

Proposal: Regression Analysis of Low-Pressure UV Irradiation for disinfecting viruses Drinking Water from Wastewater

David Aguilera, Samantha Juedemann, Tanjina Zaman

Introduction and background information:

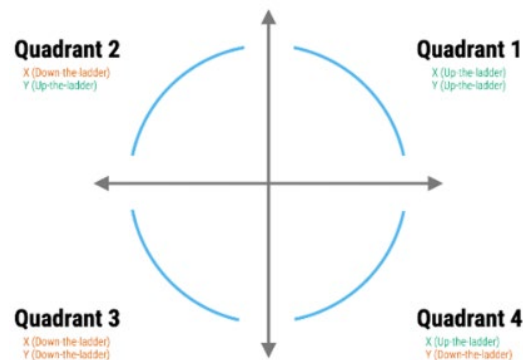
Viruses are one of the smallest and most difficult pathogens to mitigate in drinking water treatment. The challenges associated with mitigating these viruses will worsen with methods of water retrieval to combat water scarcity. To ensure sufficient virus mitigation, surrogate viruses need to be selected to be used in laboratory disinfection experiments to replace the dangerous and challenging pathogenic virus culturing. It is necessary to demonstrate inactivation relationships for many viruses with each type of disinfection system to determine which specific viruses and viral characteristics are more resistant to disinfection. Then, regulations can be set in place based on these experiments with the surrogates.

This data represents disinfection experiments using Low-Pressure UV (LP-UV) irradiation with five different viruses. The viruses are exposed to multiple UV doses, which is the predictor variable, and the surviving concentration is divided by the initial concentration to be the response variable. This analysis should determine which virus is the most resistant to LP-UV, and possibly which viral property causes this resistance. The model can then be used to determine the inactivation of viruses based on the UV-dose. This will help determine the UV dose to attain a required level of mitigation for viruses, even those which are resistant.

Planned analysis based on initial data review:

Normality analysis: The data should be normal to perform the categorical variable analysis. Additionally, non-normality may indicate the need for a variable transformation (in which case, a Box-Tidwell will be performed). The sample data will be checked for non-normality by performing partial residual plots. If necessary, a transformation will be conducted to address linearity using the bulging rule. Transformation techniques are possible by the bulging rule, as shown in the following illustration.

Rule of the Bulge: Visual Mnemonic for choosing power transformations to 'straighten' each of the four monotonic curves.



Regression diagnostics: The analysis will include confirming model adequacy and residual diagnostics, cleaning the outliers and leverage points, finding the influence point, and checking the collinearity among the variables.

Non-Parametric Analysis: One of the group members, Samantha Juedemann, obtained the raw data in a rigorous lab setting on our lovely Marquette campus premises. Due to experimental constraints, the sample size of the raw data is small, preventing us from approximating the true population. Therefore, the analysis will overcome this challenge by either employing bootstrapping techniques or conducting non-parametric tests.

Categorical Variable Analysis: The type of virus should be separated into groups. The virus MS2 will be used as the baseline since it is one of the most common viral surrogates. The remaining viruses will be assigned a dummy variable to distinguish them. The following linear regression and table represent the initial categorical model with corresponding dummy variables.

The regression Categorical Model:

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e$$

Type Viruses	x2	x3	x4	x5
MS2	0	0	0	0
PhiX174	1	0	0	0
P22	0	1	0	0
Phi6	0	0	1	0
FCV	0	0	0	1