HW2 – DS for Mechanical Systems

Sachin Francis Dsouza
UNI: sfd2121
October/19/2021

Homework #2

0.1 Problem 1 (20 points)

Given the data in this dataset, it consists of 2 columns and 200 rows. If we treat the first columnsas x, and the second column as y, we want to build a simple linear regression model: $y = \beta_0 + \beta_1 x$ to fit the data:

1. What is the point estimate of the intercept β_0 and the slope β_1, respectively?
    - Point Estimate: Intercept β_0 = -31.8043 and the slope β_1 = 16.2056.

2. What is the standard error for the intercept and the slope, respectively?
    - Standard Error: Intercept β_0 = 5.7635 and the slope β_1 = 0.5484.

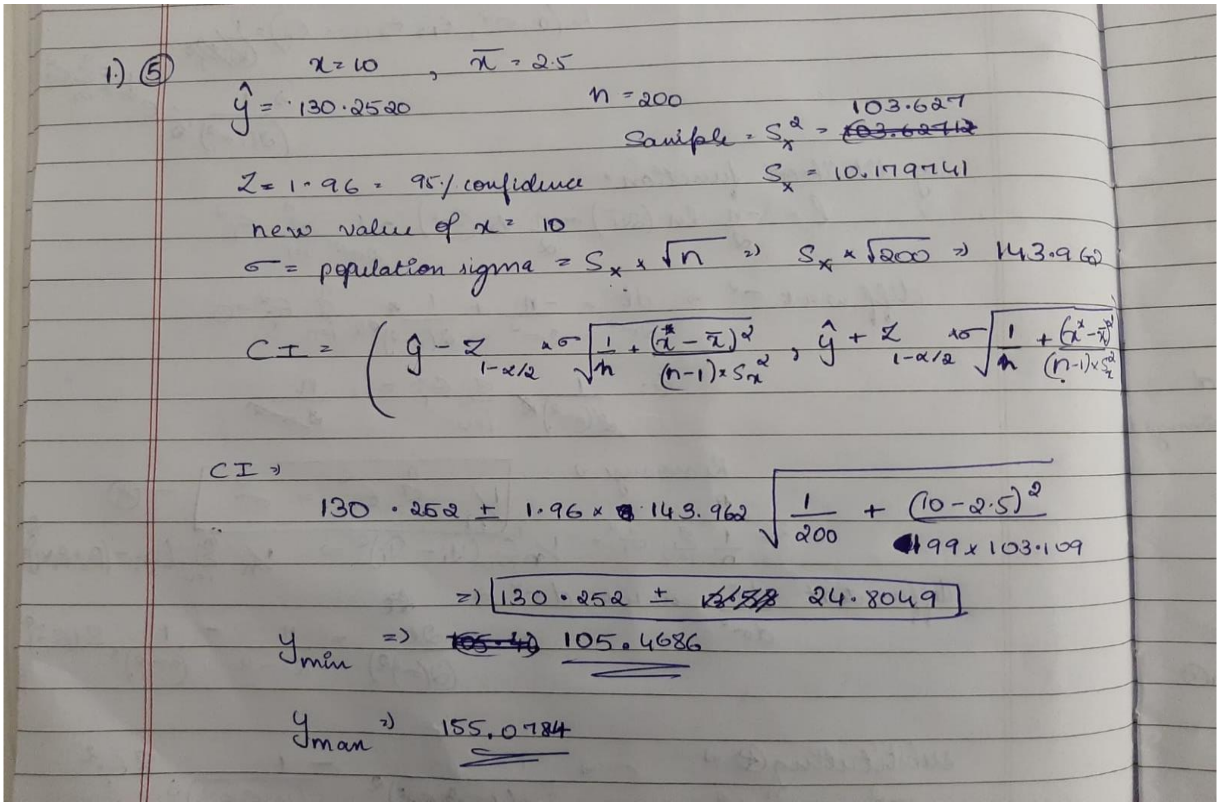
3. Is the result significant, i.e., the slope is significantly different from zero?

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.814
Model:                            OLS   Adj. R-squared:                  0.813
Method:                 Least Squares   F-statistic:                     864.6
Date:                Mon, 18 Oct 2021   Prob (F-statistic):           3.61e-74
Time:                        07:07:48   Log-Likelihood:                -1157.0
No. Observations:                 200   AIC:                             2318.
Df Residuals:                     198   BIC:                             2325.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -31.8043      5.764     -5.518      0.000     -43.170     -20.438
x             16.2056      0.551     29.404      0.000      15.119      17.292
==============================================================================
Omnibus:                        2.158   Durbin-Watson:                   1.638
Prob(Omnibus):                  0.340   Jarque-Bera (JB):                1.747
Skew:                           0.071   Prob(JB):                        0.418
Kurtosis:                       2.565   Cond. No.                         10.8
==============================================================================
```

- From the OLS Regression results, the 95% Confidence Interval for β_0 = [-43.170, -20.438] and 95% Confidence Interval for β_1 = [15.119, 17.292]. The P-Value for both the Intercept and Independent variable is tending to 0, and this means the null hypothesis can be rejected. Since the Class Intervals do not have zero, the correlation between the two doesn't exist. Thus, we can conclude that the slope β_1 is significantly different from Zero.

4. What is the R^2 of your model?
   - $R^2 = 0.8137$.

5. For a new x value of 10, with your fitted model, what is the predicted y, and its 95% confidenceinterval?
   - Predicted y-value for x = 10: 130.2520



1.) 5

$x = 10$, $\bar{x} = 2.5$

$\hat{y} = 130.2520$ $n = 200$

Sample $S_x^2 = 103.627$ ~~103.62712~~

$Z = 1.96 = 95\%$ confidence $S_x = 10.179741$

new value of $x = 10$

σ = population sigma $= S_x \times \sqrt{n} \Rightarrow S_x \times \sqrt{200} \Rightarrow 143.962$

$$CI = \left(\hat{y} - Z_{1-\alpha/2} \times \sigma \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1) \times S_x^2}},\ \hat{y} + Z_{1-\alpha/2} \times \sigma \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1) \times S_x^2}}\right)$$

$$CI \Rightarrow 130.252 \pm 1.96 \times 143.962 \sqrt{\frac{1}{200} + \frac{(10-2.5)^2}{199 \times 103.109}}$$

$$\Rightarrow \boxed{130.252 \pm 24.8049}$$

$y_{min} \Rightarrow$ ~~105.40~~ 105.4686

$y_{max} \Rightarrow 155.0784$

   - From Code: Y min = 116.6153533 & Y max = 143.888859

0.2 Problem 2 (20 points)

Using the same dataset, now we want to build a different linear regression model as: $y = \beta_0 + \beta_1 x + \beta_2 x^2$ to fit the data:

1. What is the point estimate of the β_0, β_1, and β_2, respectively?
   - Point Estimate: $\beta_0 = -18.0505$, $\beta_1 = 16.9156$ and $\beta_2 = -0.1420$.

2. What is the standard error for the three coefficients, respectively?
   - Standard Error: $\beta_0 = 8.138$, $\beta_1 = 0.622$ and $\beta_2 = 0.060$.

3. Are β_1 and β_2 significantly different from zero?
   - From the OLS Regression results, the 95% Confidence Interval for β_0 = [-34.100, -2.001], 95% Confidence Interval for β_1 = [15.689, 18.142] and the 95% Confidence Interval for β_2 = [-0.260, -0.024]. The P-Value for β_0, β_1 and β_2 are 0.028, 0.000 and 0.019, and since the P-Values are < (less than) 0.05, means the null hypothesis can be rejected. Since the Class Intervals do not have zero, the correlation between the three doesn't exist. Thus, we can conclude that β_1 and β_2 are significantly different from Zero.

4. What is the R^2 of your model?
   - $R^2 = 0.8188$.

0.3 Problem 3 (10 bonus points)

Show that, for the simple linear regression $y = \beta_0 + \beta_1 x + \epsilon$, if the error term ϵ is assumed to follow a normal distribution of $N(0, \sigma^2)$, then maximizing the likelihood leads to minimizing the mean squared error of $\sum_{i=1} (y_i - \hat{y}_i)^2$, where $\hat{y}_i = \beta_0 + \beta_1 x_i$.

(Hint: what is the likelihood of observing an error of particular error ϵ_i?)

Problem ③: $y_n = \beta_0 + \beta_1 x + \epsilon$ $\quad \epsilon \sim N(0, \sigma^2)$

$$MSE = \sum_{i=1} (y_i - \hat{y}_i)^2 \qquad \hat{y}_i = \beta_0 + \beta_1 X_i$$

For any input (x_i, y_i), ~~Error~~ $\Rightarrow \epsilon_i = y_i - (\beta_0 + \beta_1 x)$

$\Rightarrow y_i - \hat{y}_i$

Assume $\Rightarrow$

$$\epsilon_i = y_i - \hat{y}_i \sim N(0, \sigma^2)$$

For any ϵ_i, $f(\epsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \sum_{i=0}^{1} (\epsilon_i)^2}$

Likelihood function $\Rightarrow$

$$L(0, \sigma^2, \epsilon_1, \epsilon_2, \ldots \epsilon_n) = \frac{1}{\cancel{(2\pi\sigma^2)^{n/2}}}$$

$$\Rightarrow \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \epsilon_i^2}$$

Log likelihood function:

$$l = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \epsilon_i^2$$

Scanned with CamScanner

Log likelihood function:

$$l = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\epsilon_i^2$$

diff w.r.t σ^2 $\Rightarrow$ $$\frac{dl}{d\sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=0}^{n}\epsilon_i^2 = 0$$

$$\frac{1}{\cancel{2}(\sigma^2)^{\cancel{2}}}\sum_{i=0}^{n}\epsilon_i^2 = \frac{n}{\cancel{2}\cancel{\sigma^2}}$$

Rearrange $\Rightarrow$ $\boxed{\frac{1}{n}\sum_{i=0}^{n}\epsilon_i^2 = \sigma^2}$ — (A)

$$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n}\epsilon_i^2 = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 \Rightarrow \frac{1}{n}\sum_{i=0}^{m}(y_i - (\beta_0 + \beta_1 x_i))^2$$

differentiat $\frac{dl}{d\sigma^2}$ w.r.t $d\sigma^2$ $\Rightarrow$

$$\frac{\partial l}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{(\sigma^2)^3}\sum(\epsilon_i)^2$$

substituting (A) $\Rightarrow$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2$$

$$\Rightarrow \frac{n}{2\times\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\right)^2} - \frac{1}{\left(\frac{1}{n}\sum\epsilon_i^2\right)^3}\sum\epsilon_i^2$$

$$\Rightarrow \frac{n^3}{2\left(\sum_{i=0}^{n}\epsilon_i^2\right)^2} - \frac{n^3}{\left(\sum_{i=0}^{n}\epsilon_i^2\right)^2} = \frac{-n^3}{2\left(\sum_{i=0}^{n}\epsilon_i^2\right)^2} < 0$$

$\therefore$ likelihood error term is maximized when. MSE is min. ~~squared~~

$$\sigma^2 = \frac{1}{n}\sum(y_i - \hat{y}_i)^2$$

Scanned with CamScanner

Thus when Maximum likelihood is maximized the Mean square error is minimum

$$\sigma^2 = \frac{1}{n}\left(\sum_{i=0}^{n}(y_i - \hat{y}_i)^2\right) \rightarrow \text{Mean square error}$$

$\rightarrow MSE \rightarrow \sigma^2$

$$f(Y_1 \cdots Y_n) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)}$$

$\therefore$ when $e^{-(\text{min})}$ $\rightarrow$ leads to maximizing likelihood

• Simple linear regression $y = \beta_0 + \beta_1 x + \epsilon$, if the error term ϵ is assumed to normal distribution of $N(0, \sigma^2)$, then maximizing the likelihood leads to maximizing the mean squared error of $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, where $\hat{y}_i = \beta_0 + \beta_1 x_i$

Linear regression => $y = \beta_0 + \beta_1 x + \epsilon = f(x)$

$\epsilon \rightarrow$ error term.

$\epsilon \rightarrow N(0, \sigma^2)$ - assumed to normal distribution.

Prove that => ~~M.L.E~~ Maximizing M.L.E = minimizes Mean squared error.

We have 3 unknown parameters β_0, β_1 and σ^2 and they need to be estimated by using a given sample.

MEAN SQUARE ERROR = $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ — (1) & $\hat{y}_i = \beta_0 + \beta_1 x_i$ — (2)

For any input (x_i, y_i) => $\epsilon_i = y_i - (\beta_0 + \beta_1 x)$ — (3)

$\epsilon_i = y_i - \hat{y}_i \sim N(0, \sigma^2)$

Scanned with CamScanner

For any input $(x_i, y_i) \Rightarrow \epsilon_i = y_i - (\beta_0 + \beta_1 x)$ — (3)

$$\epsilon_i = y_i - \hat{y}_i \sim N(0, \sigma^2)$$

For a fixed X_i, the distribution of Y_i is equal to $N(f(X_i), \sigma^2)$ with p.d.f

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y - f(x_i))^2}{2\sigma^2}}$$

and the likelihood function of the sequence $Y_1 \dots Y_n$ is :-

$$f(Y_1 \dots Y_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \times e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - f(X_i))^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2} \quad \text{— (4)}$$

For points $X_i, \dots X_n$ which is fixed and non random and deal with randomness from the error ϵ.

The maximum likelihood function :)

To maximize the likelihood, we need to ~~minimize~~ maximize the maximum likelihood estimates of β_0, β_1 and σ^2.

Scanned with CamScanner

Page ___

Maximum likelihood function :- From equation (4)

$$f(y_1, \ldots, y_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}$$

To maximize the function $f(y_1 \ldots y_n)$, we need to minimize from equation (4).

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \quad — (5)$$

because $\Rightarrow \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \rightarrow$ constant

$\frac{-1}{2\sigma^2} \rightarrow$ constant.

$\therefore$ From equation (5) $\Rightarrow \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$

we need to minimize it

Minimizing equation (5) $\Rightarrow \sum_{i=0}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$

From equation (2) $\hat{y}_i = \beta_0 + \beta_1 x_i$ put in (5)

Scanned with CamScanner

$i=1$

we need to minimize it

Minimizing equation (5) $\Rightarrow \sum_{i=0}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$

From equation (2) $\hat{y}_i = \beta_0 + \beta_1 x_i$ put in (5)

$$\sum_{i=0}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2 \Rightarrow \sum_{i=0}^{n} (Y_i - \hat{y}_i)^2$$ should be minimi

Thus by maximizing the M.L.E $\Rightarrow$ The equation
likelihood

$$\sum_{i=0}^{n} (Y_i - \hat{y}_i)^2$$

needs to be minimized.

$$\therefore \sum_{i=0}^{n} (Y_i - \hat{y}_i)^2 = \text{Mean square error} \Rightarrow \text{equation (1)}$$

$\therefore$ Thus proved,
Maximizing likelihood leads to minimizing the mean squared error.

Scanned with CamScanner

Page

Anotherway way:-

From equation (4)

$$f(Y_1, \dots Y_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2}$$ is the

Maximum likelihood estimator (MLE).

Mean square error => $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \longrightarrow$ equation (1)

So ~~to~~ Minimizing mean square error

$\therefore$ differentiating eq.n (1) => $\frac{d}{dx}(MSE) = \frac{1}{n}\frac{d}{dx}\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)$

$\hat{y}_i = \beta_0 + \beta_1 x_i$ => equate to the above equation.

$$\frac{d(MSE)}{dx} = \frac{1}{n}\frac{d}{dx}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

To minimizing it => The derivative should be equated to zero

$$\frac{d}{dx}(MSE) = 0 \Rightarrow \frac{1}{n}\frac{d}{dx}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 = 0$$

Scanned with CamScanner

To minimizing it $\Rightarrow$ The derivative should be equated to zero.

$$\frac{d}{dx}(MSE) = 0 \Rightarrow \frac{1}{n}\frac{d}{dx}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 = 0$$

$$\sum_{i=1}^{n}\frac{d}{dx}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 = 0 \rightarrow (6)$$

Equation (4) $\Rightarrow$ $$f(Y_1, \ldots Y_n) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 + \beta_1 x_i)^2}$$

applying ln :

$$l = \frac{-n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 + \beta_1 X_i)^2 \quad (7)$$

From (6) Residual sums of squares :-

$$RSS(\beta) = \sum_{i=1}^{N}(y_i - f(x_i))^2$$

$$\Rightarrow \sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2$$

Scanned with CamScanner

Date ____
Page ____

Dealing in terms of matrices : The geometry of least squares fitting in the $\mathbb{R}^{p+1}$ dimension space.

Denote by X the $N \times (p+1)$ matrix with each row and input vector and similarly let y be the N vector of outputs in the training set.

Residual sum of squares (RSS) $\Rightarrow$ $RSS(\beta) = (y - X\beta)^T (y - X\beta)$

$$\frac{\partial RSS}{\partial \beta} = -2X^T(y - X\beta)$$

$$\frac{\partial^2 RSS}{\partial \beta \partial \beta^T} \Rightarrow 2X^TX$$

Assuming that X has a full column rank, hence X^TX is a positive definite, we set derivative to zero $\Rightarrow$ $X^T(y - X\beta) = 0$

$\therefore$ unique sol-n $\Rightarrow$ $\hat{\beta} = (X^TX)^{-1}X^Ty$

The training inputs $\Rightarrow$ $\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$.

Thus the minimizing MSE leads to maximizing zero $\Rightarrow$ $X^T(y - X\beta) = 0$

$\therefore$ unique sol^n $\hat{\beta} = (X^TX)^{-1}X^Ty$

The training inputs $\Rightarrow$ $\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$.

Thus the minimizing MSE leads to maximizing MSE

$\therefore$ from equation ⑥ & ④

Thus minimizing mean square error $\Rightarrow$ leads to

If $\frac{\partial}{\partial x}\sum_{i=0}^{n}(y_i - (0\beta_0 + \beta_1 x_i))^2 = 0$ $\rightarrow$ ⑦

(tends to zero) $\rightarrow$ ⑦

In $f(Y_1 \ldots Y_n) \Rightarrow \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2}$

Minimizing ⑦

$f(Y_1 \ldots Y_n) \Rightarrow \left(\frac{1}{2\sqrt{\pi\sigma}}\right)^n e^{-1/2\sigma^2}$ (Minimized Mean square error)

$f(Y_1 \ldots Y_n) = \left(\frac{1}{2\sqrt{\pi\sigma}}\right)^n e^{-1/2\sigma^2 \min(MSE)}$ $\Rightarrow$ leads to maximizing $f(Y_1 \ldots Y_n)$

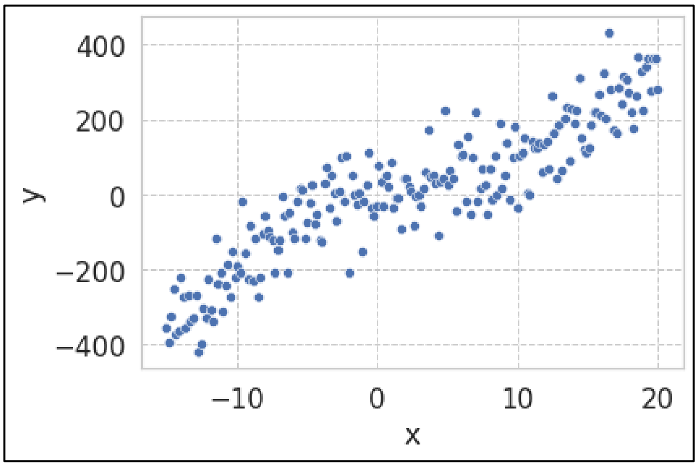
$\Rightarrow$ leads to maximizing MLE

hence proved

```
%matplotlib notebook
from typing import List
from typing import Tuple
from typing import Union

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import statsmodels.formula.api as smf

from tqdm import tqdm

sns.set(font_scale=1.5)
sns.set_style("whitegrid", {'grid.linestyle':'--'})

from google.colab import drive
drive.mount('/content/drive')


data = pd.read_csv("/content/drive/MyDrive/simple_linear_regression.csv")
data.head()


%matplotlib inline
# distribution of the dependent variable
plt.scatter(x="x", y ='y', data = data)
plt.tight_layout()
plt.show()

# correlations
continuous_variables = [
    "x",
    ]

for variable in continuous_variables:
    plt.figure()
    sns.scatterplot(x=variable, y="y", data=data)
    plt.tight_layout()

y = data["y"]
x = data["x"]

def simple_linear_regression(
    x: Union[List, np.ndarray, pd.Series],
    y: Union[List, np.ndarray, pd.Series]) -> Tuple[float, float]:
    """Return the intercept and slope of a simple linear regression."""
    beta_1 = np.cov(x, y)[0][1] / np.cov(x, x)[0][1]
    beta_0 = np.mean(y) - beta_1 * np.mean(x)

    return beta_0, beta_1

beta_0, beta_1 = simple_linear_regression(x=x, y=y)

# calculate R^2
y_pred = beta_0 + beta_1 * x
SST = np.sum(np.square(y - np.mean(y)))
residual = y - y_pred
SSE = np.sum(np.square(residual))
r2 = 1 - SSE / SST

print(f"beta_0 is: {beta_0:5.4f}")
print(f"beta_1 is: {beta_1:5.4f}")
print(f"R-square is: {r2:5.4f}")

plt.figure()
x_range = np.linspace(start=np.min(x), stop=np.max(x), num=100)
sns.scatterplot(x="x", y="y", data=data)
sns.lineplot(x=x_range, y=(beta_0 + beta_1 * x_range), color="red")
plt.tight_layout()

# confidence intervals
SE_beta_0 = (np.var(residual, ddof=2) * (1. / len(x) + (np.mean(x))**2 / np.sum((x - np.mean(x
))**2)))**0.5
SE_beta_1 = (np.var(residual) / np.sum((x - np.mean(x))**2))**0.5

print(f"The standard error for beta_0 is: {SE_beta_0:5.4f}")
```

```
print(f"The standard error for beta_1 is: {SE_beta_1:5.4f}")

# simple linear regression with the `statsmodels` library
model_1 = smf.ols(formula='y ~ x', data=data)
result_1 = model_1.fit()
print(result_1.summary())

p2 = np.polyfit(x,y,2)
print(p2)
plt.plot(x,y,'ro',label='Measured (y)')
plt.plot(x,np.polyval(p2,x),'b-',label='Predicted (y)')
plt.legend(); plt.show()

from sklearn.metrics import r2_score
r2_score(y, y_pred)

import statsmodels.api as sm
xc = np.vstack((x**2,x,np.ones(n))).T
model = sm.OLS(y,xc).fit()
predictions = model.predict(xc)
model.summary()
```

Fig 1: Scatter Plot



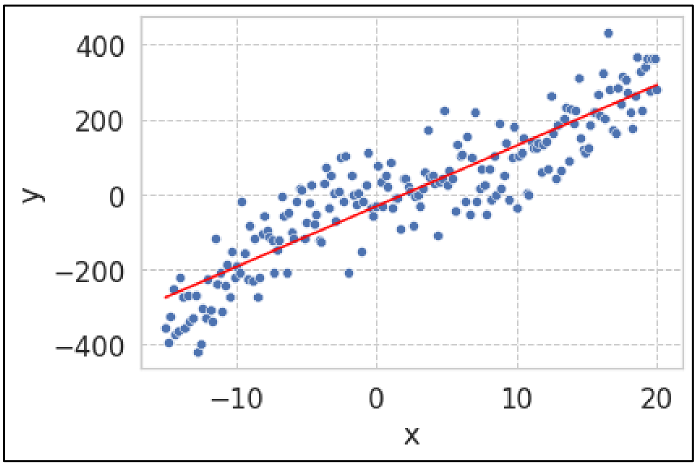


Fig 2 : Question 1 Linear Regression

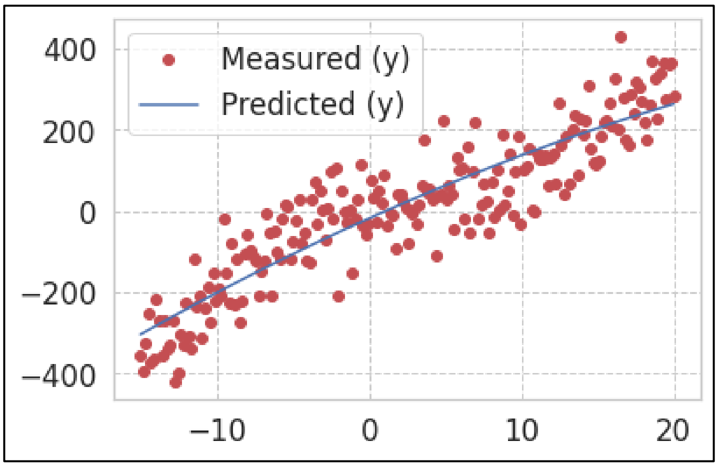


Fig 3 : Question 2 Linear Regression