BM5033 Final Exam

Submitted by:

BM23MTECH11002
BM23MTECH11006
BM23MTECH11007
EE20BTECH11016

Contents

1 Import libraries 2

2 Data Info 2
- 2.1 Columns 2
- 2.2 Conditions 2

3 Convert dat files to Dataframe 3

4 Import DataFrame 3

5 Data Analysis 4

6 Statistical Tests 31

7 Final Inference 72

8 Suggesting changes in experimental methodology. 72

1 Import libraries

[ ]:
```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import altair as alt
import scipy
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm
from scipy import stats
```

2 Data Info

2.1 Columns

- p1_lvl - Expression levels of P1 in each cell (in arbitrary units)
- p2_lvl - Expression levels of P2 in each cell (in arbitrary units)
- cell_area - Size of the molecule
- maj-axis - major axis length of each cell (approximating the shape of the cell cross-section to be elliptical)
- min_axis - minor axis length of each cell
- condition - Experimental Condition
- replication - Repitition count under each condition

2.2 Conditions

- e0x - Normal setup (WT) without any modifications
- e1x - The expression levels of the protein P1 were reduced while keeping that of P2 intact
- e2x - The expression levels of the protein P2 were reduced while keeping that of P1 intact
- e3x - The expression levels of the protein P1 were increased while keeping that of P2 intact
- e4x - The expression levels of the protein P2 were increased while keeping that of P1 intact

  x = 1, 2, 3 depending on the number of repetition

In entire code,

- e0 contains data of e01,e02,e03
- e1 contains data of e11,e12,e13
- e1 contains data of e21,e22,e23
- e1 contains data of e31,e32,e33
- e1 contains data of e41,e42,e43

---

3 Convert dat files to Dataframe

[ ]:
```
def read_data_file(path,condition, replication):
    """Reads data file based on condition, replications and returns a DataFrame

    Args:
      condition: Experimental condition.
      replication: Replication number.

    Returns: DataFrame containing the data from input file.
    """

    file_path = f'{path}{condition}{replication}.dat' # Data file path
    data=[i.strip().split() for i in open(file_path).readlines()] # splits the␣
 ↪data based on spaces and them as nested lists

    # Create DataFrame
    df=pd.DataFrame(data,columns=columns)

    # Append condition and replicaton value to respective data
    df['condition'] = condition
    df['replication'] = replication

    return df
```

[ ]:
```
columns=['p1_lvl','p2_lvl','cell_area','maj_axis','min_axis'] # Column Names
conditions = ['e0', 'e1', 'e2', 'e3', 'e4'] # Various Conditions
replications = ['1', '2', '3'] # Value representing number of replications

path="/content/drive/MyDrive/Data/Q1DataSets/Q1DataSet2/" # Replace it with your␣
 ↪File path
file_lst=os.listdir(path) # Returns list of all files in the respective directory
file_lst.sort() # Sort the list values

# Create a list of DataFrames, one for each condition and replication
df = pd.concat([read_data_file(path, c, r) for c in conditions for r in␣
 ↪replications])

df.to_csv("DataSet2.csv",index=False)
```

4 Import DataFrame

[ ]:
```
df=pd.read_csv('DataSet2.csv') # This dataframe includes all the 15 .dat file␣
 ↪combined together
df
```

```
[ ]:         p1_lvl     p2_lvl  cell_area  maj_axis  min_axis condition  \
     0     285.79645  318.38006   88.48058   9.73106   3.09819        e0
     1     228.01462  316.57763   95.36651  10.02494   3.14376        e0
     2     331.06314  352.47901   95.07759  10.57045   3.02686        e0
     3     368.32333  345.23546   57.87355   8.17404   2.39859        e0
     4     318.78085  348.54889   69.52686   9.03779   2.62459        e0
     ...         ...        ...        ...       ...       ...       ...
     2953  247.09355  508.81594   75.22882  11.25593   2.23617        e4
     2954  235.05288  506.63546   41.92317   8.41123   1.69518        e4
     2955  240.23013  483.13104   24.83344   6.33961   1.37994        e4
     2956  297.61694  490.77206   43.88618   8.48084   1.78321        e4
     2957  230.33775  621.15607   26.73256   7.44577   1.25444        e4

           replication
     0               1
     1               1
     2               1
     3               1
     4               1
     ...           ...
     2953            3
     2954            3
     2955            3
     2956            3
     2957            3

     [2958 rows x 7 columns]
```

5 Data Analysis

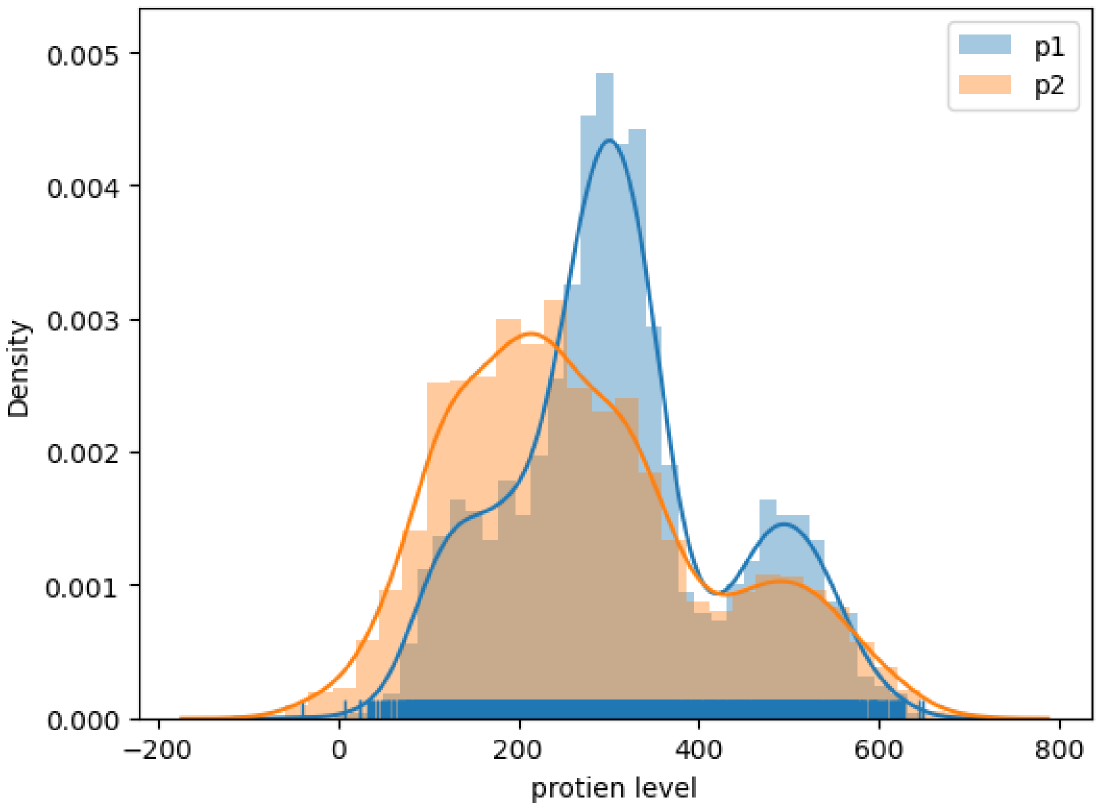
Important terminology

- distplot : It is suitable for comparing the range and distribution for groups of numerical data. Data is plotted as value points along an axis.
- heatmap : It provides an immediate visual summary of information across two axes, allowing users to quickly grasp the most important or relevant data points
- jointplot : It is a Python graph that visualizes the relationship between two variables.
- correlation : Relation between an attribute to every other attribute . It varies from -1 to 1 . Better the value better the relation .

```
[ ]: ## Distribution plot for p1 and p2 for all the 15 .dat file combined together

f, ax = plt.subplots(1, 1)

# Plot the distribution of the 'p1_lvl' column with a histogram, kernel density␣
 ↪estimate (KDE), and rug plot
sns.distplot(df['p1_lvl'],kde=True,hist=True,rug=True,label="p1")
```

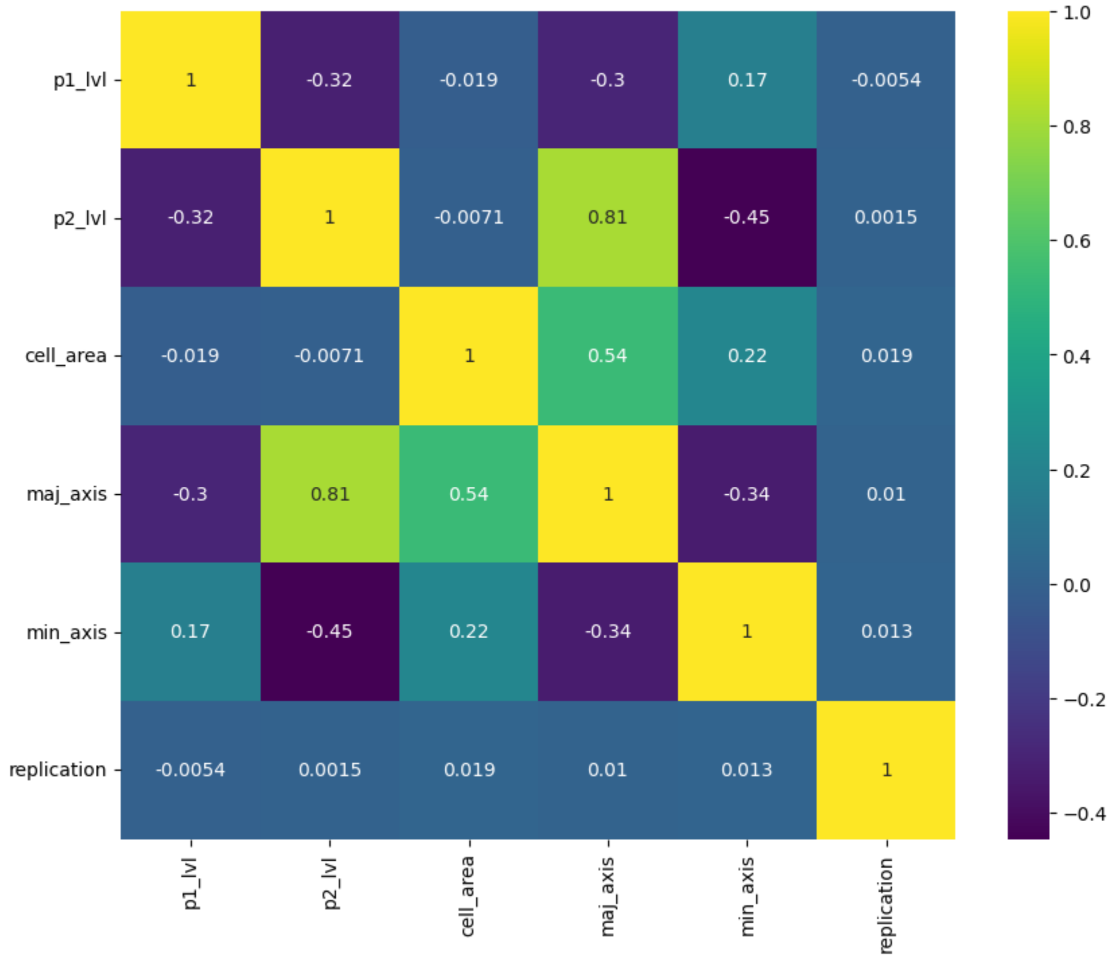
```
# Plot the distribution of the 'p2_lvl' column with a histogram and KDE, but␣
↪without a rug plot
sns.distplot(df['p2_lvl'],kde=True,hist=True,rug=False,label="p2")
plt.xlabel("protien level")

ax.legend()

plt.show() # Display plot
```

```
<ipython-input-39-de1c6d2e9e1f>:6: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['p1_lvl'],kde=True,hist=True,rug=True,label="p1")
<ipython-input-39-de1c6d2e9e1f>:9: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['p2_lvl'],kde=True,hist=True,rug=False,label="p2")
```




```
[ ]: ## heatmap for whole data (all 15 appended datasets) to have broader picture on␣
     ↪how data is correlated.

     # Compute the correlation matrix
     corr_matrix = df.corr()

     # Plot the correlation heatmap
     sns.heatmap(corr_matrix, annot=True, cmap = 'viridis')

     # Set the figure size
     fig=plt.gcf()
     fig.set_size_inches(10,8)

     plt.show() # Display plot
```

```
<ipython-input-40-91cfcd2df138>:4: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  corr_matrix = df.corr()
```



Inference : –>Close observation of the data reveals that, p1 and p2 are having near to zero corelation on cell area (size)

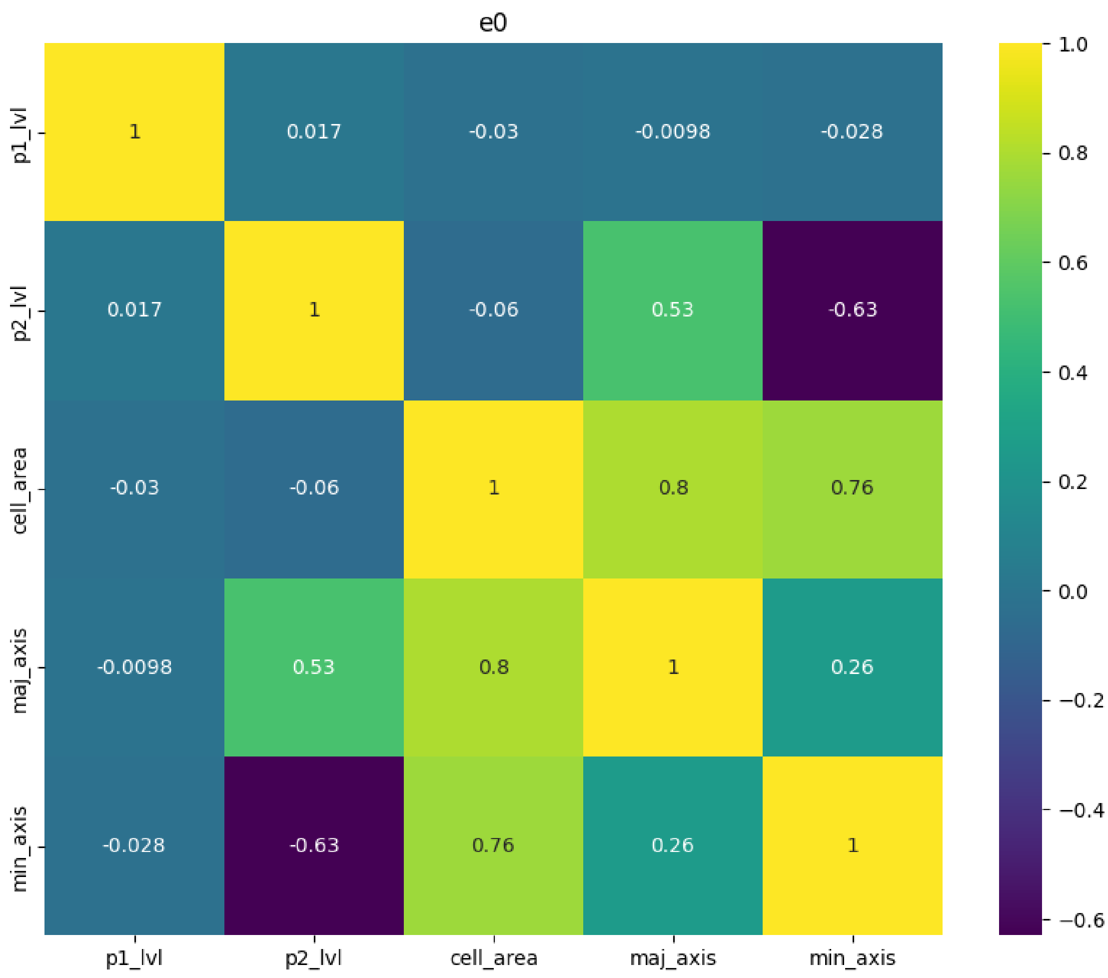
–>P1 is having impact on major & min axes.

–>P2 is having high corelation on major axis & mid negative correlation on minor axis.

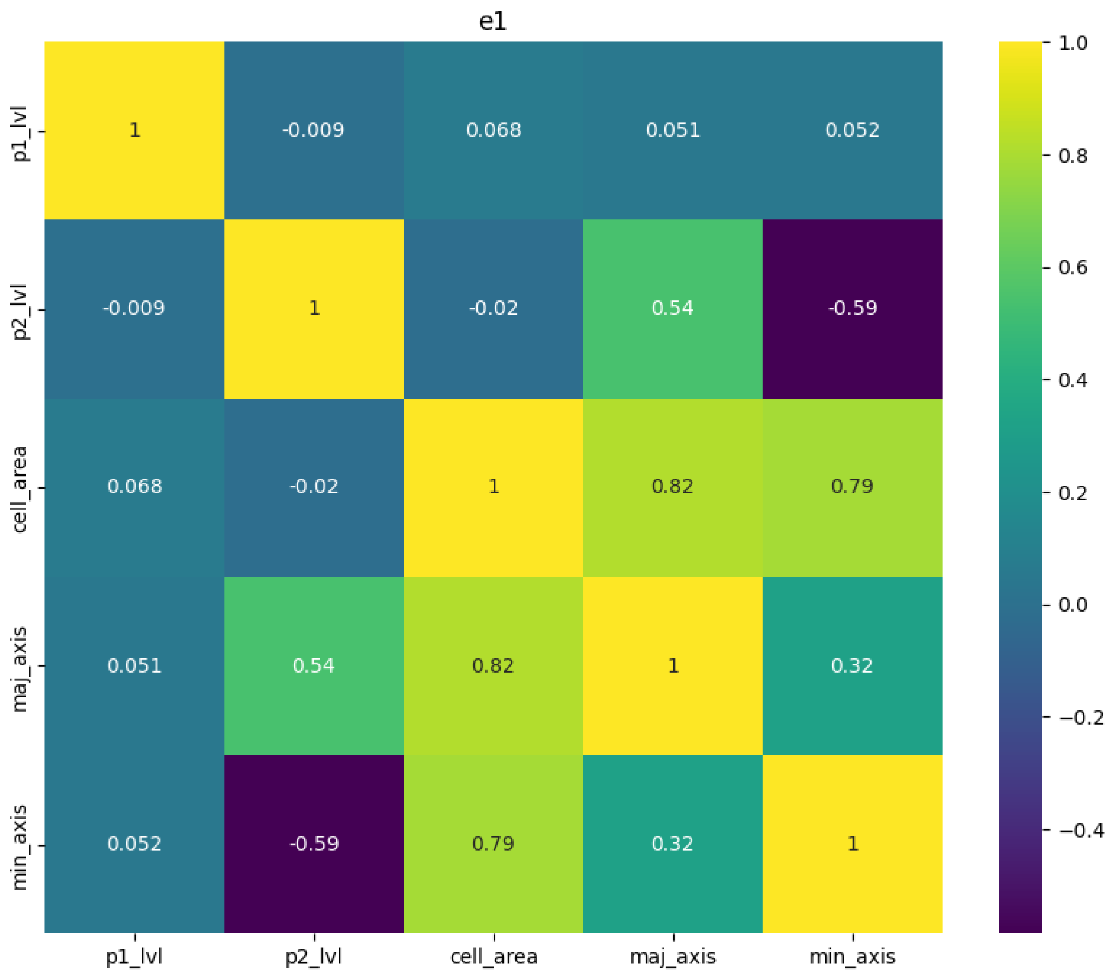
[ ]:
```
## correlation under each condition

# Iterate over the unique experimental conditions
for i in df["condition"].value_counts()[df.condition.unique()].keys():
  # Drop replication column
  df1=df.drop(["replication"],axis=1)

  # Compute the correlation matrix and plot correlaton Heatmap
  sns.heatmap(df1[df1.condition==i].corr(), annot=True, cmap = 'viridis')

  # Set the figure size and title
```

```
fig=plt.gcf()
fig.set_size_inches(10,8)
plt.title(i)

plt.show() # Display plot
```

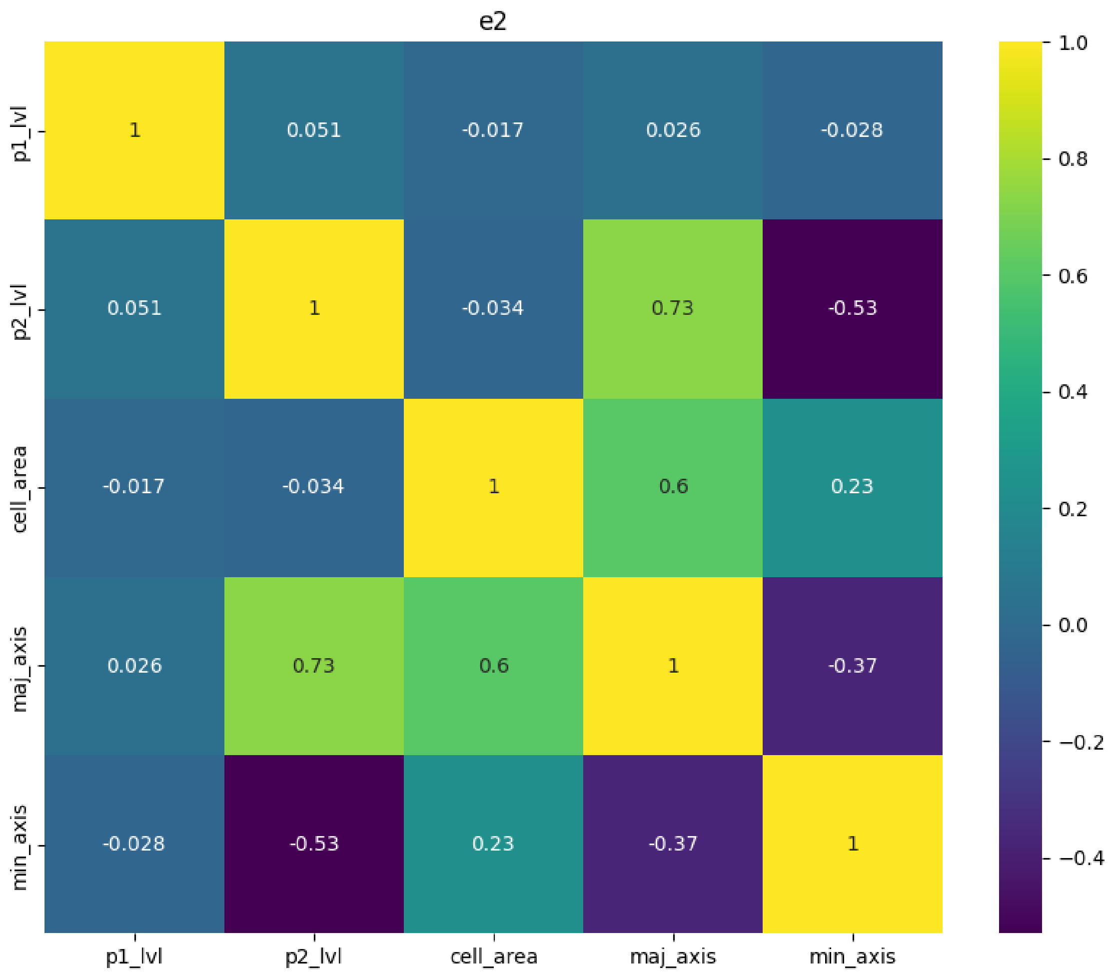
```
<ipython-input-41-521392304fce>:9: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.condition==i].corr(), annot=True, cmap = 'viridis')
```



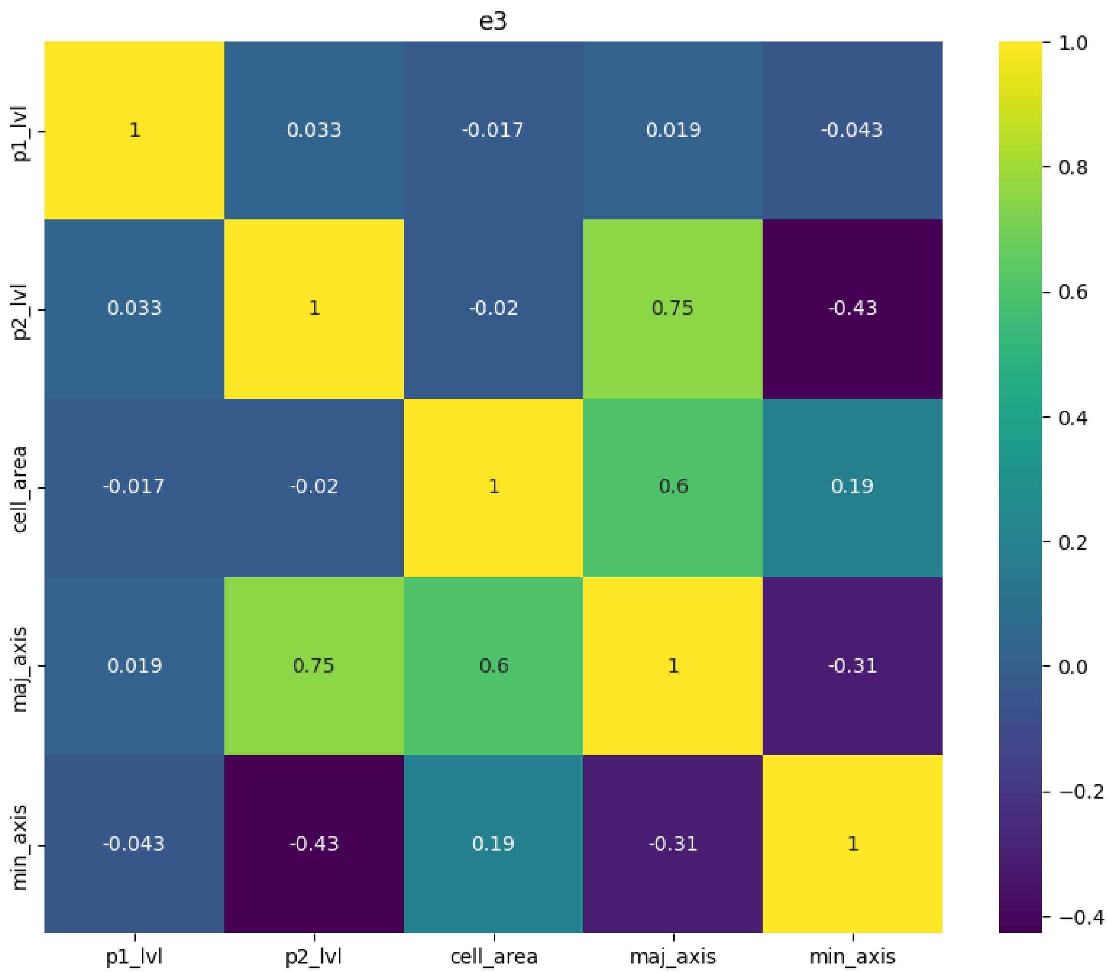

```
<ipython-input-41-521392304fce>:9: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.condition==i].corr(), annot=True, cmap = 'viridis')
```



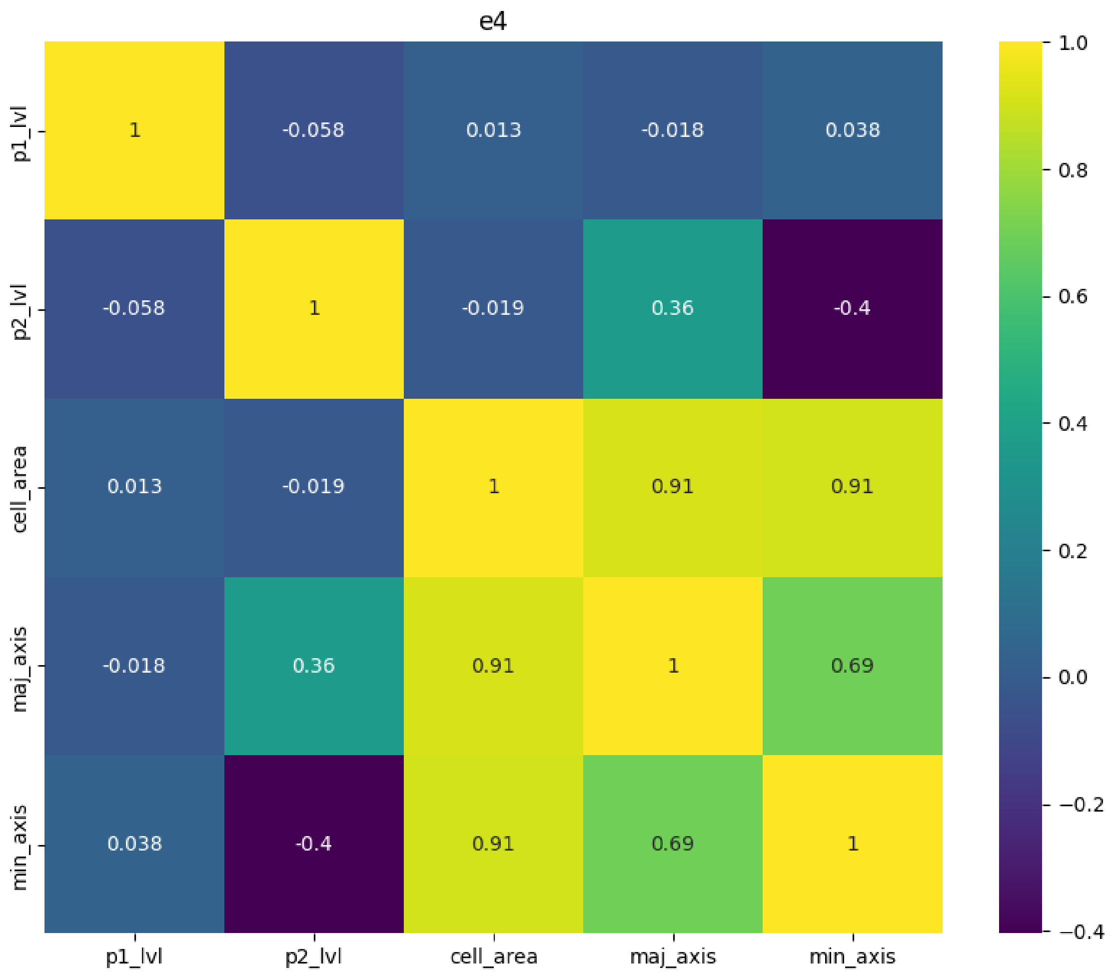

```
<ipython-input-41-521392304fce>:9: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.condition==i].corr(), annot=True, cmap = 'viridis')
```




```
<ipython-input-41-521392304fce>:9: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.condition==i].corr(), annot=True, cmap = 'viridis')
```




```
<ipython-input-41-521392304fce>:9: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.condition==i].corr(), annot=True, cmap = 'viridis')
```



Inference :

As we have 5 different experimental conditions, we will be analyzing the data with each condition using heatmap. Overall, the impact of P1 & P2 on cell area has no change and is similar across all conditions. Simillarly impact of P1 on major & minor axes follows the same. Also the impact of P2 on minor axis did not make much changes.

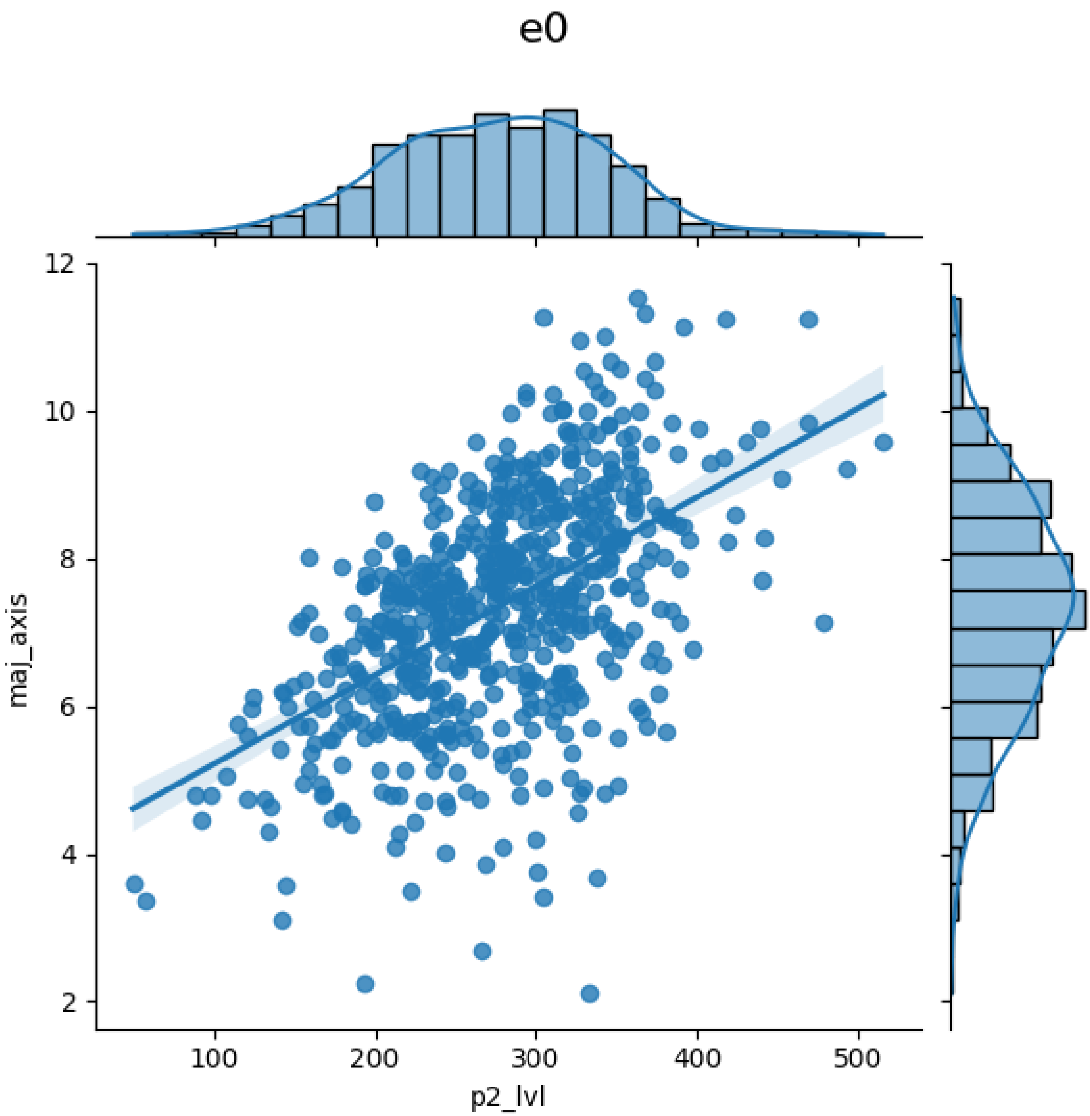
Changes were seen in P2 impact on major axis . We'll see impact in each condition :

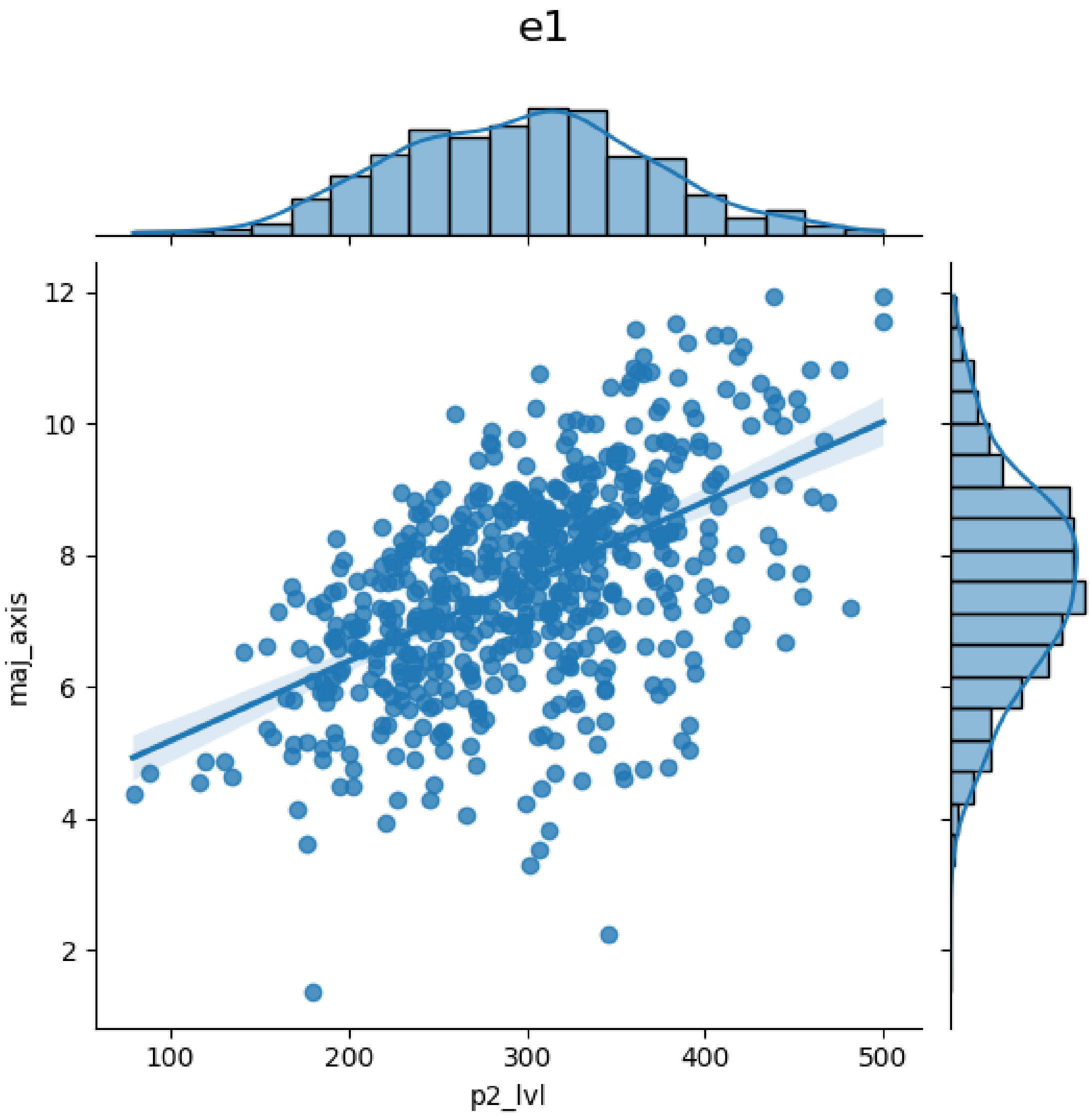
e0,e1,e4 - Here it is around mid positive corelation

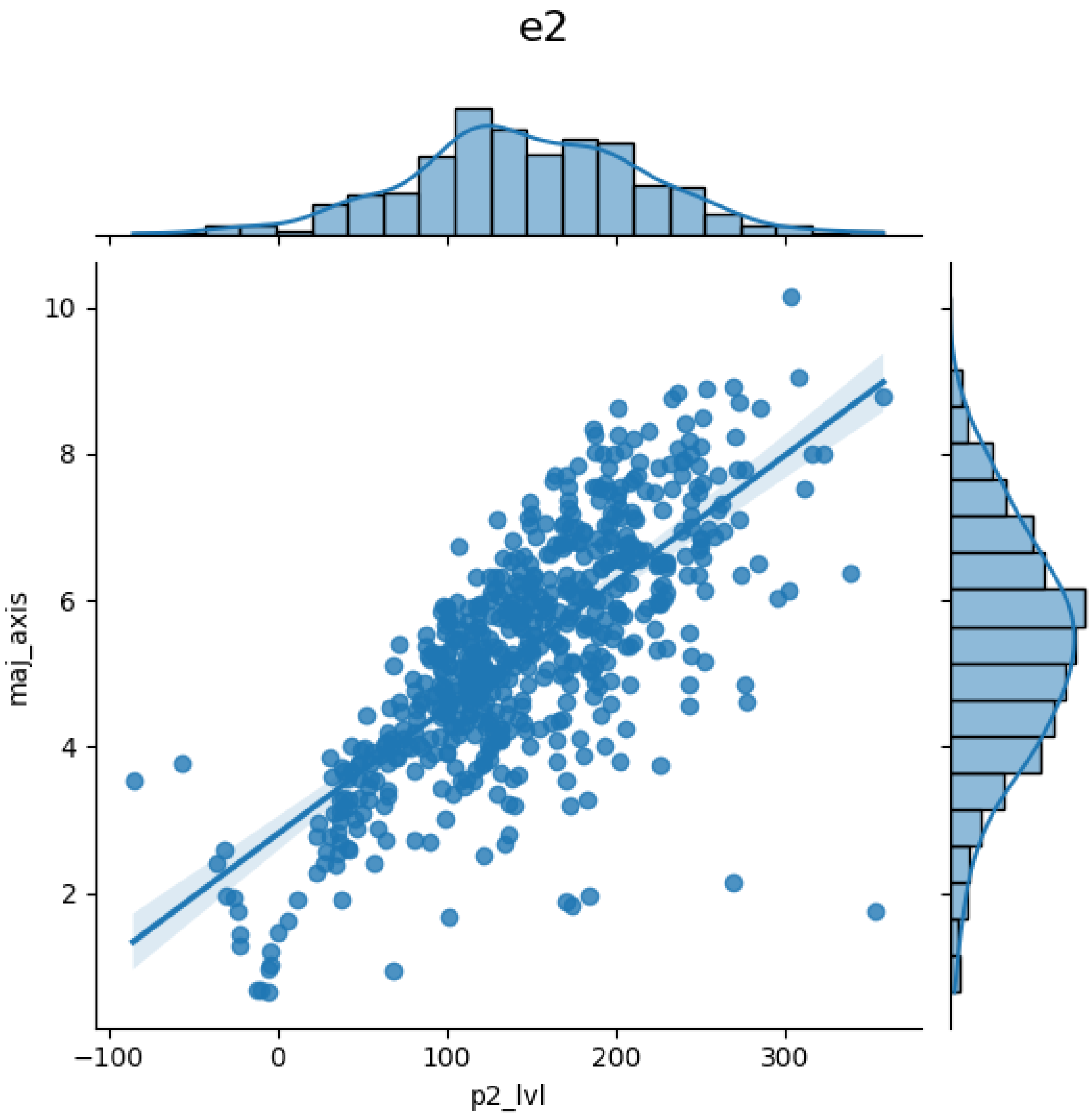
e2,e3 - It's showing high correlation in these two conditions i.e., when P2 is reduced keeping P1 intact or increasing P1 keeping P2 intact.

[ ]:
```
## p2 impact on major axis on each experimental condition (e0,e1,e2,e3,e4)

# Iterate over the unique experimental conditions
for i in df["condition"].value_counts()[df.condition.unique()].keys():
```

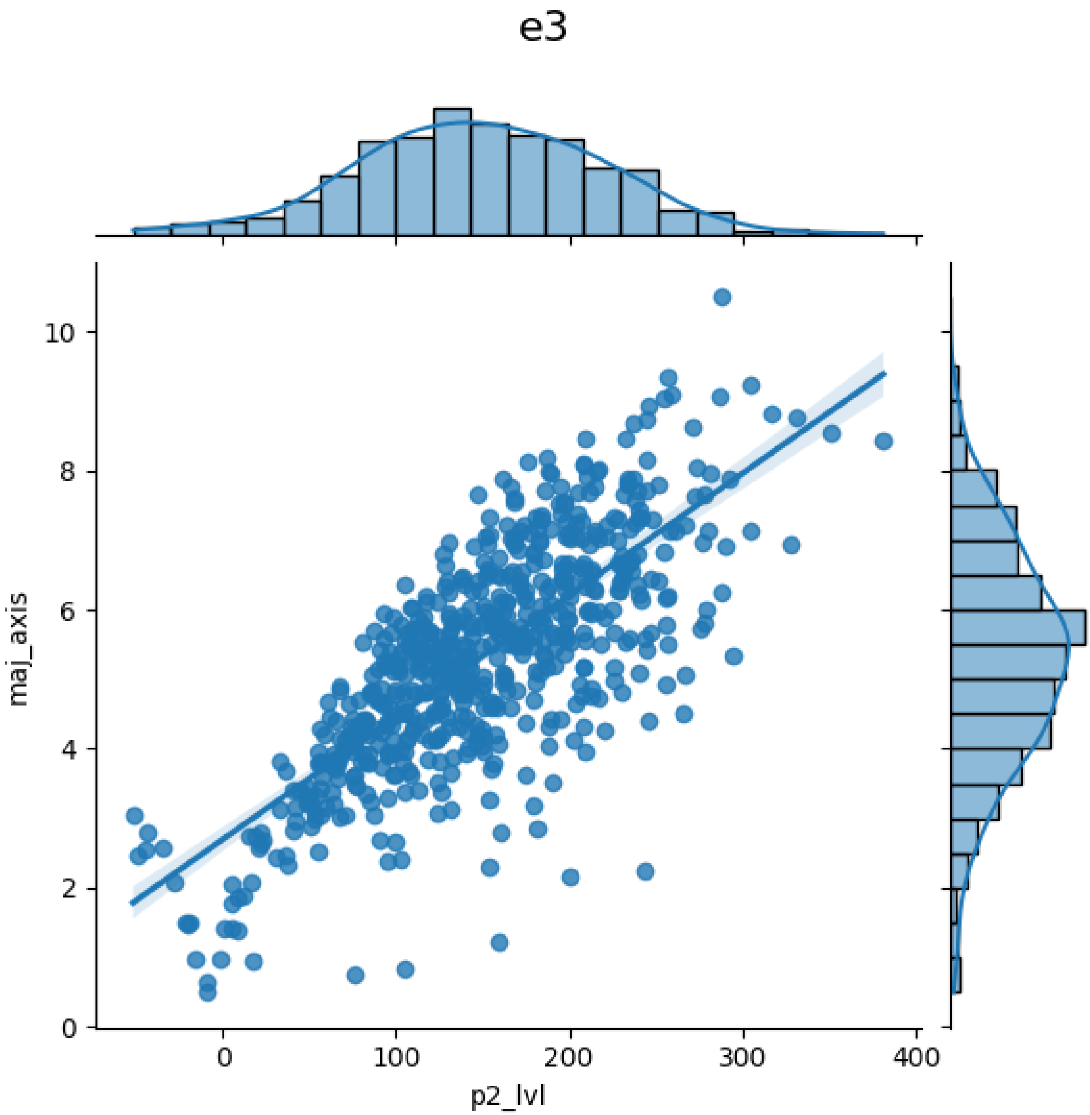
```
# Create a joint plot for the current condition
sns.jointplot(x='p2_lvl',y='maj_axis',data=df[df.condition==i],kind='reg')

# Adjust the subplots and add centred title
plt.subplots_adjust(top=0.9)
plt.suptitle(i, fontsize = 16)
```



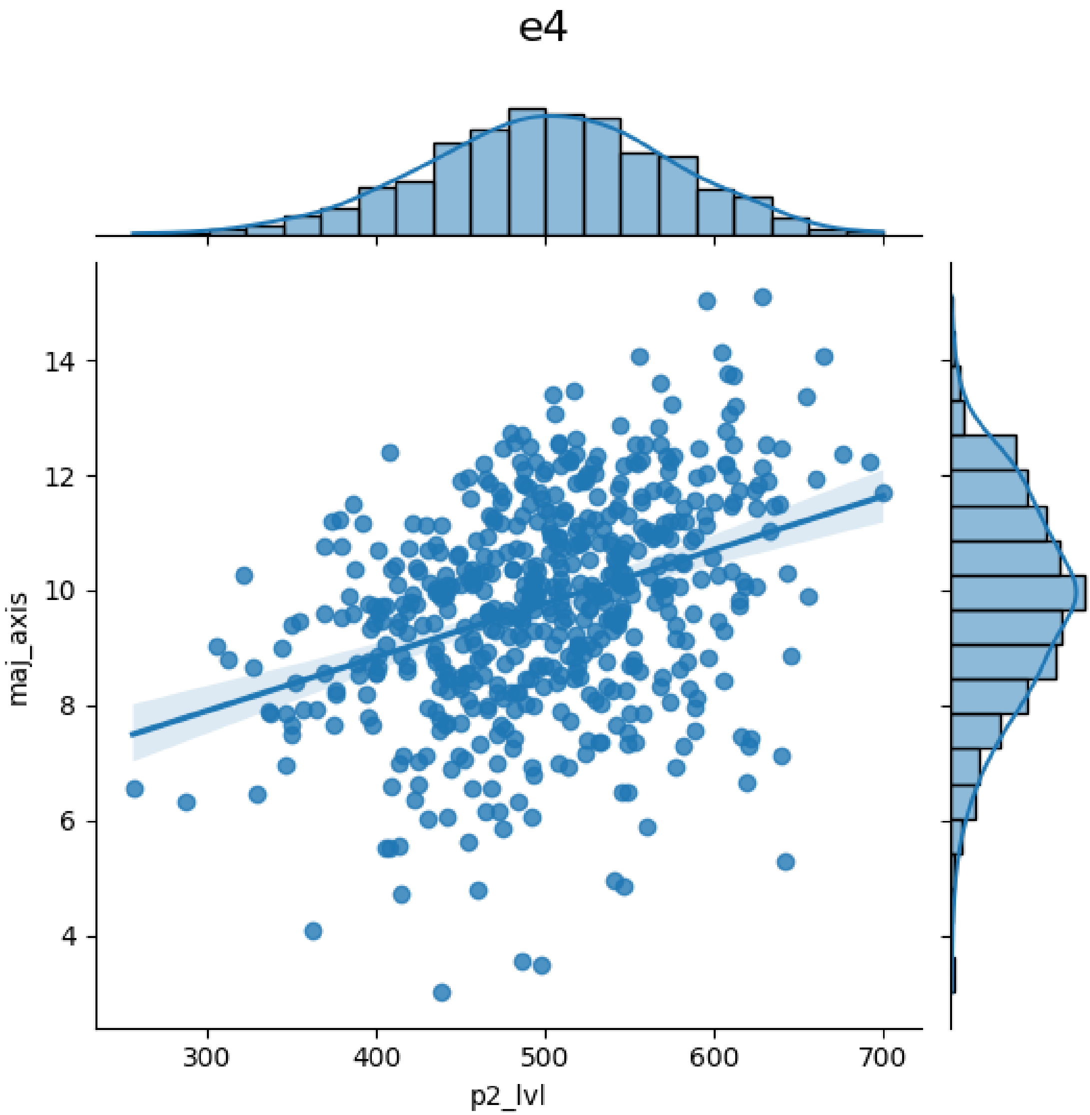

e1
maj_axis
p2_lvl
100
200
300
400
500
2
4
6
8
10
12



e2
10
8
6
4
2
maj_axis
−100
0
100
200
300
p2_lvl



e3
maj_axis
p2_lvl
0
100
200
300
400
0
2
4
6
8
10

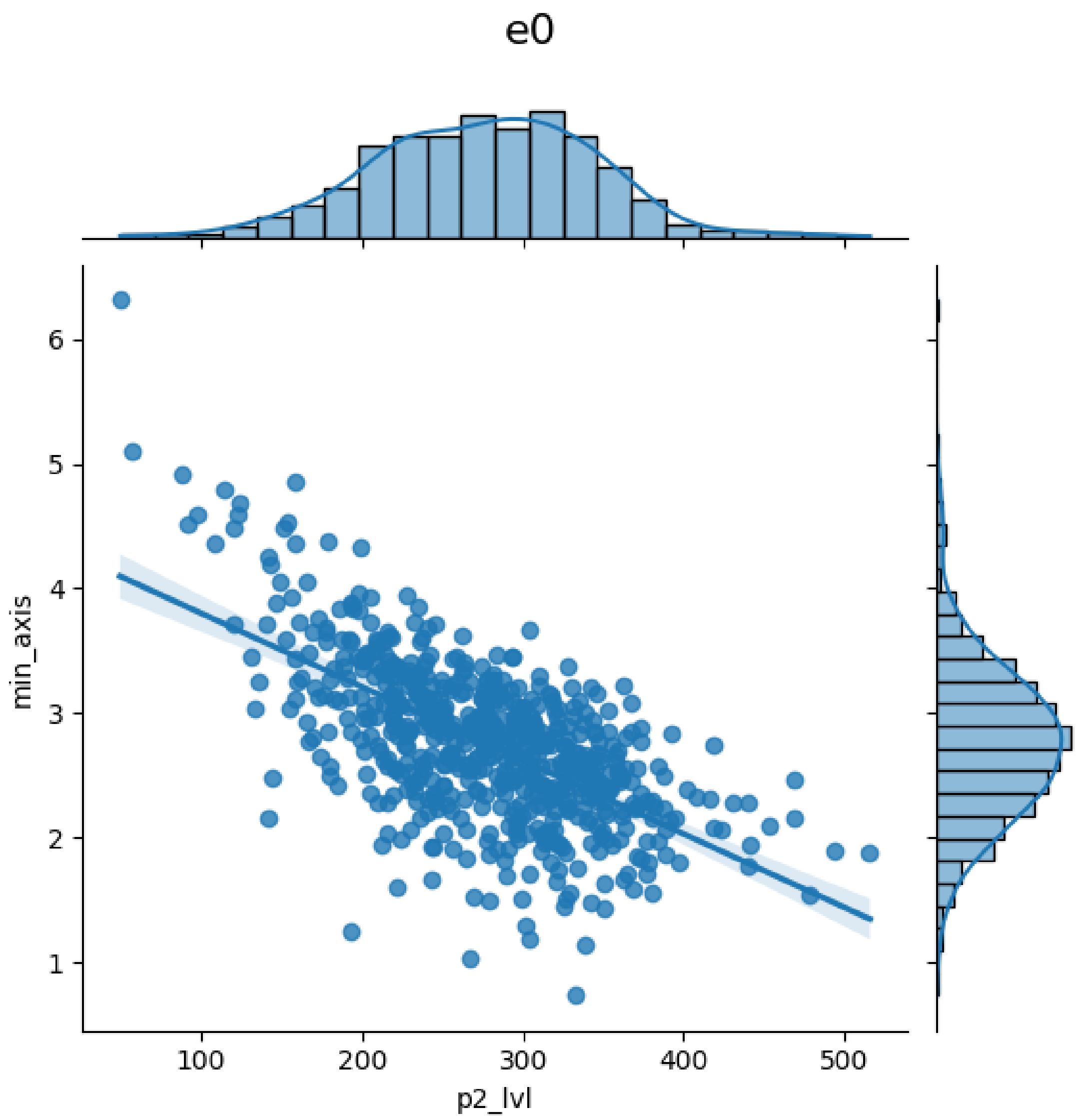


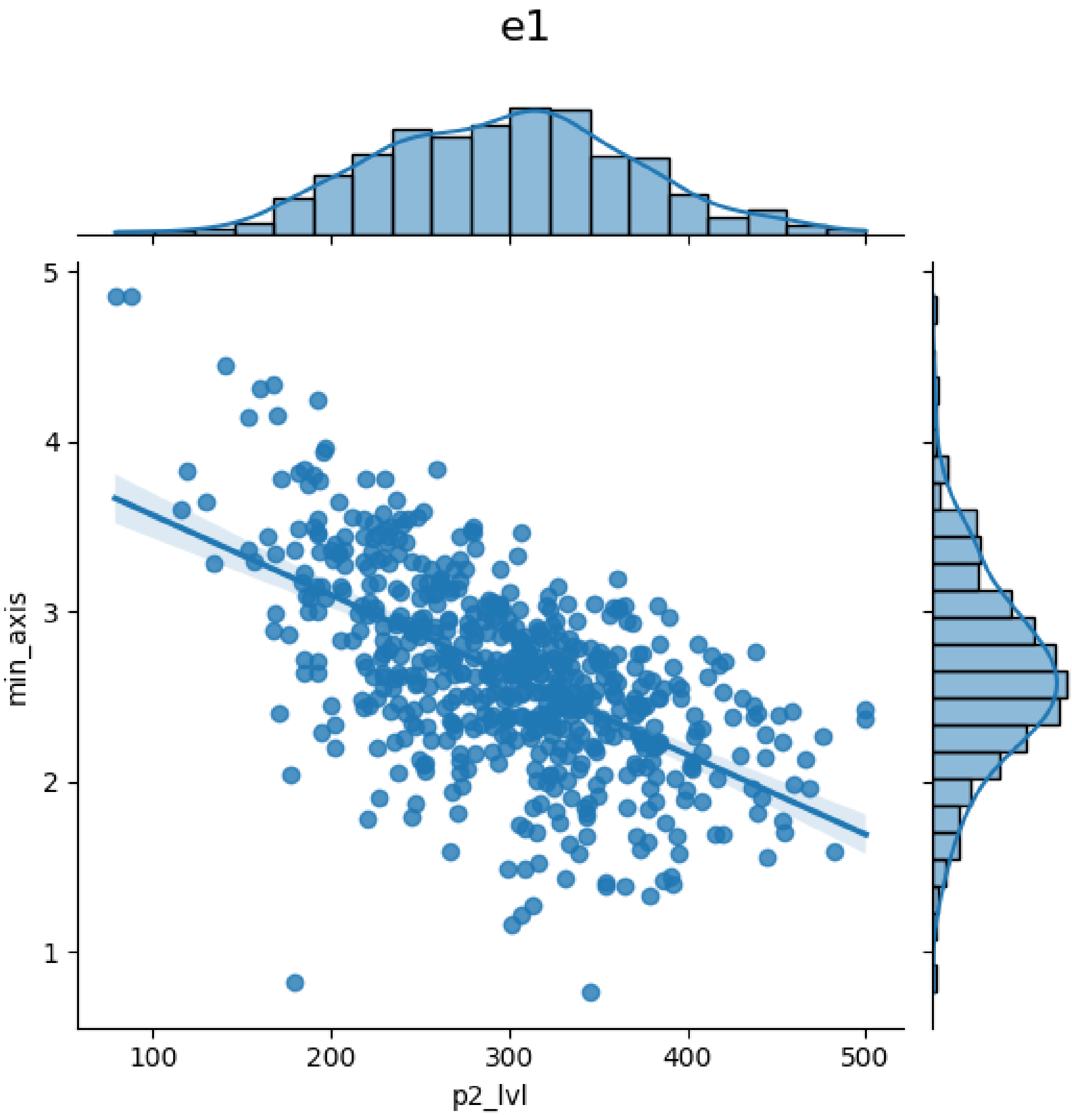
Inference :

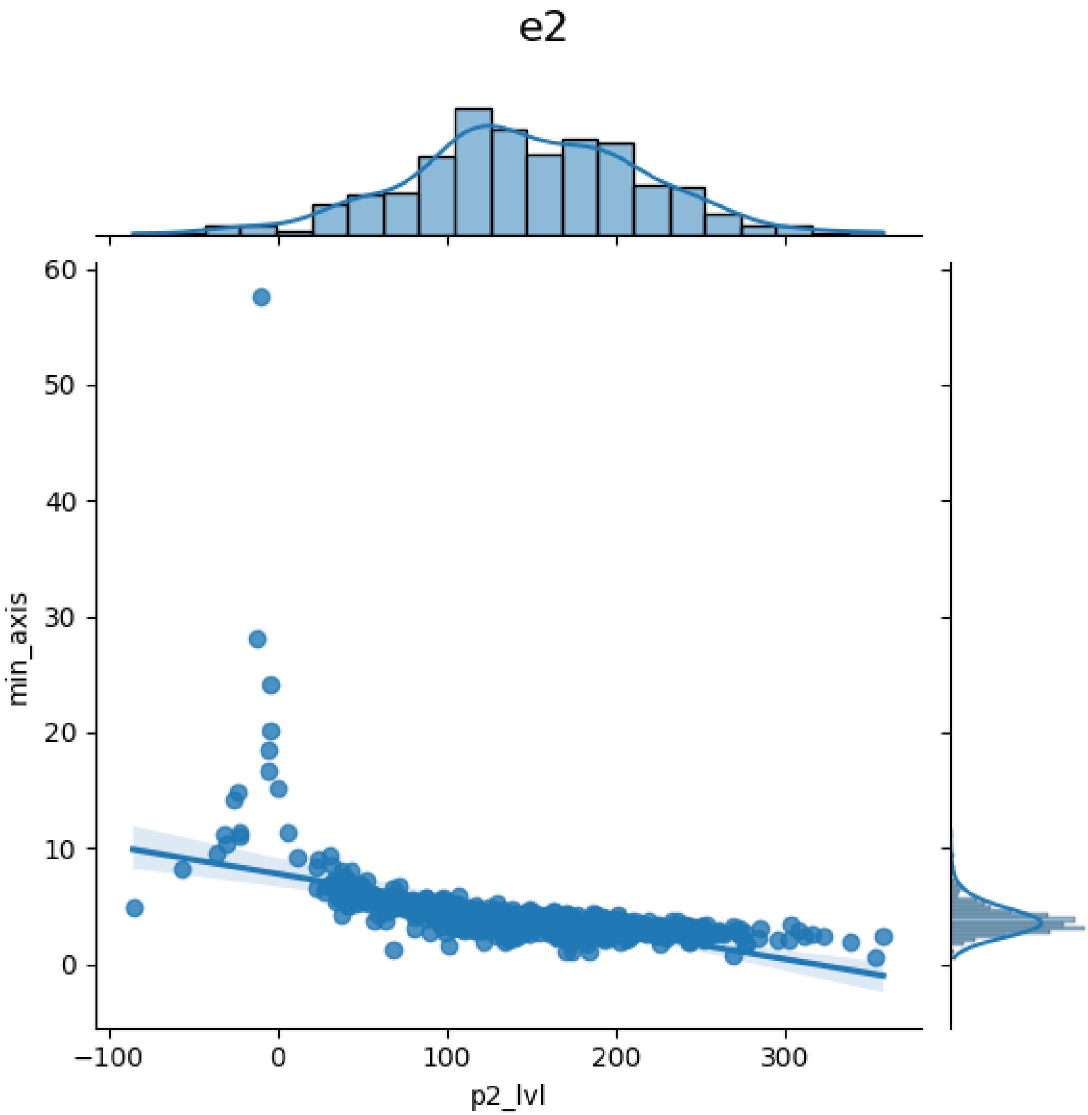
This is just a graphical represntation of correlation value we saw in heatmap. Here we considered the impact of P2 on major axis under each condition.

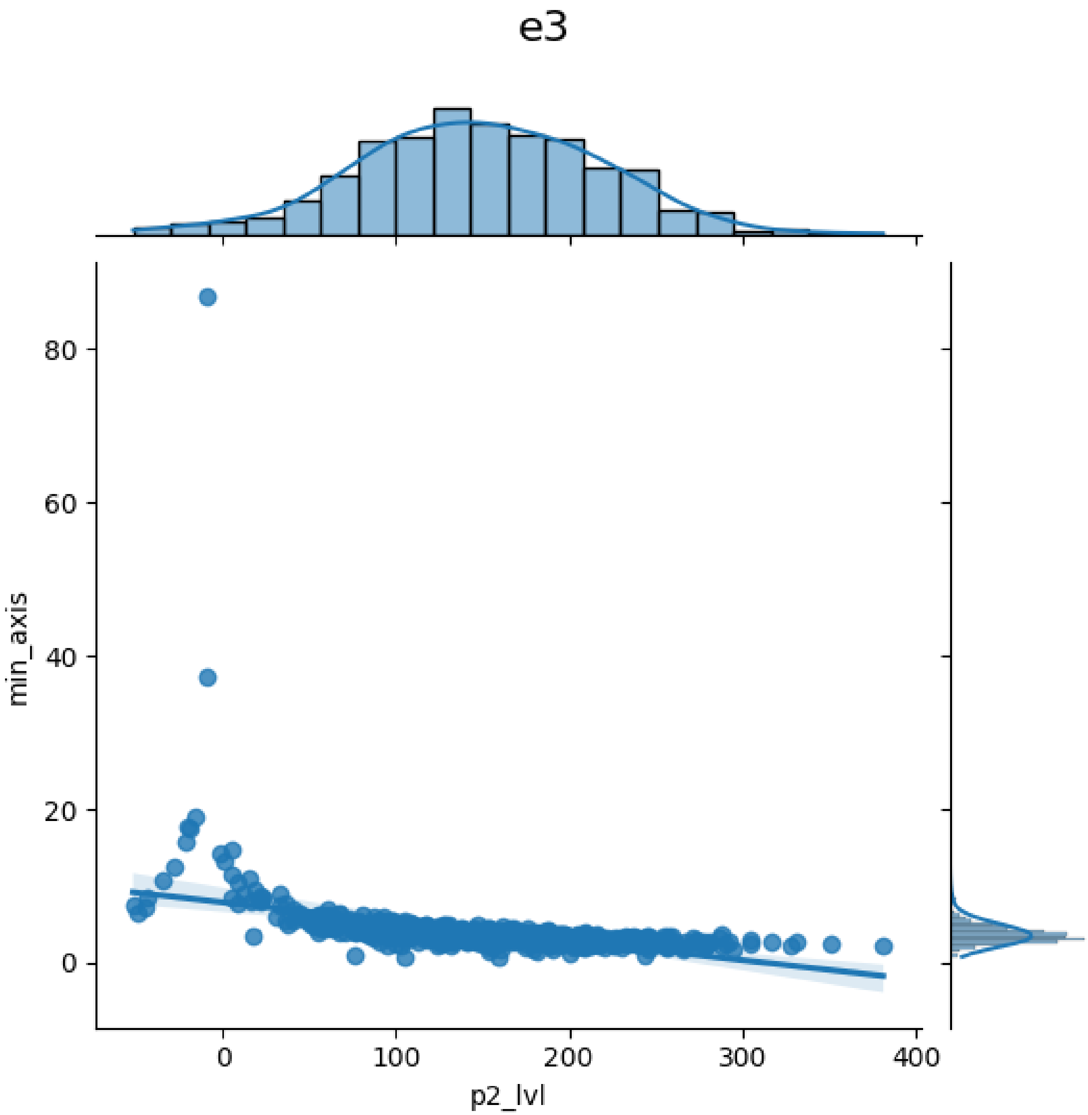
[ ]:
```
## p2 impact on minor axis on each experimental condition (e0,e1,e2,e3,e4)

# Iterate over the unique experimental conditions
for i in df["condition"].value_counts()[df.condition.unique()].keys():

  # Create a joint plot for the current condition
  sns.jointplot(x='p2_lvl',y='min_axis',data=df[df.condition==i],kind='reg')
```

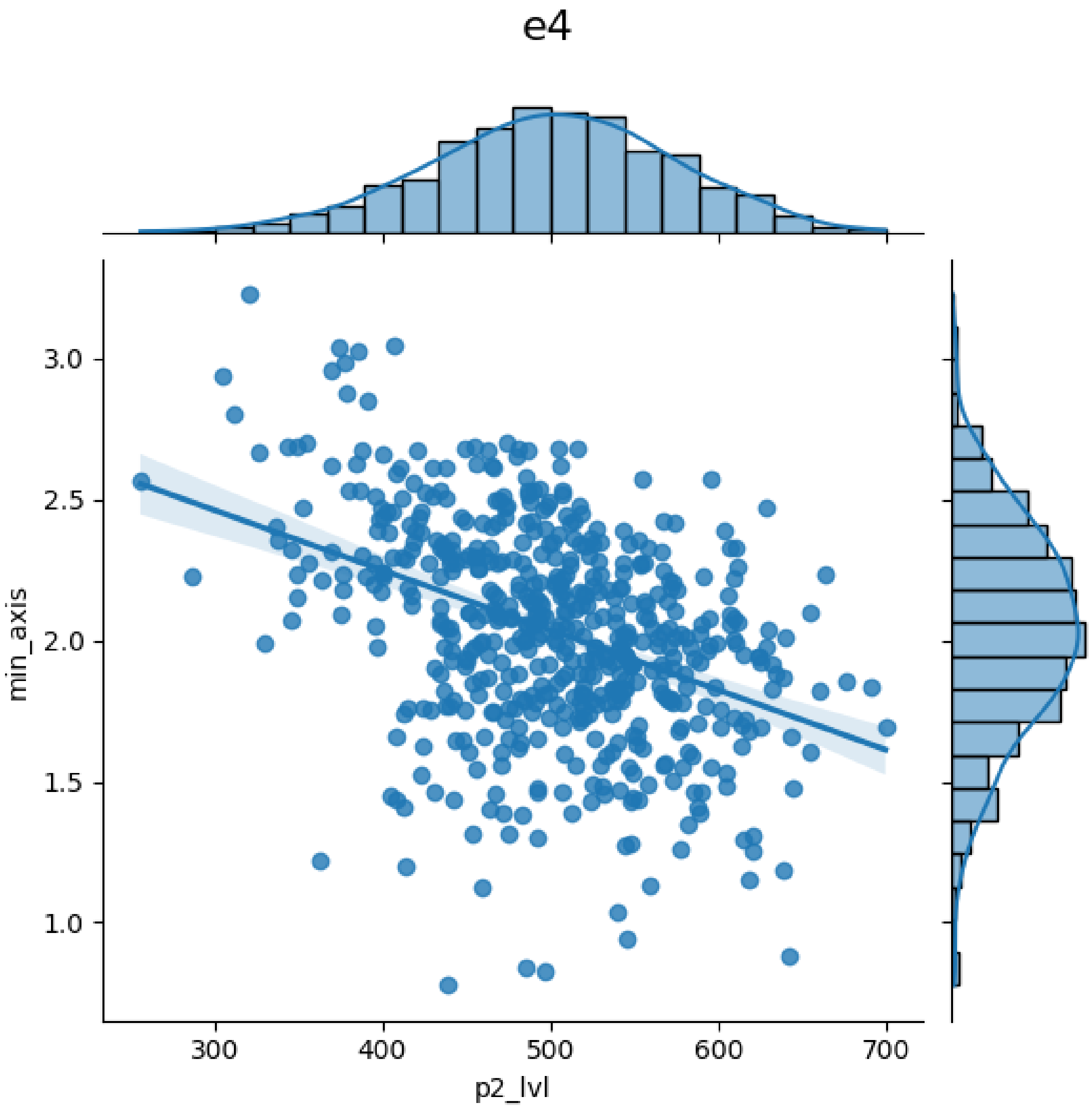
```python
# Adjust the subplots and add centred title
plt.subplots_adjust(top=0.9)
plt.suptitle(i, fontsize = 16)
```




e1
min_axis
p2_lvl
1
2
3
4
5
100
200
300
400
500


e2
60
50
40
30
20
10
0
min_axis
−100
0
100
200
300
p2_lvl


e3
80
60
40
20
0
min_axis
0
100
200
300
400
p2_lvl



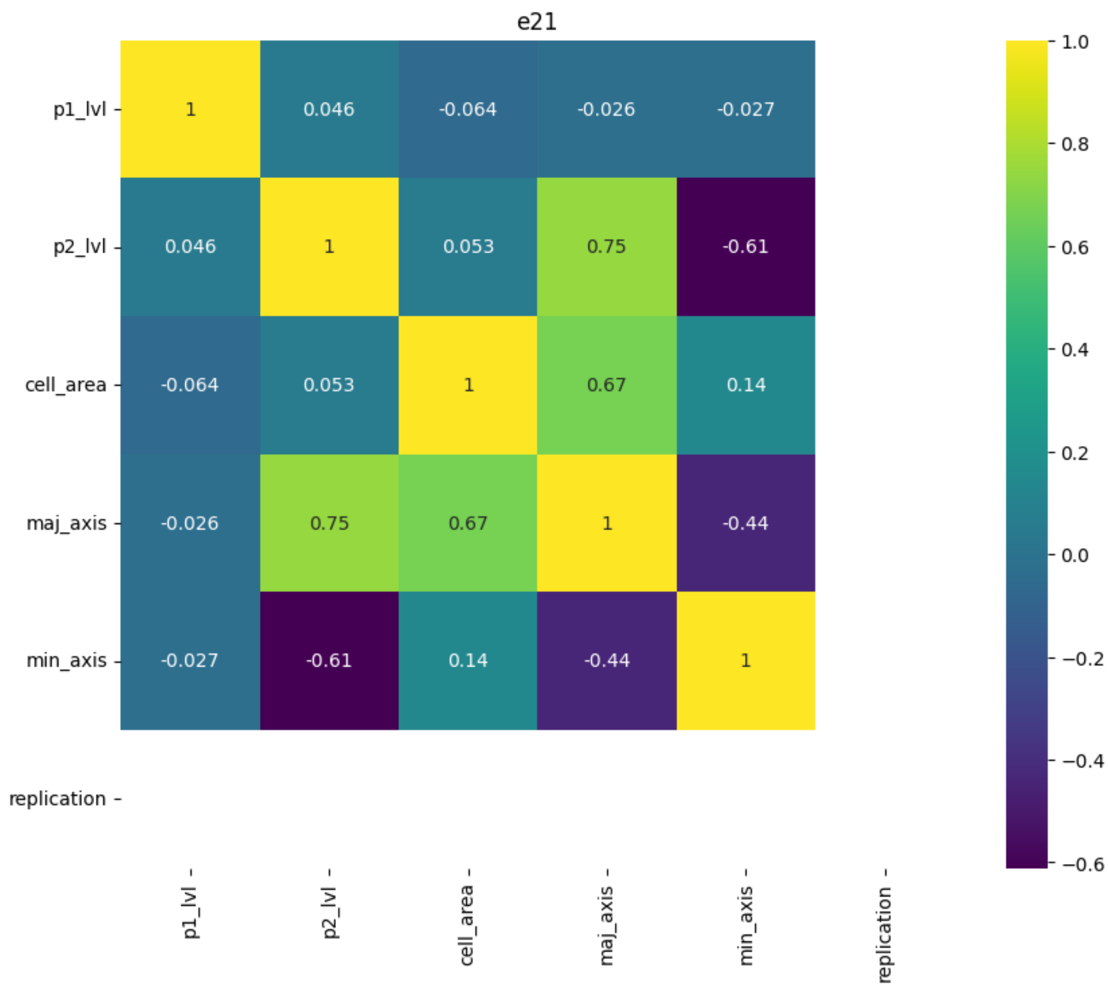
Inference :

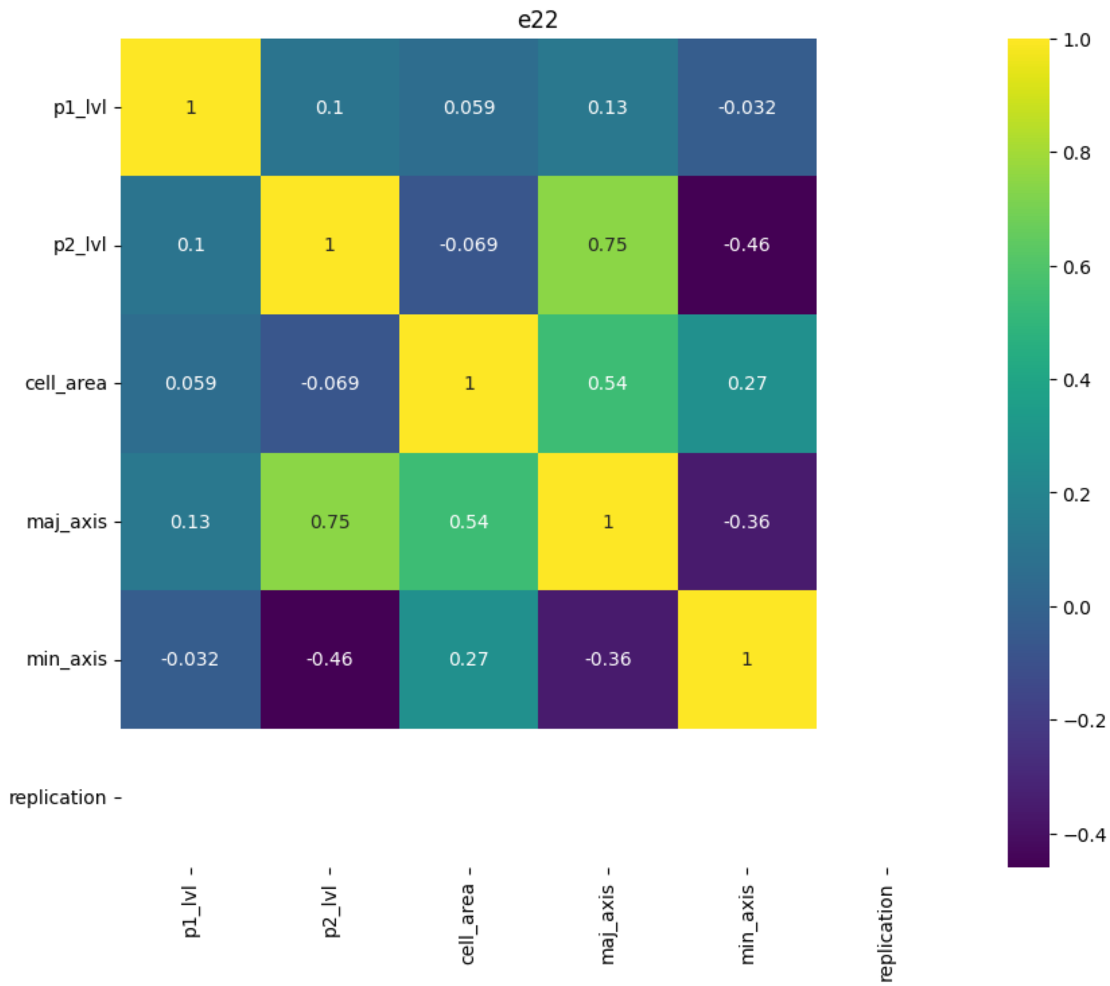
This is just a graphical represntation of correlation value we saw in heatmap. Here we considered the impact of P2 on minor axis under each condition.

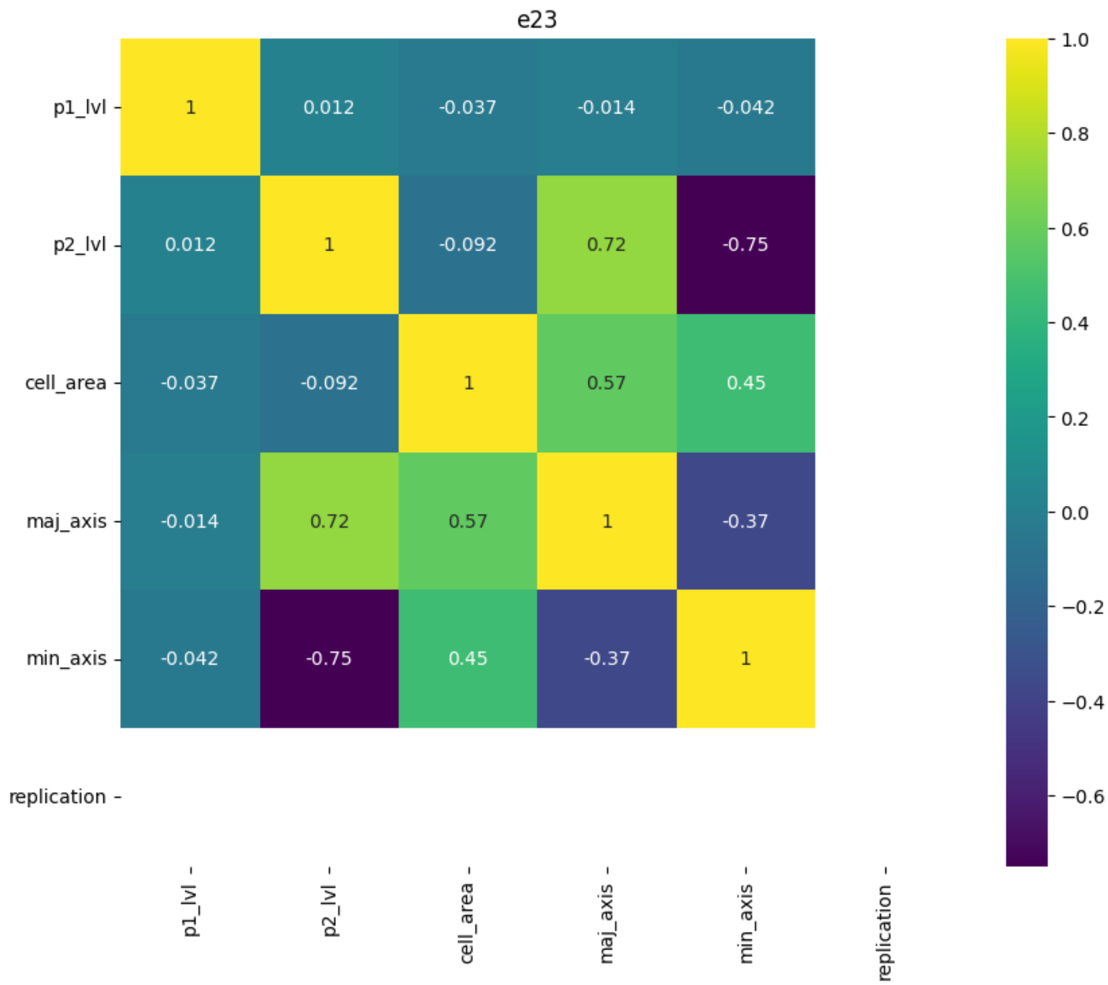
[ ]:
```
## correlation under each replication of e2

# Iterate over the unique experimental conditions
for i in df["replication"].value_counts()[df.replication.unique()].keys():

  # Filter dataFrame based on e2 condition
  df1=df[df.condition=='e2']
```

```python
# Compute the correlation matrix and plot correlaton Heatmap
sns.heatmap(df1[df1.replication==i].corr(), annot=True, cmap = 'viridis')

# Set the figure size and title
fig=plt.gcf()
fig.set_size_inches(10,8)
plt.title("e2"+str(i))

plt.show() # Display plot
```

```
<ipython-input-44-40905197934d>:10: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.replication==i].corr(), annot=True, cmap = 'viridis')
```




e22
p1_lvl
p2_lvl
cell_area
maj_axis
min_axis
replication
1
0.1
0.059
0.13
-0.032
0.1
1
-0.069
0.75
-0.46
0.059
-0.069
1
0.54
0.27
0.13
0.75
0.54
1
-0.36
-0.032
-0.46
0.27
-0.36
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4



Inference :

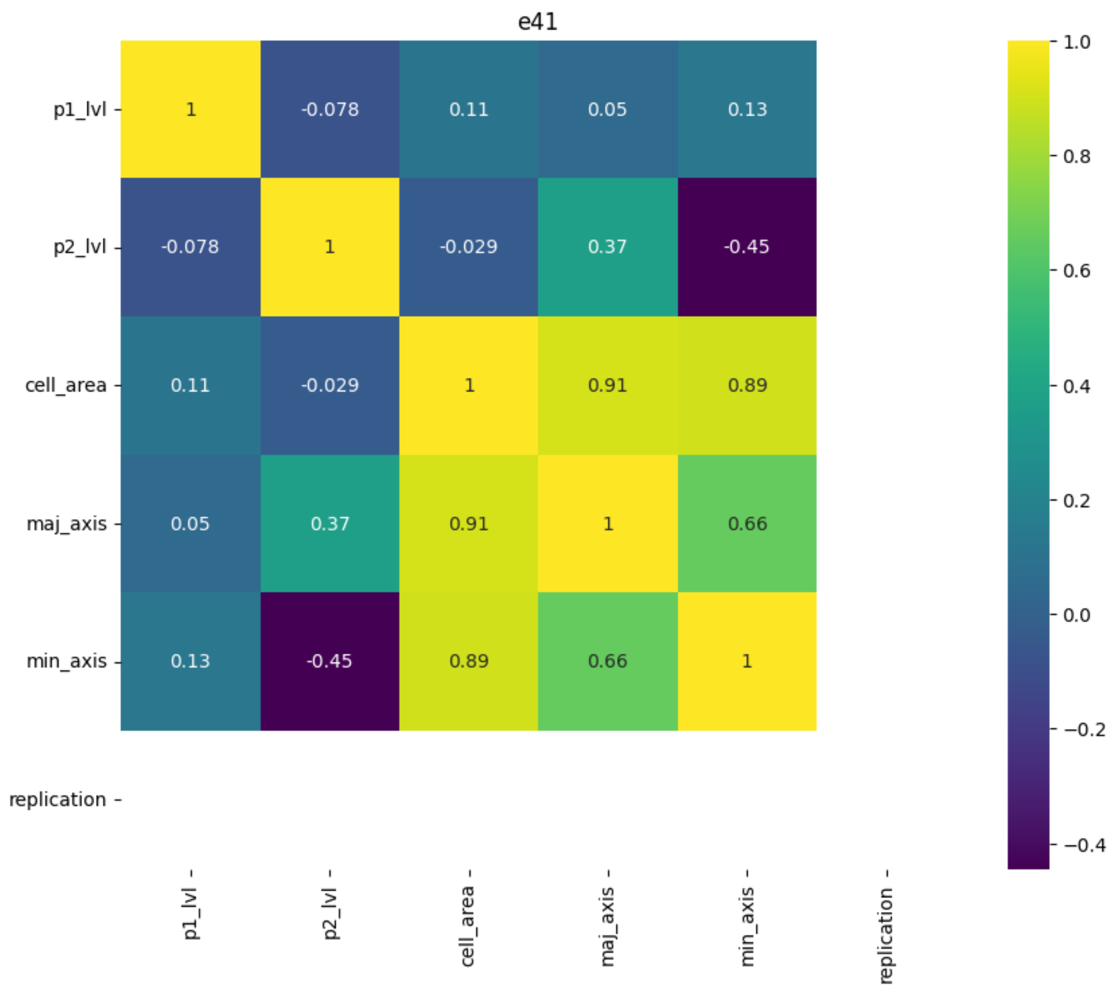
Here we're trying to see how the data is behaving with each repetition under e2 condition.In this observation we'll be considering only the impact of P2 on shape as thoase are the only ones having good outcome.

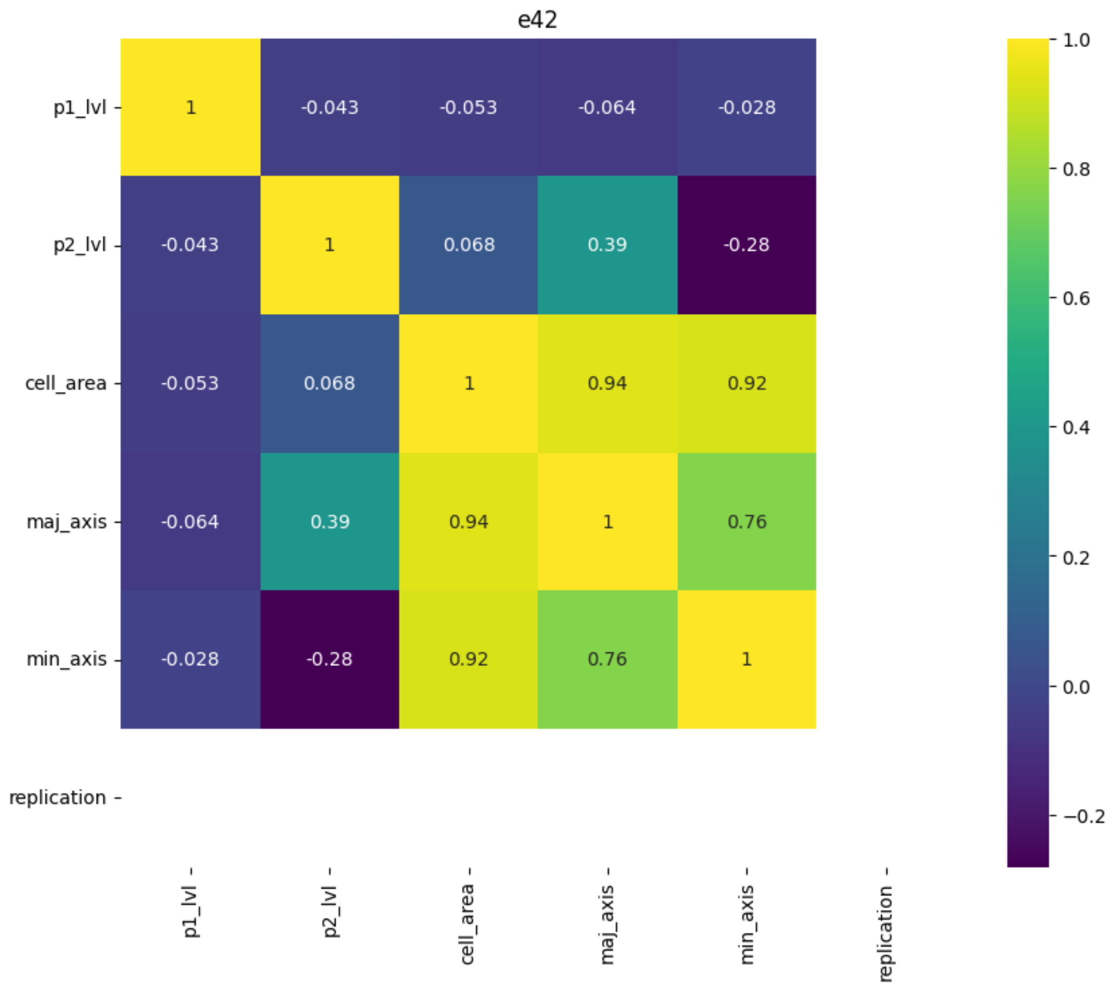
If we observe under each repitition, the impact of P2 on major axis stayed same. But the impact of P2 on the minor axis has been fluctuating.

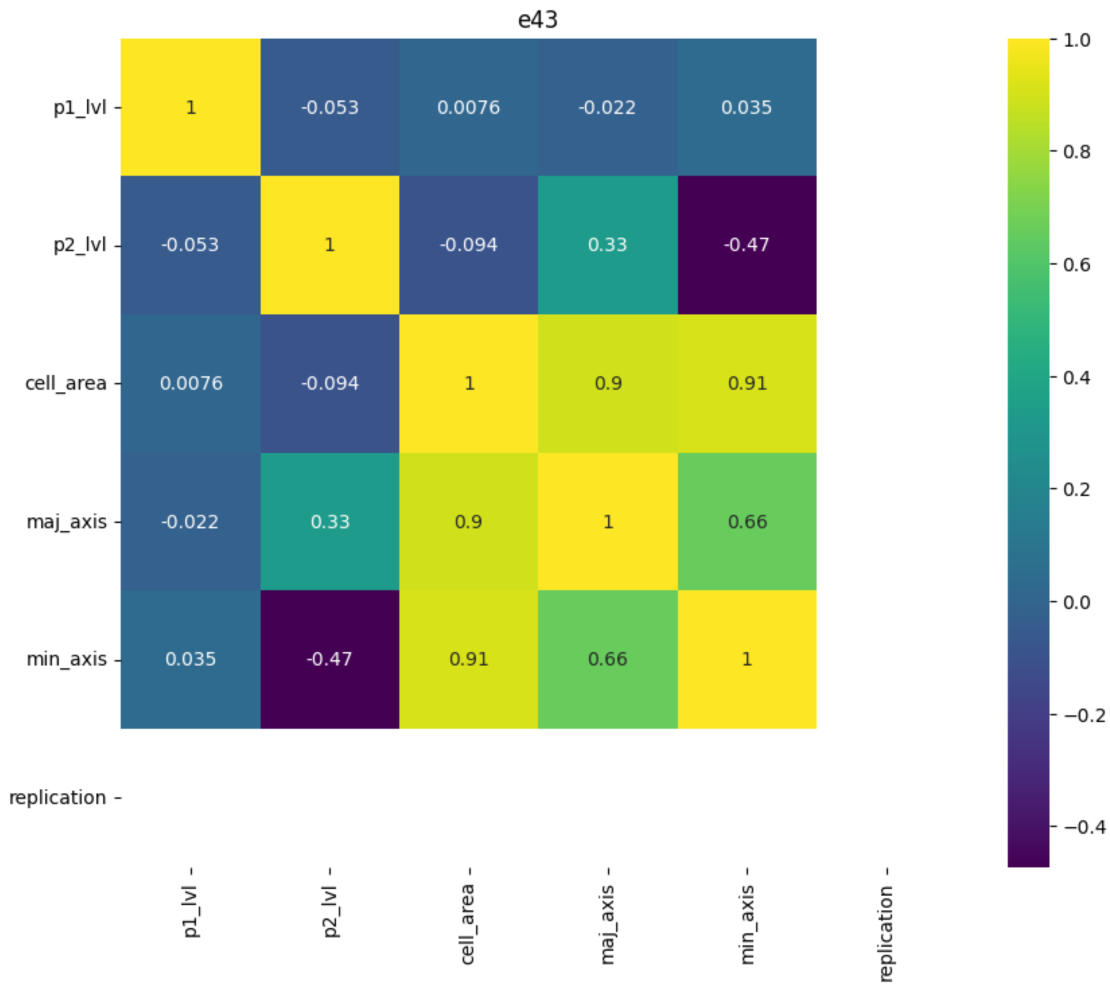
[ ]:
```
## correlation under each replications of e4 ( as we have similar results with
↪e0,e1,e4 we've considered one of it)

# Iterate over the unique replications
for i in df["replication"].value_counts()[df.replication.unique()].keys():

  # Filter dataFrame based on e4 condition
  df1=df[df.condition=='e4']
```

```
# Compute the correlation matrix and plot correlaton Heatmap
sns.heatmap(df1[df1.replication==i].corr(), annot=True, cmap = 'viridis')

# Set the figure size and title
fig=plt.gcf()
fig.set_size_inches(10,8)
plt.title("e4"+str(i))

plt.show() # Display plot
```

```
<ipython-input-45-2717166e76e7>:10: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df1[df1.replication==i].corr(), annot=True, cmap = 'viridis')
```




e42
p1_lvl
p2_lvl
cell_area
maj_axis
min_axis
replication
1
-0.043
-0.053
-0.064
-0.028
-0.043
1
0.068
0.39
-0.28
-0.053
0.068
1
0.94
0.92
-0.064
0.39
0.94
1
0.76
-0.028
-0.28
0.92
0.76
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2



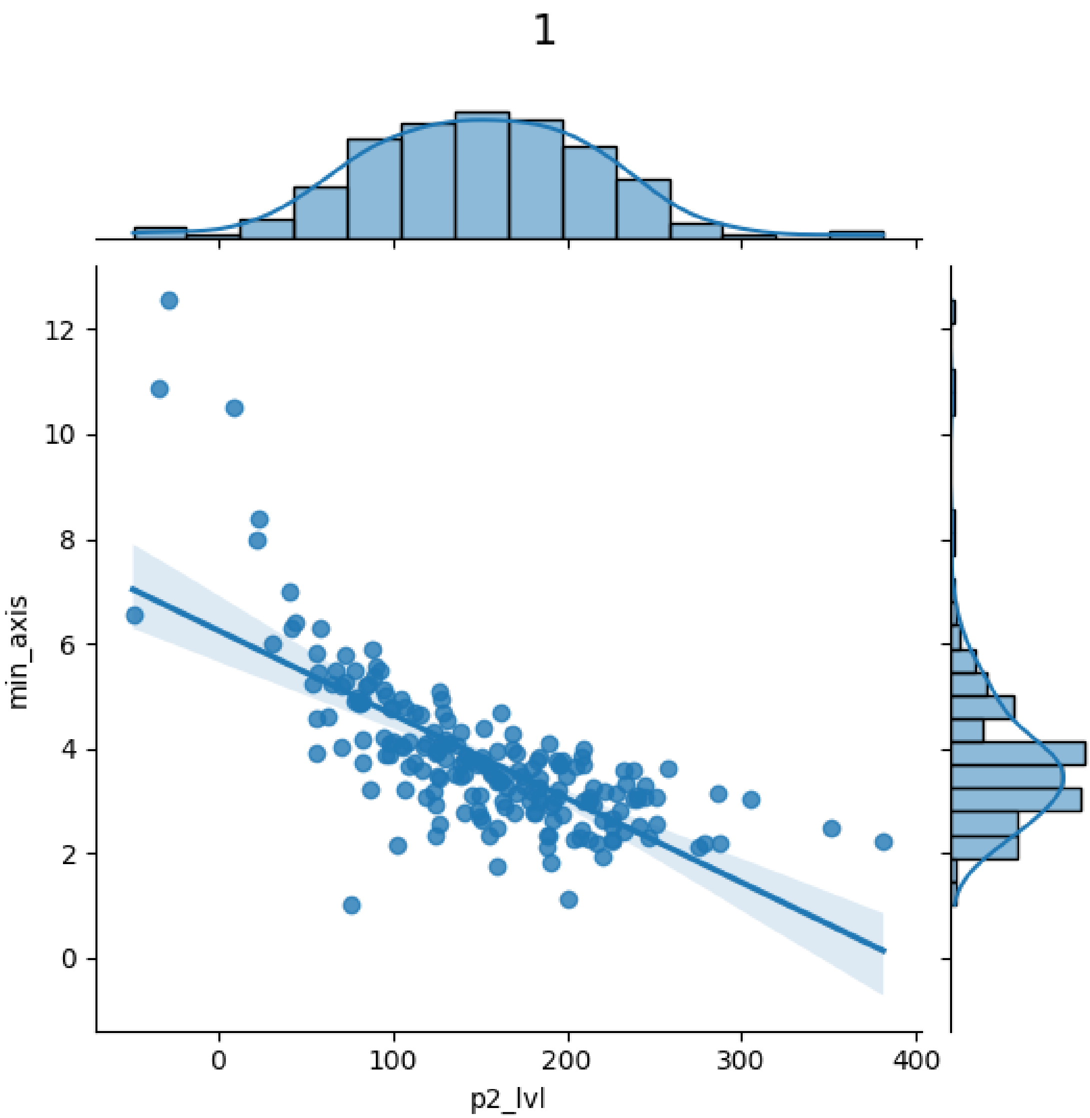
Inference :

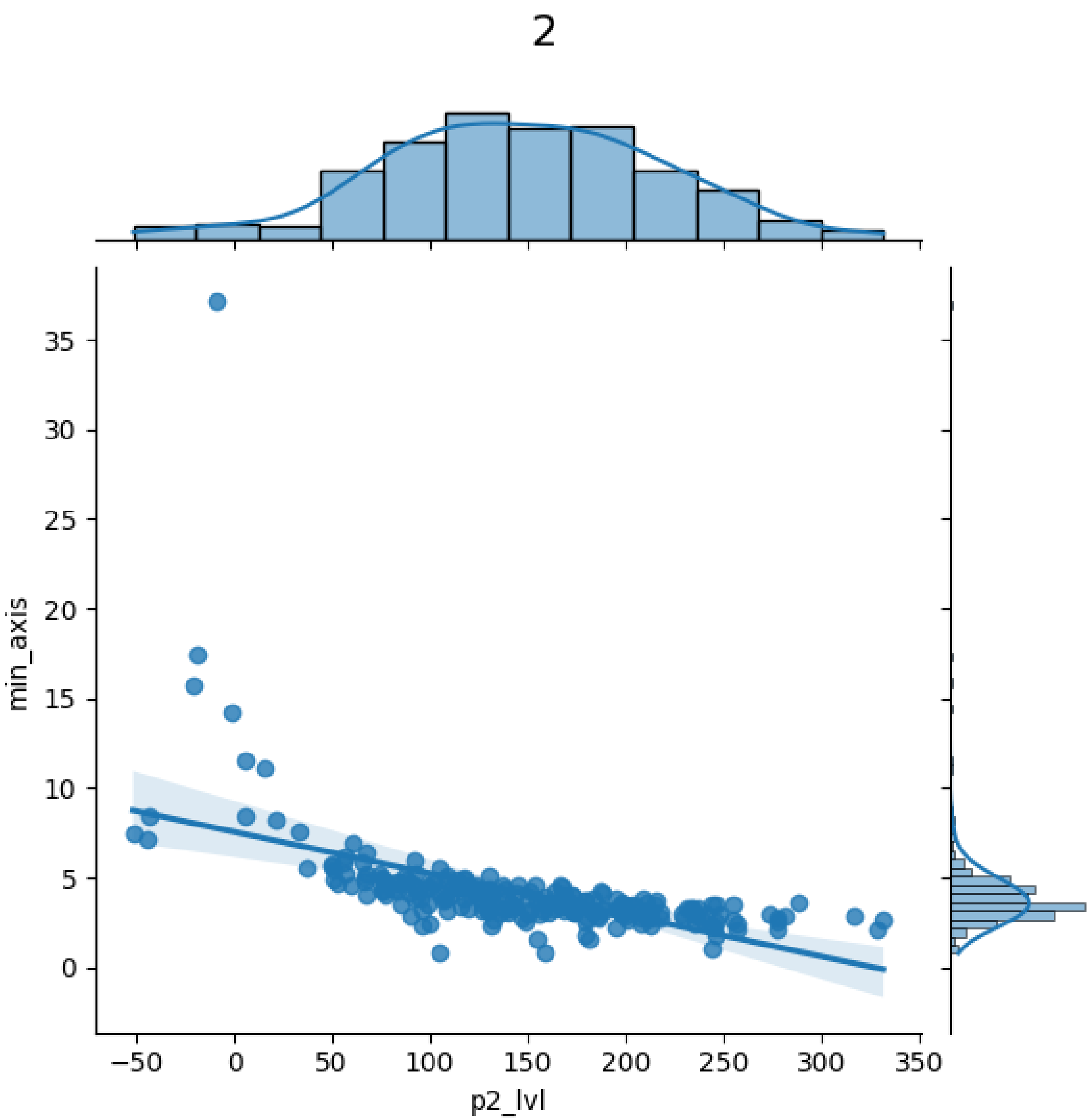
Here we're trying to see the behaviour of the data with each repetition under e4 condition.In this observation we'll be considering only the impact of P2 on shape as those are the only ones having good outcome.

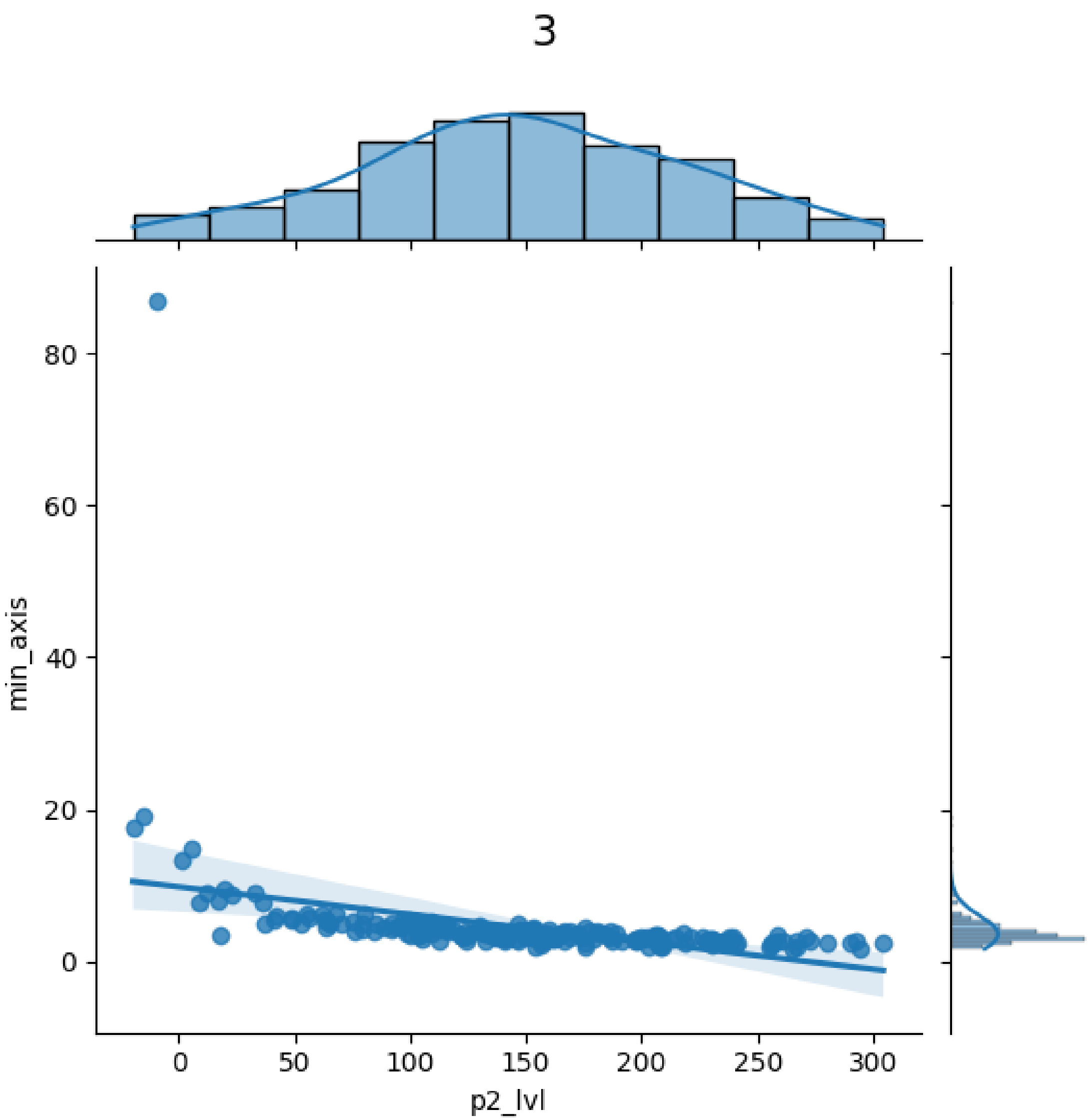
Results are pretty much similar to what we inferred in the above scenario on e2.

[ ]:
```
## p2 impact on minor axis in condition under each repetition of e3 ( as we have␣
↪similar results with e2,e3 we've considered one of it)

# Iterate over the unique replications
for i in df["replication"].value_counts()[df.replication.unique()].keys():

  # Filter dataFrame based on e3 condition
  df1=df[df.condition=='e3']

  # Create a joint plot for the current condition
```

```
sns.jointplot(x='p2_lvl',y='min_axis',data=df1[df1.replication==i],kind='reg')

# Adjust the subplots and add centred title
plt.subplots_adjust(top=0.9)
plt.suptitle(i, fontsize = 16)
```




2
35
30
25
20
15
10
5
0
min_axis
−50
0
50
100
150
200
250
300
350
p2_lvl



Inference :

As we saw fluctuations in the imact of P2 on minor axis in each condition, we decided to analyse on the reason for fluctuations. For that, we used jointplot between these 2 under each repitition.

As per the graphs we got, we found outliers as the key reason for these fluctuations.

6 Statistical Tests

1. **Conducting a linear model analysis for each of the three datasets under all conditions**

linear regression(ols) applied to e0 contains e01,e02,e03

linear regression(ols) applied to e1 contains e11,e12,e13 linear regression(ols) applied to e1 contains e21,e22,e23

linear regression(ols) applied to e1 contains e31,e32,e33

linear regression(ols) applied to e1 contains e41,e42,e43

i) It allows the model to the characteristics and patterns specific to each dataset.So, it can captures unique relationships and variations that may be present in individual datasets.

ii) Test-specific variations may be captured, and you can examine whether the relationships differ across tests.

iii) Smaller sample sizes within each dataset may result in reduced statistical power. It results that the limited ability to detect significant effects or associations due to limited data.

iv) Conducting multiple hypothesis tests (one for each dataset) without correction may increase the risk of making a Type I error.

v) If multiple models need to be interpreted, it can increase the complexity of the overall analysis. Difficulty in providing a unified interpretation when dealing with multiple models.

2. **Combining all datasets into a single model and conducting the linear model analysis.**

Assumptions

a. Assumes that the relationships and characteristics across datasets are homogenous

b. The residuals (the differences between observed and predicted values) are normally distributed.

Advantages

i) Combining datasets leads to a larger sample size, enhancing statistical power and improving the precision of parameter estimates. Greater ability to detect significant effects and relationships.

ii) Simplifies the analysis by treating all datasets as a single entity, streamlining model development and interpretation. Easier to manage and communicate results.

iii) Estimates of coefficients and other parameters are based on a larger dataset, potentially leading to more reliable and stable results.

Draw backs

iv) The combined model may become complex, especially if datasets have different structures or patterns. Challenges in interpreting a complex model and potential overfitting.

v) Outliers in one dataset may disproportionately influence the combined model. Sensitivity to extreme values, potentially affecting the robustness of the analysis.

[ ]:
```
df.describe() # Returns description of the data in the DataFrame
```

[ ]:
```
             p1_lvl       p2_lvl    cell_area     maj_axis     min_axis  \
count   2958.000000  2958.000000  2958.000000  2958.000000  2958.000000
mean     311.740225   269.658793    59.790264     7.012926     3.184773
std      123.479216   145.441257    20.217504     2.331285     2.535062
```

```
min     -40.024410   -85.400030    1.391440     0.489870     0.560590
25%     236.763385   160.295973   46.033330     5.380665     2.276547
50%     300.931940   245.970225   59.601295     6.951340     2.816520
75%     368.821860   352.039562   73.545325     8.559145     3.548015
max     649.025540   699.673480  125.802950    15.110750    86.944080

       replication
count  2958.000000
mean      1.993915
std       0.810376
min       1.000000
25%       1.000000
50%       2.000000
75%       3.000000
max       3.000000
```

[ ]:
```
def ols(x,y):
  """ Performs an ordinary least squares (OLS) regression on two variables.

  Args:
    x: Independent variable.
    y: Dependent variable.

  """

  # Adds a constant term to independent variable
  X2 = sm.add_constant(x)

  # Creates OLS model
  est = sm.OLS(y, X2)

  # Fits OLS model
  est2 = est.fit()

  print("R-Squared : "+ str(est.fit().rsquared))
  print("\n")
  print(est2.summary())
```

[ ]:
```
## On cell_area

X=df[['p1_lvl','p2_lvl']] # Independent variables
y=df['cell_area'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.0005716013235134154
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              cell_area   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.8450
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.430
Time:                        11:57:07   Log-Likelihood:                -13089.
No. Observations:                2958   AIC:                         2.618e+04
Df Residuals:                    2955   BIC:                         2.620e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         61.5766      1.453     42.375      0.000      58.727      64.426
p1_lvl        -0.0039      0.003     -1.241      0.215      -0.010       0.002
p2_lvl        -0.0021      0.003     -0.765      0.444      -0.007       0.003
==============================================================================
Omnibus:                        4.159   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.125   Jarque-Bera (JB):                3.675
Skew:                          -0.004   Prob(JB):                        0.159
Kurtosis:                       2.827   Cond. No.                     1.67e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.67e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[ ]:
```python
## on cell area with each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p1_lvl','p2_lvl']] # Independent variables
  y=df1['cell_area'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 0.00440136126296764


                            OLS Regression Results
==============================================================================
Dep. Variable:              cell_area   R-squared:                       0.004
```

```
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.344
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.262
Time:                        11:57:07   Log-Likelihood:                -2707.3
No. Observations:                 611   AIC:                             5421.
Df Residuals:                     608   BIC:                             5434.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         68.8900      5.925     11.628      0.000      57.255      80.525
p1_lvl        -0.0118      0.016     -0.720      0.472      -0.044       0.020
p2_lvl        -0.0176      0.012     -1.461      0.145      -0.041       0.006
==============================================================================
Omnibus:                        0.594   Durbin-Watson:                   2.093
Prob(Omnibus):                  0.743   Jarque-Bera (JB):                0.691
Skew:                           0.024   Prob(JB):                        0.708
Kurtosis:                       2.843   Cond. No.                     2.96e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.96e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e1

R-Squared : 0.0050087454605418635


                            OLS Regression Results
==============================================================================
Dep. Variable:              cell_area   R-squared:                       0.005
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     1.467
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.231
Time:                        11:57:07   Log-Likelihood:                -2576.1
No. Observations:                 586   AIC:                             5158.
Df Residuals:                     583   BIC:                             5171.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         57.3082      4.296     13.340      0.000      48.871      65.746
```

```
p1_lvl         0.0270      0.016      1.647      0.100      -0.005       0.059
p2_lvl        -0.0052      0.011     -0.457      0.648      -0.028       0.017
==============================================================================
Omnibus:                        0.752   Durbin-Watson:                   1.899
Prob(Omnibus):                  0.687   Jarque-Bera (JB):                0.835
Skew:                          -0.029   Prob(JB):                        0.659
Kurtosis:                       2.824   Cond. No.                     1.79e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.79e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
condition = e2

R-Squared : 0.0013884102004716459
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              cell_area   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.4102
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.664
Time:                        11:57:07   Log-Likelihood:                -2635.4
No. Observations:                 593   AIC:                             5277.
Df Residuals:                     590   BIC:                             5290.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         63.5399      5.171     12.288      0.000      53.385      73.695
p1_lvl        -0.0062      0.016     -0.378      0.706      -0.038       0.026
p2_lvl        -0.0099      0.012     -0.803      0.422      -0.034       0.014
==============================================================================
Omnibus:                        0.753   Durbin-Watson:                   2.028
Prob(Omnibus):                  0.686   Jarque-Bera (JB):                0.851
Skew:                          -0.053   Prob(JB):                        0.653
Kurtosis:                       2.848   Cond. No.                     2.06e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[2] The condition number is large, 2.06e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e3

R-Squared : 0.0006569729794903001

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              cell_area   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.003
Method:                 Least Squares   F-statistic:                    0.2012
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.818
Time:                        11:57:07   Log-Likelihood:                -2726.8
No. Observations:                 615   AIC:                             5460.
Df Residuals:                     612   BIC:                             5473.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         62.6593      8.153      7.686      0.000      46.649      78.670
p1_lvl        -0.0065      0.016     -0.406      0.685      -0.038       0.025
p2_lvl        -0.0056      0.012     -0.474      0.636      -0.029       0.018
==============================================================================
Omnibus:                        0.784   Durbin-Watson:                   2.027
Prob(Omnibus):                  0.676   Jarque-Bera (JB):                0.873
Skew:                           0.044   Prob(JB):                        0.646
Kurtosis:                       2.838   Cond. No.                     5.19e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.19e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e4

R-Squared : 0.0004969588608149111

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              cell_area   R-squared:                       0.000
```

```
Model:                            OLS   Adj. R-squared:                 -0.003
Method:                 Least Squares   F-statistic:                    0.1367
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.872
Time:                        11:57:07   Log-Likelihood:                -2438.4
No. Observations:                 553   AIC:                             4883.
Df Residuals:                     550   BIC:                             4896.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         60.7716      8.305      7.318      0.000      44.459      77.085
p1_lvl         0.0052      0.018      0.286      0.775      -0.030       0.041
p2_lvl        -0.0050      0.012     -0.421      0.674      -0.028       0.018
==============================================================================
Omnibus:                        0.705   Durbin-Watson:                   2.014
Prob(Omnibus):                  0.703   Jarque-Bera (JB):                0.786
Skew:                           0.008   Prob(JB):                        0.675
Kurtosis:                       2.816   Cond. No.                     5.76e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.76e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Expression Level of the protein P1

H0: βP1=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP10 (There is significant relationship between expression level P1 and the cell size)

Expression Level of the protein P2

H0: βP2=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP20 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

For the Expression level of P1 & P2 the p-value was greater than 0.05, so null hypotheses cannot be rejected the relationship between expression level of P1 & P2 with 'Cells' is not statistically significant. But, Expression level of P2 the p-value was slightly near to 0.05, so expression level of P2 and 'Cells' has some relation.

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

For the Expression level of P1 & P2 the p-value was greater than 0.05, so null hypotheses cannot be rejected the relationship between expression level of P1 & P2 with 'Cells' is not statistically significant. But, Expression level of P1 the p-value was slightly near to 0.05, so expression level of P1 and 'Cells' has some relation.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

For the Expression level of P1 & P2 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 & P2 with 'Cells' is not statistically significant.

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

For the Expression level of P1 & P2 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 & P2 with 'Cells' is not statistically significant.

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

For the Expression level of P1 & P2 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 & P2 with 'Cells' is not statistically significant.

[ ]:
```python
## p1 & p2 on major_axis

X=df[['p1_lvl','p2_lvl']] # Independent variables
y=df['maj_axis'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.6595729479514083


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.660
Model:                            OLS   Adj. R-squared:                  0.659
Method:                 Least Squares   F-statistic:                     2863.
Date:                Thu, 19 Oct 2023   Prob (F-statistic):               0.00
Time:                        11:57:07   Log-Likelihood:                -5106.7
No. Observations:                2958   AIC:                         1.022e+04
Df Residuals:                    2955   BIC:                         1.024e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.8221      0.098     39.083      0.000       3.630       4.014
p1_lvl        -0.0008      0.000     -3.825      0.000      -0.001      -0.000
p2_lvl         0.0128      0.000     70.362      0.000       0.012       0.013
```

```
==============================================================================
Omnibus:                      253.912   Durbin-Watson:                   1.984
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              407.967
Skew:                          -0.639   Prob(JB):                     2.58e-89
Kurtosis:                       4.296   Cond. No.                     1.67e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.67e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[ ]:
```
## p1 & p2 on major_axis under each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p1_lvl','p2_lvl']] # Independent variables
  y=df1['maj_axis'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 0.2764628411349941


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.276
Model:                            OLS   Adj. R-squared:                  0.274
Method:                 Least Squares   F-statistic:                     116.2
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           1.89e-43
Time:                        11:57:07   Log-Likelihood:                -1044.3
No. Observations:                 611   AIC:                             2095.
Df Residuals:                     608   BIC:                             2108.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.1873      0.390     10.748      0.000       3.422       4.952
p1_lvl        -0.0006      0.001     -0.544      0.586      -0.003       0.002
p2_lvl         0.0120      0.001     15.239      0.000       0.010       0.014
==============================================================================
Omnibus:                       36.219   Durbin-Watson:                   2.066
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               44.161
```

```
Skew:                          -0.546   Prob(JB):                     2.57e-10
Kurtosis:                       3.737   Cond. No.                     2.96e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.96e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e1

R-Squared : 0.2912451095026556

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.291
Model:                            OLS   Adj. R-squared:                  0.289
Method:                 Least Squares   F-statistic:                     119.8
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           2.63e-44
Time:                        11:57:07   Log-Likelihood:                -1006.5
No. Observations:                 586   AIC:                             2019.
Df Residuals:                     583   BIC:                             2032.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.6921      0.295     12.516      0.000       3.113       4.271
p1_lvl         0.0018      0.001      1.613      0.107      -0.000       0.004
p2_lvl         0.0121      0.001     15.408      0.000       0.011       0.014
==============================================================================
Omnibus:                       44.729   Durbin-Watson:                   1.936
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               57.616
Skew:                          -0.629   Prob(JB):                     3.08e-13
Kurtosis:                       3.881   Cond. No.                     1.79e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.79e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e2

R-Squared : 0.5402680626783148

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.539
Method:                 Least Squares   F-statistic:                     346.7
Date:                Thu, 19 Oct 2023   Prob (F-statistic):         2.75e-100
Time:                        11:57:07   Log-Likelihood:                -894.65
No. Observations:                 593   AIC:                             1795.
Df Residuals:                     590   BIC:                             1808.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.9119      0.275     10.605      0.000       2.373       3.451
p1_lvl        -0.0004      0.001     -0.414      0.679      -0.002       0.001
p2_lvl         0.0173      0.001     26.315      0.000       0.016       0.019
==============================================================================
Omnibus:                      139.662   Durbin-Watson:                   2.047
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              493.327
Skew:                          -1.066   Prob(JB):                    7.51e-108
Kurtosis:                       6.927   Cond. No.                     2.06e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e3

R-Squared : 0.5675976872546724

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.568
Model:                            OLS   Adj. R-squared:                  0.566
Method:                 Least Squares   F-statistic:                     401.7
Date:                Thu, 19 Oct 2023   Prob (F-statistic):         3.82e-112
Time:                        11:57:07   Log-Likelihood:                -910.87
No. Observations:                 615   AIC:                             1828.
```

```
Df Residuals:                     612   BIC:                             1841.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.7828      0.426      6.539      0.000       1.947       3.619
p1_lvl        -0.0002      0.001     -0.228      0.819      -0.002       0.001
p2_lvl         0.0176      0.001     28.335      0.000       0.016       0.019
==============================================================================
Omnibus:                       47.756   Durbin-Watson:                   2.022
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               67.924
Skew:                          -0.595   Prob(JB):                     1.78e-15
Kurtosis:                       4.111   Cond. No.                     5.19e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.19e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e4

R-Squared : 0.12994698934857396


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.130
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     41.07
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           2.37e-17
Time:                        11:57:07   Log-Likelihood:                -1090.1
No. Observations:                 553   AIC:                             2186.
Df Residuals:                     550   BIC:                             2199.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.0797      0.725      7.005      0.000       3.655       6.504
p1_lvl         0.0001      0.002      0.071      0.943      -0.003       0.003
p2_lvl         0.0093      0.001      9.052      0.000       0.007       0.011
==============================================================================
Omnibus:                       32.699   Durbin-Watson:                   2.016
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               37.960
```

```
Skew:                          -0.570   Prob(JB):                     5.72e-09
Kurtosis:                       3.592   Cond. No.                     5.76e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.76e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Expression Level of the protein P1

H0: βP1=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP10 (There is significant relationship between expression level P1 and the cell size)

Expression Level of the protein P2

H0: βP2=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP20 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the p-value was near to 0.05 expression level of P1 & 'major axis' may have slight statistical relation. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

[ ]:
```
## p1 on major_axis

X=df[['p1_lvl']] # Independent variable
y=df['maj_axis'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.08921328749749202


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.089
Model:                            OLS   Adj. R-squared:                  0.089
Method:                 Least Squares   F-statistic:                     289.5
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           5.10e-62
Time:                        11:57:07   Log-Likelihood:                -6562.2
No. Observations:                2958   AIC:                         1.313e+04
Df Residuals:                    2956   BIC:                         1.314e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          8.7709      0.111     78.933      0.000       8.553       8.989
p1_lvl        -0.0056      0.000    -17.016      0.000      -0.006      -0.005
==============================================================================
Omnibus:                       31.424   Durbin-Watson:                   1.156
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               32.146
Skew:                           0.251   Prob(JB):                     1.05e-07
Kurtosis:                       3.093   Cond. No.                         911.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[ ]:
```python
## p1 on major_axis under each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p1_lvl']] # Independent variables
  y=df1['maj_axis'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 9.523890659746126e-05


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                   0.05801
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.810
Time:                        11:57:07   Log-Likelihood:                -1143.1
No. Observations:                 611   AIC:                             2290.
Df Residuals:                     609   BIC:                             2299.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          7.4261      0.384     19.363      0.000       6.673       8.179
p1_lvl        -0.0003      0.001     -0.241      0.810      -0.003       0.002
==============================================================================
Omnibus:                        3.134   Durbin-Watson:                   2.033
Prob(Omnibus):                  0.209   Jarque-Bera (JB):                3.019
Skew:                          -0.171   Prob(JB):                        0.221
Kurtosis:                       3.045   Cond. No.                     1.83e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.83e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e1

R-Squared : 0.002642302246965267
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.547
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.214
Time:                        11:57:07   Log-Likelihood:                -1106.6
No. Observations:                 586   AIC:                             2217.
Df Residuals:                     584   BIC:                             2226.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          7.3286      0.210     34.947      0.000       6.917       7.740
p1_lvl         0.0017      0.001      1.244      0.214      -0.001       0.004
==============================================================================
Omnibus:                        3.238   Durbin-Watson:                   1.955
Prob(Omnibus):                  0.198   Jarque-Bera (JB):                3.214
Skew:                          -0.111   Prob(JB):                        0.201
Kurtosis:                       3.286   Cond. No.                         499.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


condition = e2

R-Squared : 0.00067803146174705290


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.4010
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.527
Time:                        11:57:07   Log-Likelihood:                -1124.9
No. Observations:                 593   AIC:                             2254.
Df Residuals:                     591   BIC:                             2262.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
const          5.0862      0.386     13.186      0.000       4.329       5.844
p1_lvl         0.0008      0.001      0.633      0.527      -0.002       0.003
==============================================================================
Omnibus:                        4.687   Durbin-Watson:                   2.067
Prob(Omnibus):                  0.096   Jarque-Bera (JB):                4.658
Skew:                          -0.217   Prob(JB):                       0.0974
Kurtosis:                       2.999   Cond. No.                     1.76e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.76e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e3

R-Squared : 0.0003486172354670858


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.2138
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.644
Time:                        11:57:07   Log-Likelihood:                -1168.6
No. Observations:                 615   AIC:                             2341.
Df Residuals:                     613   BIC:                             2350.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.0015      0.635      7.870      0.000       3.754       6.249
p1_lvl         0.0006      0.001      0.462      0.644      -0.002       0.003
==============================================================================
Omnibus:                        3.055   Durbin-Watson:                   1.969
Prob(Omnibus):                  0.217   Jarque-Bera (JB):                2.869
Skew:                          -0.157   Prob(JB):                        0.238
Kurtosis:                       3.113   Cond. No.                     4.89e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
[2] The condition number is large, 4.89e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
condition = e4

R-Squared : 0.00032497132317454499
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.1791
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.672
Time:                        11:57:07   Log-Likelihood:                -1128.5
No. Observations:                 553   AIC:                             2261.
Df Residuals:                     551   BIC:                             2270.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          9.9954      0.515     19.425      0.000       8.985      11.006
p1_lvl        -0.0007      0.002     -0.423      0.672      -0.004       0.003
==============================================================================
Omnibus:                       14.533   Durbin-Watson:                   2.063
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               15.327
Skew:                          -0.356   Prob(JB):                     0.000470
Kurtosis:                       3.397   Cond. No.                     1.98e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.98e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Expression Level of the protein P1

H0: βP1=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP10 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the p-value was near to 0.05 expression level of P1 & 'major axis' may have slight statistical relation.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant.

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant.

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'major axis' is not statistically significant.

[ ]:
```
## p2 on major_axis

X=df[['p2_lvl']] # Independent variable
y=df['maj_axis'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.6578878775216196


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.658
Model:                            OLS   Adj. R-squared:                  0.658
Method:                 Least Squares   F-statistic:                     5684.
Date:                Thu, 19 Oct 2023   Prob (F-statistic):               0.00
Time:                        11:57:07   Log-Likelihood:                -5114.0
No. Observations:                2958   AIC:                         1.023e+04
Df Residuals:                    2956   BIC:                         1.024e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
```

```
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.5070      0.053     66.383      0.000       3.403       3.611
p2_lvl         0.0130      0.000     75.395      0.000       0.013       0.013
==============================================================================
Omnibus:                      253.081   Durbin-Watson:                   1.970
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              406.230
Skew:                          -0.637   Prob(JB):                     6.14e-89
Kurtosis:                       4.293   Cond. No.                         645.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[ ]:
```python
## p2 on major_axis under each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p2_lvl']] # Independent variables
  y=df1['maj_axis'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 0.27611023370510224


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.276
Model:                            OLS   Adj. R-squared:                  0.275
Method:                 Least Squares   F-statistic:                     232.3
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           1.14e-44
Time:                        11:57:07   Log-Likelihood:                -1044.4
No. Observations:                 611   AIC:                             2093.
Df Residuals:                     609   BIC:                             2102.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.0140      0.225     17.877      0.000       3.573       4.455
p2_lvl         0.0120      0.001     15.241      0.000       0.010       0.014
==============================================================================
Omnibus:                       35.987   Durbin-Watson:                   2.066
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.699
Skew:                          -0.545   Prob(JB):                     3.24e-10
Kurtosis:                       3.726   Cond. No.                     1.18e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.18e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e1

R-Squared : 0.28808256923892095

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.288
Model:                            OLS   Adj. R-squared:                  0.287
Method:                 Least Squares   F-statistic:                     236.3
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           4.97e-45
Time:                        11:57:07   Log-Likelihood:                -1007.8
No. Observations:                 586   AIC:                             2020.
Df Residuals:                     584   BIC:                             2028.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.9663      0.241     16.431      0.000       3.492       4.440
p2_lvl         0.0121      0.001     15.373      0.000       0.011       0.014
==============================================================================
Omnibus:                       44.131   Durbin-Watson:                   1.926
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               55.552
Skew:                          -0.635   Prob(JB):                     8.65e-13
Kurtosis:                       3.813   Cond. No.                     1.32e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.32e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e2

R-Squared : 0.5401347970445641

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.539
Method:                 Least Squares   F-statistic:                     694.2
Date:                Thu, 19 Oct 2023   Prob (F-statistic):          9.05e-102
Time:                        11:57:07   Log-Likelihood:                -894.74
No. Observations:                 593   AIC:                             1793.
Df Residuals:                     591   BIC:                             1802.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.8071      0.106     26.559      0.000       2.600       3.015
p2_lvl         0.0173      0.001     26.347      0.000       0.016       0.019
==============================================================================
Omnibus:                      139.440   Durbin-Watson:                   2.045
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              492.048
Skew:                          -1.064   Prob(JB):                    1.42e-107
Kurtosis:                       6.922   Cond. No.                         379.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

condition = e3

R-Squared : 0.5675608492675719

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.568
Model:                            OLS   Adj. R-squared:                  0.567
Method:                 Least Squares   F-statistic:                     804.5
Date:                Thu, 19 Oct 2023   Prob (F-statistic):          1.10e-113
Time:                        11:57:07   Log-Likelihood:                -910.89
No. Observations:                 615   AIC:                             1826.
Df Residuals:                     613   BIC:                             1835.
Df Model:                           1
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.6884      0.101     26.510      0.000       2.489       2.888
p2_lvl         0.0176      0.001     28.364      0.000       0.016       0.019
==============================================================================
Omnibus:                       47.619   Durbin-Watson:                   2.021
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               67.681
Skew:                          -0.594   Prob(JB):                     2.01e-15
Kurtosis:                       4.109   Cond. No.                         386.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


condition = e4

R-Squared : 0.1299389384279791


                            OLS Regression Results
==============================================================================
Dep. Variable:               maj_axis   R-squared:                       0.130
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     82.29
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           2.07e-18
Time:                        11:57:07   Log-Likelihood:                -1090.1
No. Observations:                 553   AIC:                             2184.
Df Residuals:                     551   BIC:                             2193.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.1157      0.520      9.847      0.000       4.095       6.136
p2_lvl         0.0093      0.001      9.071      0.000       0.007       0.011
==============================================================================
Omnibus:                       32.689   Durbin-Watson:                   2.015
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               37.950
Skew:                          -0.569   Prob(JB):                     5.74e-09
Kurtosis:                       3.592   Cond. No.                     3.55e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[2] The condition number is large, 3.55e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Expression Level of the protein P2

H0: βP2=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP20 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant.

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'major axis' is statistically significant

[ ]:
```
## p1 & p2 on minor_axis

X=df[['p1_lvl','p2_lvl']] # Independent variables
y=df['min_axis'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.19996975090308888


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.200
Model:                            OLS   Adj. R-squared:                  0.199
Method:                 Least Squares   F-statistic:                     369.3
```

```
Date:                Thu, 19 Oct 2023   Prob (F-statistic):          6.91e-144
Time:                        11:57:07   Log-Likelihood:                -6618.3
No. Observations:                2958   AIC:                         1.324e+04
Df Residuals:                    2955   BIC:                         1.326e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.0193      0.163     30.790      0.000       4.700       5.339
p1_lvl         0.0007      0.000      1.909      0.056   -1.86e-05       0.001
p2_lvl        -0.0076      0.000    -25.071      0.000      -0.008      -0.007
==============================================================================
Omnibus:                     6828.684   Durbin-Watson:                   1.956
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         57333198.996
Skew:                          21.977   Prob(JB):                         0.00
Kurtosis:                     683.622   Cond. No.                     1.67e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.67e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[ ]:
```
## p1 & p2 on minor_axis under each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p1_lvl','p2_lvl']] # Independent variables
  y=df1['min_axis'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 0.39690527561649114


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.397
Model:                            OLS   Adj. R-squared:                  0.395
Method:                 Least Squares   F-statistic:                     200.1
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           1.73e-67
Time:                        11:57:07   Log-Likelihood:                -442.36
No. Observations:                 611   AIC:                             890.7
```

```
Df Residuals:                     608   BIC:                             904.0
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.4554      0.145     30.628      0.000       4.170       4.741
p1_lvl        -0.0002      0.000     -0.548      0.584      -0.001       0.001
p2_lvl        -0.0059      0.000    -19.983      0.000      -0.006      -0.005
==============================================================================
Omnibus:                       26.568   Durbin-Watson:                   2.165
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.818
Skew:                          -0.349   Prob(JB):                     1.37e-09
Kurtosis:                       4.056   Cond. No.                     2.96e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.96e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e1

R-Squared : 0.34544291593515397


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.345
Model:                            OLS   Adj. R-squared:                  0.343
Method:                 Least Squares   F-statistic:                     153.8
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           2.23e-54
Time:                        11:57:07   Log-Likelihood:                -375.15
No. Observations:                 586   AIC:                             756.3
Df Residuals:                     583   BIC:                             769.4
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.9546      0.100     39.376      0.000       3.757       4.152
p1_lvl         0.0005      0.000      1.398      0.163      -0.000       0.001
p2_lvl        -0.0047      0.000    -17.472      0.000      -0.005      -0.004
==============================================================================
Omnibus:                       48.153   Durbin-Watson:                   1.848
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               76.871
```

```
Skew:                          -0.575   Prob(JB):                     2.03e-17
Kurtosis:                       4.351   Cond. No.                     1.79e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.79e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e2

R-Squared : 0.28127340793531685

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.279
Method:                 Least Squares   F-statistic:                     115.4
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           4.86e-43
Time:                        11:57:07   Log-Likelihood:                -1430.2
No. Observations:                 593   AIC:                             2866.
Df Residuals:                     590   BIC:                             2880.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          7.8167      0.677     11.538      0.000       6.486       9.147
p1_lvl     -5.538e-05      0.002     -0.026      0.979      -0.004       0.004
p2_lvl        -0.0246      0.002    -15.174      0.000      -0.028      -0.021
==============================================================================
Omnibus:                     1083.432   Durbin-Watson:                   2.042
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           988361.324
Skew:                          11.890   Prob(JB):                         0.00
Kurtosis:                     201.584   Cond. No.                     2.06e+03
==============================================================================
```

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are strong multicollinearity or other numerical problems.

condition = e3

```
R-Squared : 0.18441431541156605


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.184
Model:                            OLS   Adj. R-squared:                  0.182
Method:                 Least Squares   F-statistic:                     69.19
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           8.12e-28
Time:                        11:57:07   Log-Likelihood:                -1674.8
No. Observations:                 615   AIC:                             3356.
Df Residuals:                     612   BIC:                             3369.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          9.0205      1.474      6.121      0.000       6.127      11.914
p1_lvl        -0.0022      0.003     -0.781      0.435      -0.008       0.003
p2_lvl        -0.0251      0.002    -11.706      0.000      -0.029      -0.021
==============================================================================
Omnibus:                     1319.865   Durbin-Watson:                   1.962
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          3022799.432
Skew:                          16.752   Prob(JB):                         0.00
Kurtosis:                     344.819   Cond. No.                     5.19e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.19e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e4

R-Squared : 0.16366513500009283


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.164
Model:                            OLS   Adj. R-squared:                  0.161
Method:                 Least Squares   F-statistic:                     53.82
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           4.51e-22
Time:                        11:57:07   Log-Likelihood:                -197.67
No. Observations:                 553   AIC:                             401.3
```

```
Df Residuals:                     550   BIC:                             414.3
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.0586      0.144     21.181      0.000       2.775       3.342
p1_lvl         0.0001      0.000      0.388      0.698      -0.000       0.001
p2_lvl        -0.0021      0.000    -10.327      0.000      -0.003      -0.002
==============================================================================
Omnibus:                       34.977   Durbin-Watson:                   2.023
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               41.968
Skew:                          -0.577   Prob(JB):                     7.71e-10
Kurtosis:                       3.698   Cond. No.                     5.76e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.76e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Expression Level of the protein P1

H0: βP1=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP10 (There is significant relationship between expression level P1 and the cell size)

Expression Level of the protein P2

H0: βP2=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP20 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the p-value was near to 0.05 expression level of P1 & 'minor axis' may have slight statistical relation. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant. But the Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

[ ]:
```
## p1 on minor_axis

X=df[['p1_lvl']] # Independent variable
y=df['min_axis'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.02980122683987152


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.030
Model:                            OLS   Adj. R-squared:                  0.029
Method:                 Least Squares   F-statistic:                     90.80
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           3.20e-21
Time:                        11:57:08   Log-Likelihood:                -6903.6
No. Observations:                2958   AIC:                         1.381e+04
Df Residuals:                    2956   BIC:                         1.382e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const          2.0799      0.125     16.678      0.000       1.835       2.324
p1_lvl         0.0035      0.000      9.529      0.000       0.003       0.004
==============================================================================
Omnibus:                     6304.019   Durbin-Watson:                   1.835
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         32576778.945
Skew:                          18.247   Prob(JB):                         0.00
Kurtosis:                     515.819   Cond. No.                         911.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[ ]:

```python
## p1 on minor_axis under each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p1_lvl']] # Independent variables
  y=df1['min_axis'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 0.0007881755774573618


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.4804
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.489
Time:                        11:57:08   Log-Likelihood:                -596.61
No. Observations:                 611   AIC:                             1197.
Df Residuals:                     609   BIC:                             1206.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.8696      0.157     18.301      0.000       2.562       3.177
p1_lvl        -0.0004      0.001     -0.693      0.489      -0.001       0.001
==============================================================================
Omnibus:                       66.224   Durbin-Watson:                   2.116
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              148.519
Skew:                           0.605   Prob(JB):                     5.62e-33
```

```
Kurtosis:                       5.090   Cond. No.                     1.83e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.83e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

condition = e1

R-Squared : 0.002715398259708679

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.590
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.208
Time:                        11:57:08   Log-Likelihood:                -498.52
No. Observations:                 586   AIC:                             1001.
Df Residuals:                     584   BIC:                             1010.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.5506      0.074     34.331      0.000       2.405       2.697
p1_lvl         0.0006      0.000      1.261      0.208      -0.000       0.002
==============================================================================
Omnibus:                       17.248   Durbin-Watson:                   1.828
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               27.333
Skew:                           0.226   Prob(JB):                     1.16e-06
Kurtosis:                       3.957   Cond. No.                         499.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

condition = e2

R-Squared : 0.000784819011580562

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.4642
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.496
Time:                        11:57:08   Log-Likelihood:                -1527.9
No. Observations:                 593   AIC:                             3060.
Df Residuals:                     591   BIC:                             3069.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.7233      0.761      6.205      0.000       3.228       6.218
p1_lvl        -0.0017      0.003     -0.681      0.496      -0.007       0.003
==============================================================================
Omnibus:                      982.563   Durbin-Watson:                   2.015
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           508140.529
Skew:                           9.868   Prob(JB):                         0.00
Kurtosis:                     145.042   Cond. No.                     1.76e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.76e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e3

R-Squared : 0.0018108156129023056


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.112
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.292
Time:                        11:57:08   Log-Likelihood:                -1736.9
No. Observations:                 615   AIC:                             3478.
Df Residuals:                     613   BIC:                             3487.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
const          5.8465      1.601      3.651      0.000       2.702       8.991
p1_lvl        -0.0034      0.003     -1.055      0.292      -0.010       0.003
==============================================================================
Omnibus:                     1244.186   Durbin-Watson:                   1.980
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          2017683.423
Skew:                          14.746   Prob(JB):                         0.00
Kurtosis:                     282.050   Cond. No.                     4.89e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 4.89e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e4

R-Squared : 0.001481638141927144


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.8176
Date:                Thu, 19 Oct 2023   Prob (F-statistic):              0.366
Time:                        11:57:08   Log-Likelihood:                -246.67
No. Observations:                 553   AIC:                             497.3
Df Residuals:                     551   BIC:                             506.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.9417      0.104     18.588      0.000       1.736       2.147
p1_lvl         0.0003      0.000      0.904      0.366      -0.000       0.001
==============================================================================
Omnibus:                        7.016   Durbin-Watson:                   1.981
Prob(Omnibus):                  0.030   Jarque-Bera (JB):                7.177
Skew:                          -0.219   Prob(JB):                       0.0276
Kurtosis:                       3.346   Cond. No.                     1.98e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
[2] The condition number is large, 1.98e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Expression Level of the protein P1

H0: βP1=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP10 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the p-value was near to 0.05 expression level of P1 & 'minor axis' may have slight statistical relation.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant.

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant.

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

For the Expression level of P1 the p-value was greater than 0.05, so null hypotheses cannot be rejected and the relationship between expression level of P1 with 'minor axis' is not statistically significant.

[ ]:
```
## p2 on minor_axis

X=df[['p2_lvl']] # Independent variable
y=df['min_axis'] # Dependent variable

ols(X,y) # Call custom Ordinary Least Sqaues function
```

```
R-Squared : 0.19898345388212302
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.199
Model:                            OLS   Adj. R-squared:                  0.199
Method:                 Least Squares   F-statistic:                     734.3
Date:                Thu, 19 Oct 2023   Prob (F-statistic):          1.26e-144
Time:                        11:57:08   Log-Likelihood:                -6620.2
No. Observations:                2958   AIC:                         1.324e+04
Df Residuals:                    2956   BIC:                         1.326e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.2814      0.088     60.081      0.000       5.109       5.454
p2_lvl        -0.0078      0.000    -27.098      0.000      -0.008      -0.007
==============================================================================
Omnibus:                     6825.149   Durbin-Watson:                   1.948
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         57160801.785
Skew:                          21.949   Prob(JB):                         0.00
Kurtosis:                     682.597   Cond. No.                         645.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[ ]:
```
## p2 on minor_axis under each condition

for i in df["condition"].value_counts()[df.condition.unique()].keys():
  df1=df[df.condition==i]
  X=df1[['p2_lvl']] # Independent variables
  y=df1['min_axis'] # Dependent variable
  print('\033[1m'+"condition = "+i+'\033[0m\n')
  ols(X,y) # Call custom Ordinary Least Sqaues function
  print("\n")
```

```
condition = e0

R-Squared : 0.39660697903837494


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.397
Model:                            OLS   Adj. R-squared:                  0.396
Method:                 Least Squares   F-statistic:                     400.3
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           7.98e-69
```

```
Time:                        11:57:08   Log-Likelihood:                -442.51
No. Observations:                 611   AIC:                             889.0
Df Residuals:                     609   BIC:                             897.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.3902      0.084     52.365      0.000       4.226       4.555
p2_lvl        -0.0059      0.000    -20.007      0.000      -0.006      -0.005
==============================================================================
Omnibus:                       26.557   Durbin-Watson:                   2.164
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.629
Skew:                          -0.350   Prob(JB):                     1.50e-09
Kurtosis:                       4.051   Cond. No.                     1.18e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.18e+03. This might indicate that there are
strong multicollinearity or other numerical problems.


condition = e1

R-Squared : 0.34324909673931303


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.343
Model:                            OLS   Adj. R-squared:                  0.342
Method:                 Least Squares   F-statistic:                     305.2
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           2.69e-55
Time:                        11:57:08   Log-Likelihood:                -376.13
No. Observations:                 586   AIC:                             756.3
Df Residuals:                     584   BIC:                             765.0
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.0355      0.082     49.130      0.000       3.874       4.197
p2_lvl        -0.0047      0.000    -17.471      0.000      -0.005      -0.004
==============================================================================
Omnibus:                       46.407   Durbin-Watson:                   1.841
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               71.099
```

```
Skew:                          -0.573   Prob(JB):                     3.64e-16
Kurtosis:                       4.264   Cond. No.                     1.32e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.32e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
condition = e2

R-Squared : 0.2812725876101525
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.280
Method:                 Least Squares   F-statistic:                     231.3
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           2.54e-44
Time:                        11:57:08   Log-Likelihood:                -1430.2
No. Observations:                 593   AIC:                             2864.
Df Residuals:                     591   BIC:                             2873.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          7.8005      0.261     29.916      0.000       7.288       8.313
p2_lvl        -0.0246      0.002    -15.208      0.000      -0.028      -0.021
==============================================================================
Omnibus:                     1083.437   Durbin-Watson:                   2.042
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           988402.909
Skew:                          11.890   Prob(JB):                         0.00
Kurtosis:                     201.588   Cond. No.                         379.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

```
condition = e3

R-Squared : 0.18360210768763652
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.184
Model:                            OLS   Adj. R-squared:                  0.182
Method:                 Least Squares   F-statistic:                     137.9
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           7.43e-29
Time:                        11:57:08   Log-Likelihood:                -1675.1
No. Observations:                 615   AIC:                             3354.
Df Residuals:                     613   BIC:                             3363.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          7.9032      0.351     22.495      0.000       7.213       8.593
p2_lvl        -0.0252      0.002    -11.741      0.000      -0.029      -0.021
==============================================================================
Omnibus:                     1321.351   Durbin-Watson:                   1.965
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          3045660.970
Skew:                          16.794   Prob(JB):                         0.00
Kurtosis:                     346.114   Cond. No.                         386.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.


condition = e4

R-Squared : 0.16343642882360243


                            OLS Regression Results
==============================================================================
Dep. Variable:               min_axis   R-squared:                       0.163
Model:                            OLS   Adj. R-squared:                  0.162
Method:                 Least Squares   F-statistic:                     107.6
Date:                Thu, 19 Oct 2023   Prob (F-statistic):           3.71e-23
Time:                        11:57:08   Log-Likelihood:                -197.74
No. Observations:                 553   AIC:                             399.5
Df Residuals:                     551   BIC:                             408.1
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const          3.0976      0.103     29.937      0.000       2.894       3.301
p2_lvl        -0.0021      0.000    -10.375      0.000      -0.003      -0.002
==============================================================================
Omnibus:                       34.984   Durbin-Watson:                   2.022
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               42.027
Skew:                          -0.577   Prob(JB):                     7.48e-10
Kurtosis:                       3.703   Cond. No.                     3.55e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 3.55e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Expression Level of the protein P2

H0: βP2=0 (There is no significant linear between expression level P1 and the cell size)

HA: βP20 (There is significant relationship between expression level P1 and the cell size)

1. **Condition = e0 (Normal setup (WT) without any modifications)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

2. **Condition = e1 (The expression levels of the protein P1 were reduced while keeping that of P2 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

3. **Condition = e2 (The expression levels of the protein P2 were reduced while keeping that of P1 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

4. **Condition = e3 (The expression levels of the protein P1 were increased while keeping that of P2 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant.

5. **Condition = e4 (The expression levels of the protein P2 were increased while keeping that of P1 intact)**

The Expression level of P2 the p-value was less than 0.05, so null hypotheses can be rejected and the relationship between expression level of P2 with 'minor axis' is statistically significant

7 Final Inference

- Both P1 and P2 are not having much direct impact on the size (cell area).
- P1 does not have much direct impact on the shape (major and minor axes).
- P2 holds the key, having more impact on the shape. >* On the combined data, P2 has high positive correlation on the major axis and close to mid negative correlation on the minor axis. >* On the major axis, P2 is having high positive correlation for conditions e2x, e3x (either reducing P2 or increasing P1) . For the remaining 3, it's showing close to mid positive correlation. >* On minor axis, P2 has close to mid negative correlation. >* Impact during replication : »* On major axis, it's almost same in all replications. »* On minor axis, values are unstabe due to outliers.

8 Suggesting changes in experimental methodology.

Since P1 has no direct impact on the **size and shape** we can neglect it's effect. The computation can be made easier by reducing that input dimensionality without any loss of accuracy.

Minor axis values are unstable in replications due to outliers, it is better to record the minor axis values by removing the outliers.