# Volts Analysis Report

Francisco Solis

Objectives:

1. Examine the direction, strength and relationship between variables for the original data
2. Check linearity, constant variance and normality of data for the original scatterplot
3. Examine direction, strength and relationship between variables for the transformed data
4. Check linearity, constant variance, and normality of data for the transformed data
5. Determine if the transformed data is more suitable for a simple linear regression model
6. Identify any outliers and or influential points on the transformed data
7. Model limitations of the log transformed data

*All content is referring to data in “Volts_Statistical_Analysis.pdf”

**Discussion/Analysis:**

1. The relationship between Voltage and Time is negative, as time increases the voltage decreases. Points on the plot slope downward, showing that higher times correspond to lower voltages. The relationship is strong because the data follows a smooth, consistent, curved path with very little scatter around the general trend. Indicating that time explains most of the variation in voltage. The relationship is non-linear, following an exponential decay. The curve starts steep and then levels off. This is typical of capacitor discharge, where voltage decreases exponentially over time.

2. Based on the Residuals vs Fitted diagnostic plot which checks the linearity and constant variance assumptions, the scatter plot shows a nonlinear trend resembling a parabola and the spread of data is concentrated near fitted values of 4. Thus, the linearity and constant variance assumptions are violated. The QQ plot which checks the normality of the data shows a heavy left tail indicating a strong deviation from the normal distribution. Ultimately, the linearity, constant variance and normality conditions are violated so using a simple linear regression model would not be an effective method to represent the data. A transformation of the data is recommended to remedy the issues.

3. The relationship between the log(Voltage) and Time is still negative, as time increases the voltage decreases. After transforming the data, the plot shows a strong, straight path with no scatter around the general trend. A strong diagonal path signals that time explains most of the variation in voltage. By taking the log of Voltage, the exponential decay has been linearized, making it suitable for fitting a simple linear regression model.

4. Based on the Residuals vs Fitted diagnostic plots, the data still follows a nonlinear, parabolic shape and the spread of data is not even. Hence, the linearity and constant variance conditions are still not satisfied. On the QQ plot, there is still a heavy left skew meaning the data has more extreme low values than expected under a normal distribution. Although, the linearity, constant variance and normality conditions are still violated the log transformation of voltage showed improvement in these assumptions which makes the log transformation more suitable for a simple linear regression.

5. The log transformation of voltage is more suitable for a simple linear regression model because the data no longer has a nonlinear trend on the scatterplot. Despite the

main three assumptions for simple linear regression being violated, there was improvement in the diagnostic plots such as a more even spread of data in the Residuals vs Fitted plot, and marginally improved normality of data in the QQ plot.

6. The Cook's distance plot checks for influential points (if removed they heavily change the model), in this case none of the points exceed the +/- 1 threshold thus, there are no influential points. The Residuals vs Leverage plot checks for any outliers in the data and none cross the +/- 3 threshold so there are no outliers. Observation 49 and 50 are the points closest to the threshold so they have notable influence but require further analysis to be confirm if they are truly problematic or simply reflect a valid structure. The further analysis of observations 49 and 50 will not be covered in this report.

7. Using a log transformation can improve linearity and stabilize variance, but it restricts analysis to strictly positive data, changes the interpretation of coefficients to multiplicative effects, and assumes multiplicative error structure. It may also overemphasize very small values and does not guarantee a good fit if the underlying relationship is not exponential. Considering the model's limitations, a non linear regression model is recommended to account for the curvature and model the data more accurately.