**3. Exploration**  
  
For investigating the effect of a factor on a variable we can use Analysis of variance (ANOVA) which is a collection of statistical models and their associated estimation procedures used to analyze the differences among means. For following this procedure, we started exploration by making boxplots for the levels of each factor (Fig.1) It is obvious the mean of leak for species 1 (s1) and 2 (s2) are different, showing difference in mean leak by factor variety. This seems also hold true for factors regime and temperature. In the next stage we would like to investigate if the effect of each factor depends on the other factors; The profile plot of sample means (i.e., the interaction plot) (Fig.2) shows a clear interaction between variety and regime; the difference in the mean lengths for regime R and regime C depends on the varieties, with larger variety differences for regime R than for the regime C. This indicates in addition to main effects it is better include interaction effects in the model to analyze the data and inference if the main effects and interaction effects are statistically significant.

**4. Full model**

The full model that we proceeded to fit is:

Leak\_ijkr = mu + alpha\_i + beta\_ij+ gama\_k + ……+ epsilon\_ijkr

i=1,2  
j=1,2  
k=1,2  
r=1,2,…,n\_ijk

Where leak is the response variable

alpha\_i, beta\_j, gamma \_k are the effects of level I of variety, level j of regime and level k of temperature respectively.

(Alpha\*beta)\_ij , (alpha\*gama)\_ik, (beta\*gama)\_jk are two way interaction effects of factors

And (alpha\*beta\*gama)ijk is the only possible tree way interaction

Epsilon\_ijkr is the error that cannot be controlled.

**4.1** **Model Assumptions**

We checked the assumptions for our initial model. the normality assumption was not met as seen from the normal QQ plot (Fig.3), and the Shapiro-Wilks Test ($𝑝 = 0.03256$). Also, from the plot of the residuals and the fitted values (Fig.4), and the Breusch Pagan Test ($𝑝 = 2.307\times10^{-6}$), there is evidence to conclude that the constant variance assumption was not met. We do not know enough about how the data was collected to make a definitive statement about independence, and do not see any reason to question the assumption as it pertains to this data.

**5 Remedial Measures**

To address the violations of the model assumptions, a transformation on the response variable was attempted. The Box-Cox transformation suggested log10 transformation (**Fig.5)**. We checked the assumptions for our log 10 transformed model. By fitting the $log\_10$ transformation, no violation of the equal variance assumption occurred (Fig.7). Normality still appeared to be violated based on the normal QQ plot. Also, according to the Bonferoni Outlier Test of residuals ($𝑝 = 0.036851$) the observation number 25 is an outlier. To address the normality assumption violations of the $log\_10$ transformed model, again a transformation on the $log\_10$ response variable was attempted. The Box-Cox transformation applied on $log\_10$ transformed model suggested squared transformation (**Fig.8)**. The two log transformations indicate ${log\_{10}(leak)}^2$. By fitting the ${log\_{10}(leak)}^2$ transformation, no violation of normality (Fig.9) and the equal variance assumption (Fig.10) occurred, there is also no outliers (by this transformation we keep all the observations) and this transformation was applied for the next stage of model selection.

T

**6 Reduced Model and Selection**