Assignment #2

Our calculations were made in Python, the code is attached in a notebook file.

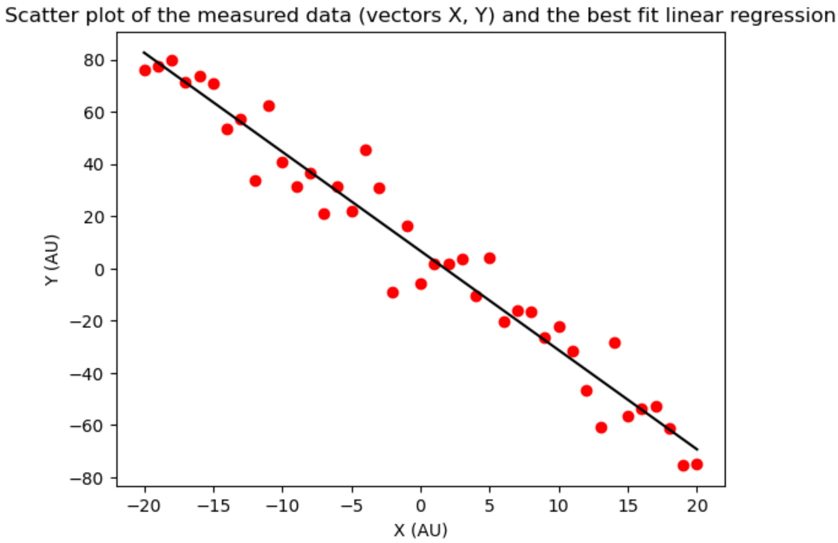
Question 1 – Linear Models and Bootstrapping

a. *Linear regression model:*

The estimated value for the slope and intercept that we've found are as follows:

```
The min MSE for this beta vector is:  96.10310512036297
The slope and intercept value with the lowest MSE: a=-3.8, b=6.7
The best fit linear regression is: y=-3.8*x+6.7
```

Figure 1:

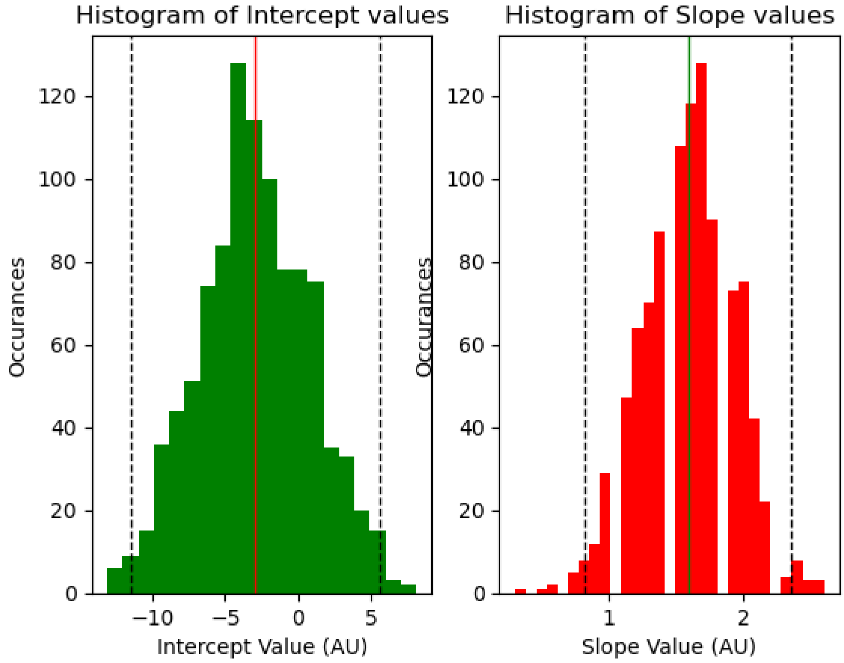


Estimating the STD of the noise distribution: In our case, the noise is defined as the difference between the predicted and measured values. So, we've calculated a distribution of those differences and its STD (here we're dealing with a sample and not a population, therefore the df=n-1).

```
Estimated Noise STD: 9.92499935184865
```

b. *Bootstrapping:*

Figure 2: Histograms of the best slope and intercept estimates for 1000 generated samples



The average value for the estimated slopes is 1.58 , and -2.75 for the intercepts, as is denoted by the solid line in Figure 2.

Answer for bonus: the dotted lines in Figure 2 represent the 97.5% CI for each distribution.

```
The boundries of the CI for the intercepts are:  -11.417141795340003 5.55714179534001
The boundries of the CI for the slopes are:  0.8238823920782311 2.3609176079217695
```

Question 2 – GLM

The set of parameters θ : [(17,6,-19,48),(1,4,-7,1),(3,0.02,0.1,-1)] the LL values are presented in a corresponding order.

a. To find the best set of parameters θ, we need to calculate the log likelihood of each set and chose the set with the highest LL value.

```
The calculated Log-Likelihood values are: [-5043703341.508967, 453.0694004659954, 411.3052242236605]
The theta vector with the max LL is: (1, 4, -7, 1)
```

b. This regularization factor "punishes" high θ values. After adding the regularization factor

```
The calculated new Log-Likelihood values are: [-5043704836.508967, 419.5694004659954, 406.3000242236605]
The theta vector with the max LL is: (1, 4, -7, 1)
```

as we can see, the LL values changed but the θ set with the highest one remains the same – the answer didn't change.

c. To set up the Poisson regression, we've used a *Log* link function (the default of the Poisson family in *statsmodels* GLM) and a noise distribution from the *Poisson-family*.

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  122
Model:                            GLM   Df Residuals:                      118
Model Family:                 Poisson   Df Model:                            3
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -294.21
Date:                Tue, 13 Dec 2022   Deviance:                       118.31
Time:                        18:44:51   Pearson chi2:                     114.
No. Iterations:                     4   Pseudo R-squ. (CS):            0.01275
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.1681      0.259      8.370      0.000       1.660       2.676
x1             2.0390      3.443      0.592      0.554      -4.709       8.787
x2            -3.3974      3.920     -0.867      0.386     -11.080       4.285
x3             1.6671      1.732      0.963      0.336      -1.728       5.062
==============================================================================
The Theta vector is: [ 2.16805029  2.03900614 -3.39743477  1.66713681]
```

d. The first value is the intercept. From the rest, as we're told that the X values are neurons: X_1, X_2, X_3 influencing our neuron's Y FR. Therefore, we can infer that negative θ values represent inhibitory inputs while positive ones represent excitatory inputs.
X_1 is excitatory, X_2 is inhibitory and X_3 is excitatory as well. Furthermore, neuron X_2 has the highest influence on Y's FR, as it has the highest absolute coefficient value.