**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\_i+ 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 $log\_10$ 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 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.

The ANOVA table of the following model

|  |  |
| --- | --- |
| ${log\_{10}(leak)}^2 = variety+regime+temp+variety:regime+variety:temp+temp:regime+ variety:regime:temp $ | (1) |

is:

**Table 1**

**6 Reduced Model and Selection**

The ANOVA table of the adequate transformed full model (Table 1) indicates the three-way interaction `variety:regime:temp` and the two way interaction `variety:temp`are not significant.

At the beginning we removed the three-way interaction effect and consider the first reduced model:

Reduced model (R1):

${log\_{10}(leak)}^2 = variety+regime+temp+variety:regime+variety:temp+temp:regime$

Reduced model (R1) meets all the model assumptions. We compared model (1) with reduced model (R1). The p-value $p=0.5585$ indicates we do not reject the reduced model (R1).

In the **second** step, the ANOVA table of reduced model (R1) indicates the two way interaction `variety:temp` is not significant so we removed it and considered the second reduced model:

Reduced model (R2):

${log\_{10}(leak)}^2 = variety+regime+temp+variety:regime $

Reduced model (R2) meets all the model assumptions. We compared model (R1) with model (R2). The p-value $p=0.5871$ indicates we do not reject the reduced model (R2).

In the **third** step, the ANOVA table of reduced model (R2) indicates the two way interaction `temp:regime` is not significant so we removed it and considered the third reduced model:

Reduced model (R3):

${log\_{10}(leak)}^2 = variety+regime+temp+variety:regime$

Reduced model (R3) meets all the model assumptions. We compared model (R2) with model (R3). The p-value $p=0.05471$ indicates we do not reject the reduced model (R3).

We decided to select the third reduced model (R3) as our final model since all its effects are significant and congruent with the assumptions very well.

**7 Final Model**

The final model we selected for this data is:

$${log\_{10}(leak)}^2 = \mu + \alpha\_i + \beta\_j+ \gama\_k +(\alpha\beta)\_{ij} + \epsilon\_{ijkr}$$

$i=1,2$  
$j=1,2$  
$k=1,2$  
$r=1,2,…,n\_{ijk}$

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

$(\alpha\beta)\_{ij}$ is the two way interaction effects of $\alpha\_i$ and $\beta\_j$

$\epsilon\_{ijkr}$ is the error that cannot be controlled.

The residuals of the model were plotted against fitted value (**Fig. 12**) and we took note of there is no outliers. We see the scatter is even overall. The Bonferoni Outlier Test of residuals ($𝑝 = 0.56156 $) also confirm that. The BP test for equal variance results in a p-value of 0.4549, indicating that the residuals from this model do not show evidence of unequal variance, and confirmed our visual analysis of the plot.

The diagnostics on the normality of the residuals points to **no** violation of this assumption. The  
normal QQ plot of the residuals (**Fig. 11**) shows that all values are following the straight line without any significant departure. At the left tail, there are two values that fall outside the simulation. The Shapiro-Wilks test for normality on these residuals results in a p-value of 0.6591, providing strong evidence that the residuals do follow the normal distribution.

**8 Model Interpretation and Conclusions**