# Examining if Linear regression is a good model for predicting Received RF Signal Strength with frequency and bandwidth as predictors.

**Michael S. Odom***
CpE 590
University of Alabama in Huntsville
Huntsville, AL 35899
mso0007@uah.edu

## Abstract

This paper will explore my findings on using Linear Regression to predict received signal strength, based on frequency and bandwidth. Overall, based on the results, this was found to be a poor prediction method, but may have been due to what data was found in the dataset.

## 1 Background

When analyzing an Radio Frequency(RF) communication channel, Signal-to-Noise (SNR) ratio is the figure of merit. SNR is calculated with a version of the *Friis Transmission Equation* as follows:

$$SNR = \frac{ReceivedSignalStrength}{NoisePower}$$

Where:

$$ReceivedSignalstrength(Watts) = TransmitPower(Watts) * TransmitAntennaGain(Watts) * ReceiveAntennaGa$$

And:

$$NoisePower = Boltzmannconstant * temperature * Bandwidth$$

The dataset "RF Signal Data" by SURAJ on Kaggle was used to perform a linear regression regression.

### 1.1 RF Background

A few notes on the nature of the data used. Signal Strength is presented in dBm, so Frequency and Bandwidth are translated from Hertz into Log space using this relationship:

$$dBspace = 10 * Log10(LinearValue)$$

The reason these values were chosen as predictors is due to Frequency's proportional relationship to signal strength, as well as its relationship to wavelength:

$$wavelength = \frac{SpeedofLight}{Frequency}$$

which as a squared term could be "overpowering. Bandwidth was chosen as a predictor due to its quality of "Spreading the power."

*Graduate Student.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

### 1.2 Linear Regression

Linear Regression is a "kind of supervised machine learning algorithm where a linear discriminant model g(x) is used to fit on the Given DataSet."(Bhadani)

More Generally:

$$g(X) = w^T * x + w_0$$

Where: $w^T$ is the weights transposed and $w_0$ is the bias.

This is commonly seen as the "easiest" Machine learning algorithm and as such one that is commonly used.

### 1.3 Dataset

The dataset was downloaded from Kaggle as a CSV. Due to the massive size of the file, I used the first 1300 rows to perform my analysis.

To clean the data, I added two additional columns to translate Frequency and Bandwidth to "logspace."

The Dataset only contained one set of coordinates per entry sand the same location name, so an assumption was made that all the distances were equal.

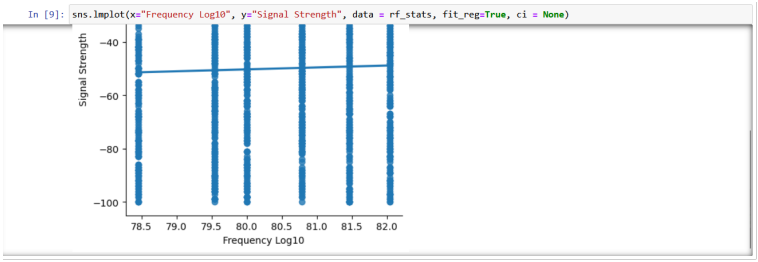
Also worth noting is that the dataset lacked Transmit power data, which is a key parameter in link analysis.

## 2 Results

First I looked at the correlation between various features.

```
Frequency           0.028303
Signal Strength     1.000000
Bandwidth          -0.052850
Temperature         0.041512
Bandwidth Log10    -0.018863
Frequency Log10     0.029379
Name: Signal Strength, dtype: float64
```

And looked at graphically:




It seems there is an incredibly weak correlation, in fact here is the mean absolute error and statistics of the dataset:

```
In [19]: mae
Out[19]: 25.555800692462977

         The mean absolute error is 24 dBm which is really large(remember this is in log scale)

In [20]: rf_stats.describe()["Signal Strength"]
Out[20]: count    1367.000000
         mean      -50.092904
         std        29.154674
         min      -100.000000
         25%       -75.000000
         50%       -50.000000
         75%       -25.500000
         max         0.000000
         Name: Signal Strength, dtype: float64
```

The Mean Absolute error is lower than the standard deviation, despite the correlation being relatively low when calculated.

## 3 Conclusion

It seems that linear regression is a poor fit when using frequency and bandwidth as predictors. For further investigation into this topic, I plan to look for a data set containing transmit power and distance as features. Since distance is included as a term in receive signal strength that is squared, it is most like more overpowering as it grows larger. Likewise, transmit power and gain provide an upper limit on what the receive strength should be. This project has really exposed to me the importance (and difficulty!) of finding a quality dataset.

## References

[1] Rahul Bhadani Assistant Professor Electrical and Computer Engineering The University of Alabama in Huntsville's notes

[2] The RF Signal Dataset found at: https://www.kaggle.com/datasets/suraj520/rf-signal-data?resource=download