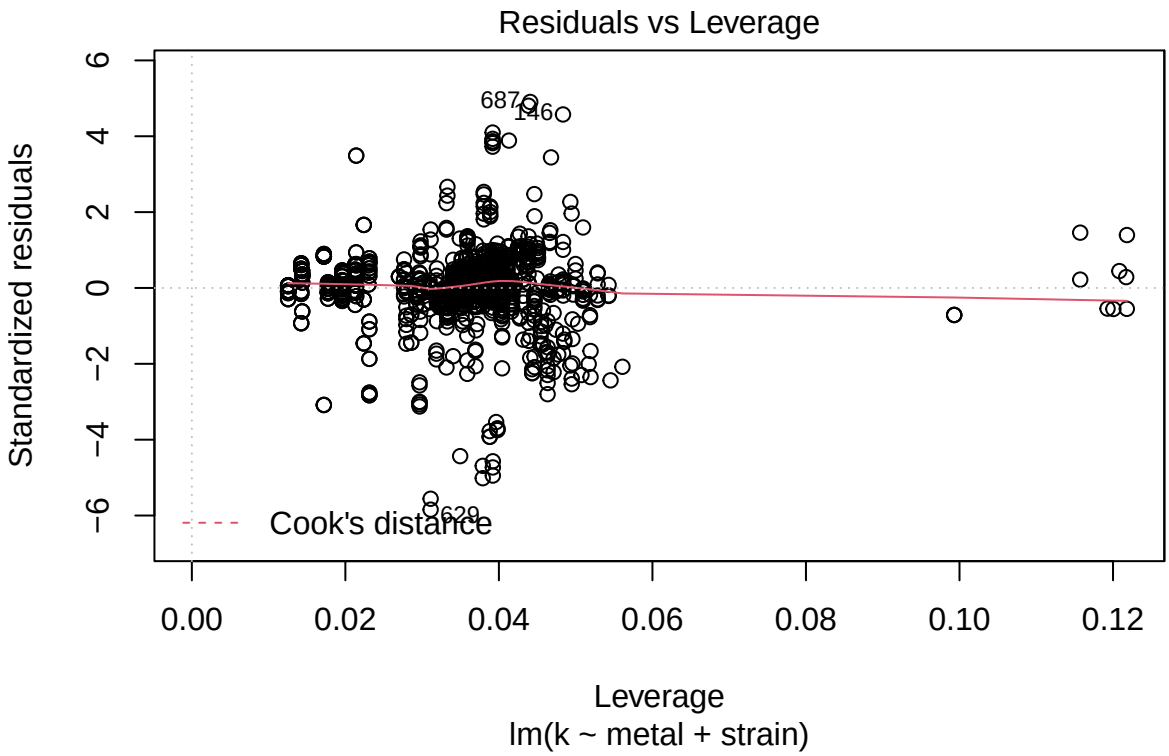
Results show that both variables have significant effects on k. Also we should check the assumptions:

```
plot (k_line_lm)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(k ~ metal + strain)
582
631
629
−6
−4
−2
0
2
4
6
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
629
631
582
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(k ~ metal + strain)



The QQ plot is better now. Even though the distribution of residuals is not completely normal now but we can't do anything about it because the transformation of k is not very meaningful. The underlying reason of violation is that there are still other fixed effect explanatory variables (like interaction of mutation and metals) and random effects (like replicates of the experiment) that should be accounted for later.

Conclusion

In this report we are not testing an specific hypothesis because our study is exploratory. Results suggest that that both the metal and the mutation are affecting the carrying capacity reached. When we run this analysis by mutation as one of the categorical variables the distribution of residuals is very different from normal but if we replace mutation with strain, this distribution will be more like normal. It suggests that there is some variation in the genetic background of different strains that is not captured in the mutations that we monitored. Looking at the relation ship of k with different strains suggests that for most of the strains k is very similar, except 3 of them which have significantly high k. This suggests that patterns that we observed in the first analysis (where we had mutation as one of the explanatory variables) are not reliable because maybe these three strain accidentally have some mutations and create a false pattern in the results of the first analysis which is independent of the true effect of those mutations. The individual effect of each mutation on k is not easily inferable from this data.