Jared Malto
Exploratory Data Analysis

1. **Internal Resistance Measure 1 and Internal Resistance Measure 2 measure the resistance of the same cell. By analyzing the data, is the Internal Resistance Measure 1 significantly different from Internal Resistance Measure 2?**

Through a z-test for a difference in means, there is strong evidence which suggests that Internal Resistance Measure 1 is significantly different from Internal Resistance Measure 2. Using the hypotheses:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

we test whether the means are significantly different from each other. To do this, we group each of the batteries together and perform the test on each cell treating them as individual samples. Before we begin testing, we need to check conditions on if this is an appropriate test. The conditions for a z-test for a difference in means are:

1. *Both n are sufficiently large ($n_1, n_2 > 30$)*
2. *σ_1, σ_2 are known*

OR

1. *Both populations are normally distributed.*

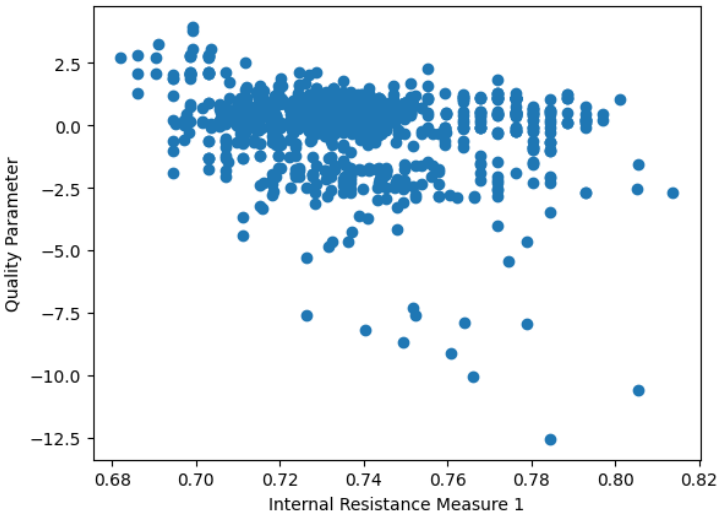
Given that the sample size is sufficiently large (n = 108 per battery), we can use Central Limit Theorem to infer the sample means are approximately normal. Using this assumption, we can use a standard z-distribution to calculate the probability of a significant difference in sample mean. Running the z-test for a difference in means for each cell in each battery, we get the results below with the corresponding p-values:

Battery ID	P-Value
1	1.1781572367060033e-112
2	7.755871927203019e-157
3	0.0*
4	3.314432075067432e-116
5	7.819149138490991e-269
6	4.507962695590614e-96

The returned values show the probability of there being no mean difference from each sample. From these p-values and using a significance level of 0.05, **we reject the null hypothesis for each of the batteries and conclude that there is evidence that shows that there is a significant difference between Internal Resistance Measure 1 and 2.**

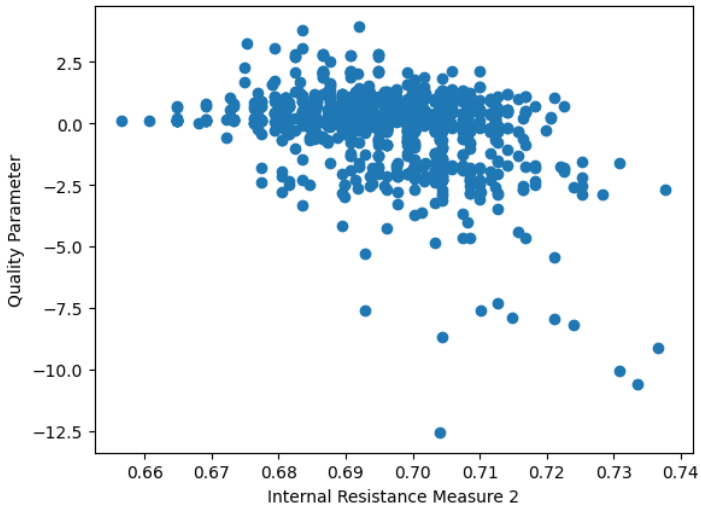
2. **Is there a correlation between Internal Resistance Measure and Quality Parameter?**

To check whether there is a correlation between Internal Resistance Measure and the Quality Parameter, we use a scatter plot to visually inspect get the correlation coefficient between the dependent variable (Internal Resistance) and explanatory variable (Quality Parameter). Plotting Quality Parameter against Internal Resistance Measure 1 results with the following plot:



By calling the data.corr() function on this data set returns a correlation coefficient of -0.22. This means that there is a **very weak negative linear correlation** between the two variables.

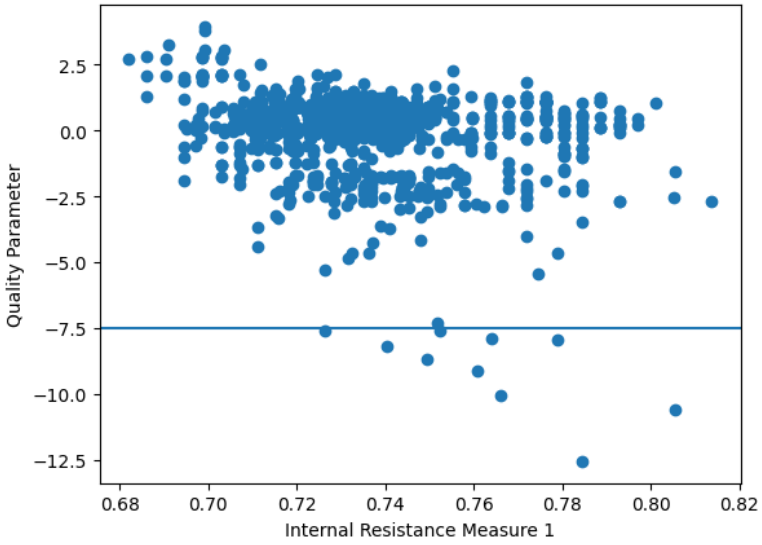
Similarly, plotting the Quality Parameter against Internal Resistance measure 2 results in:

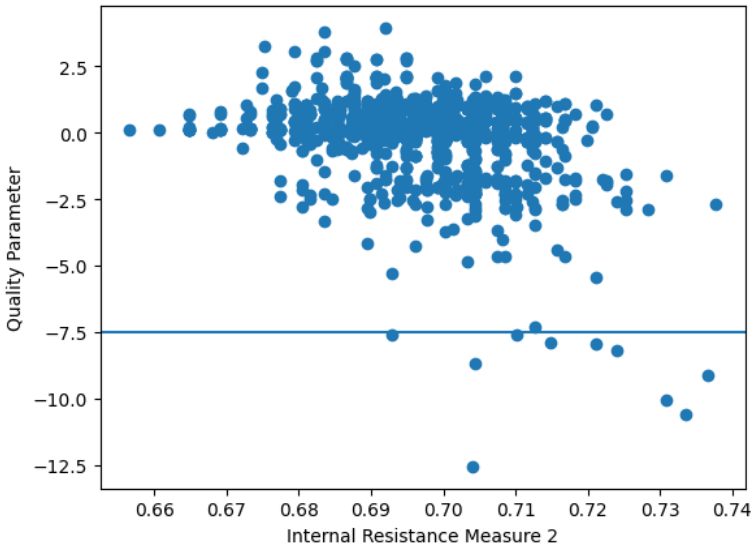


Calling data.corr() on this gives us a correlation coefficient of -0.30, giving this set a **very weak negative linear correlation.**

3. **Can we use Internal Resistance to determine whether the cell pass or fail the test?**

Using the plots from above, **<u>we generally cannot use Internal Resistance Measurements to determine whether the cell will pass or fail the test.</u>** Using the plots below:





we observe that there is no linear correlation between Internal Resistance and the Quality Parameter. In addition, both the passes and failures throughout both measurements are roughly distributed evenly, leading us to believe that a failure is possible from any Internal Resistance Measure and that one measurement is not more likely than another. Inversely, the distribution of the cells that pass the test are also evenly distributed. While most cells do pass the test, there is not a strong enough linear correlation to use the relationship of Internal Resistance Measure and Quality Parameter to predict the outcome of the test.

---

*Python would not print the answer in consistent exponential notation when performing the test. However, given how small the other comparing p-values are, it would be a safe assumption to reject the null hypothesis for this one as well