

# Hands-on Activity 11.2 Classification using Logistic Regression

Submitted by: Dela Cruz, Eugene D.G.  
Submitted to: Engr. Roman Richard  
Section: CPE22S3

## Import libraries, dataset and install ucimlrepo package

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
from scipy import stats
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn import metrics
from sklearn.linear_model import LinearRegression

pip install ucimlrepo

Collecting ucimlrepo
  Downloading ucimlrepo-0.0.6-py3-none-any.whl (8.0 kB)
Installing collected packages: ucimlrepo
Successfully installed ucimlrepo-0.0.6

from ucimlrepo import fetch_ucirepo

# fetch dataset
cervical_cancer_risk_factors = fetch_ucirepo(id=383)

# data (as pandas dataframes)
X = cervical_cancer_risk_factors.data.features
y = cervical_cancer_risk_factors.data.targets

# metadata
print(cervical_cancer_risk_factors.metadata)

# variable information
print(cervical_cancer_risk_factors.variables)
```

|    |                                  |               |            |            |      |
|----|----------------------------------|---------------|------------|------------|------|
| 17 |                                  | STDs:syphilis | Feature    | Continuous | None |
| 18 | STDs:pelvic inflammatory disease | Feature       | Continuous | None       |      |
| 19 | STDs:genital herpes              | Feature       | Continuous | None       |      |
| 20 | STDs:molluscum contagiosum       | Feature       | Continuous | None       |      |
| 21 | STDs:AIDS                        | Feature       | Continuous | None       |      |
| 22 | STDs:HIV                         | Feature       | Continuous | None       |      |
| 23 | STDs:Hepatitis B                 | Feature       | Continuous | None       |      |
| 24 | STDs:HPV                         | Feature       | Continuous | None       |      |
| 25 | STDs: Number of diagnosis        | Feature       | Integer    | None       |      |
| 26 | STDs: Time since first diagnosis | Feature       | Continuous | None       |      |
| 27 | STDs: Time since last diagnosis  | Feature       | Continuous | None       |      |
| 28 | Dx:Cancer                        | Feature       | Integer    | None       |      |
| 29 | Dx:CIN                           | Feature       | Integer    | None       |      |
| 30 | Dx:HPV                           | Feature       | Integer    | None       |      |
| 31 | Dx                               | Feature       | Integer    | None       |      |
| 32 | Hinselmann                       | Feature       | Integer    | None       |      |
| 33 | Schiller                         | Feature       | Integer    | None       |      |
| 34 | Citology                         | Feature       | Integer    | None       |      |
| 35 | Biopsy                           | Feature       | Integer    | None       |      |

|    | description | units | missing_values |
|----|-------------|-------|----------------|
| 0  | None        | None  | no             |
| 1  | None        | None  | yes            |
| 2  | None        | None  | yes            |
| 3  | None        | None  | yes            |
| 4  | None        | None  | yes            |
| 5  | None        | None  | yes            |
| 6  | None        | None  | yes            |
| 7  | None        | None  | yes            |
| 8  | None        | None  | yes            |
| 9  | None        | None  | yes            |
| 10 | None        | None  | yes            |
| 11 | None        | None  | yes            |
| 12 | None        | None  | yes            |
| 13 | None        | None  | yes            |
| 14 | None        | None  | yes            |

|    |      |      |    |
|----|------|------|----|
| 34 | None | None | no |
| 35 | None | None | no |

▼ Data Wrangling/cleaning & Perform Exploratory Data Analysis

```
cancer = pd.concat([X,y], axis = 1)
cancer
```

|     | Age | Number of sexual partners | First sexual intercourse | Num of pregnancies | Smokes | Smokes (years) | Smokes (packs/year) | Hormonal Contraceptives | Hormonal Contraceptives (years) | IUD | ... | di |
|-----|-----|---------------------------|--------------------------|--------------------|--------|----------------|---------------------|-------------------------|---------------------------------|-----|-----|----|
| 0   | 18  | 4.0                       | 15.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.00                            | 0.0 | ... |    |
| 1   | 15  | 1.0                       | 14.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.00                            | 0.0 | ... |    |
| 2   | 34  | 1.0                       | NaN                      | 1.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.00                            | 0.0 | ... |    |
| 3   | 52  | 5.0                       | 16.0                     | 4.0                | 1.0    | 37.0           | 37.0                | 1.0                     | 3.00                            | 0.0 | ... |    |
| 4   | 46  | 3.0                       | 21.0                     | 4.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 15.00                           | 0.0 | ... |    |
| ... | ... | ...                       | ...                      | ...                | ...    | ...            | ...                 | ...                     | ...                             | ... | ... |    |
| 853 | 34  | 3.0                       | 18.0                     | 0.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.00                            | 0.0 | ... |    |
| 854 | 32  | 2.0                       | 19.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 8.00                            | 0.0 | ... |    |
| 855 | 25  | 2.0                       | 17.0                     | 0.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 0.08                            | 0.0 | ... |    |
| 856 | 33  | 2.0                       | 24.0                     | 2.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 0.08                            | 0.0 | ... |    |
| 857 | 29  | 2.0                       | 20.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 0.50                            | 0.0 | ... |    |

```
cancer.head()
```

|   | Age | Number of sexual partners | First sexual intercourse | Num of pregnancies | Smokes | Smokes (years) | Smokes (packs/year) | Hormonal Contraceptives | Hormonal Contraceptives (years) | IUD | ... | diag |
|---|-----|---------------------------|--------------------------|--------------------|--------|----------------|---------------------|-------------------------|---------------------------------|-----|-----|------|
| 0 | 18  | 4.0                       | 15.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.0                             | 0.0 | ... |      |
| 1 | 15  | 1.0                       | 14.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.0                             | 0.0 | ... |      |
| 2 | 34  | 1.0                       | NaN                      | 1.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.0                             | 0.0 | ... |      |
| 3 | 52  | 5.0                       | 16.0                     | 4.0                | 1.0    | 37.0           | 37.0                | 1.0                     | 3.0                             | 0.0 | ... |      |
| 4 | 46  | 3.0                       | 21.0                     | 4.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 15.0                            | 0.0 | ... |      |

```
cancer.tail()
```

|     | Age | Number of sexual partners | First sexual intercourse | Num of pregnancies | Smokes | Smokes (years) | Smokes (packs/year) | Hormonal Contraceptives | Hormonal Contraceptives (years) | IUD | ... | di |
|-----|-----|---------------------------|--------------------------|--------------------|--------|----------------|---------------------|-------------------------|---------------------------------|-----|-----|----|
| 853 | 34  | 3.0                       | 18.0                     | 0.0                | 0.0    | 0.0            | 0.0                 | 0.0                     | 0.00                            | 0.0 | ... |    |
| 854 | 32  | 2.0                       | 19.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 8.00                            | 0.0 | ... |    |
| 855 | 25  | 2.0                       | 17.0                     | 0.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 0.08                            | 0.0 | ... |    |
| 856 | 33  | 2.0                       | 24.0                     | 2.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 0.08                            | 0.0 | ... |    |
| 857 | 29  | 2.0                       | 20.0                     | 1.0                | 0.0    | 0.0            | 0.0                 | 1.0                     | 0.50                            | 0.0 | ... |    |

```
cancer.describe(include='all')
```

|       | Age        | Number of sexual partners | First sexual intercourse | Num of pregnancies | Smokes     | Smokes (years) | Smokes (packs/year) | Hormonal Contraceptives | Hormo Contracepti (yea |
|-------|------------|---------------------------|--------------------------|--------------------|------------|----------------|---------------------|-------------------------|------------------------|
| count | 858.000000 | 832.000000                | 851.000000               | 802.000000         | 845.000000 | 845.000000     | 845.000000          | 750.000000              | 750.000                |
| mean  | 26.820513  | 2.527644                  | 16.995300                | 2.275561           | 0.145562   | 1.219721       | 0.453144            | 0.641333                | 2.256                  |
| std   | 8.497948   | 1.667760                  | 2.803355                 | 1.447414           | 0.352876   | 4.089017       | 2.226610            | 0.479929                | 3.764                  |
| min   | 13.000000  | 1.000000                  | 10.000000                | 0.000000           | 0.000000   | 0.000000       | 0.000000            | 0.000000                | 0.000                  |
| 25%   | 20.000000  | 2.000000                  | 15.000000                | 1.000000           | 0.000000   | 0.000000       | 0.000000            | 0.000000                | 0.000                  |
| 50%   | 25.000000  | 2.000000                  | 17.000000                | 2.000000           | 0.000000   | 0.000000       | 0.000000            | 1.000000                | 0.500                  |
| 75%   | 32.000000  | 3.000000                  | 18.000000                | 3.000000           | 0.000000   | 0.000000       | 0.000000            | 1.000000                | 3.000                  |
| max   | 84.000000  | 28.000000                 | 32.000000                | 11.000000          | 1.000000   | 37.000000      | 37.000000           | 1.000000                | 30.000                 |

8 rows × 36 columns

```
cancer.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 858 entries, 0 to 857
Data columns (total 36 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                    858 non-null   int64
1   Number of sexual partners            832 non-null   float64
```

|    |                                    |              |         |
|----|------------------------------------|--------------|---------|
| 2  | First sexual intercourse           | 851 non-null | float64 |
| 3  | Num of pregnancies                 | 802 non-null | float64 |
| 4  | Smokes                             | 845 non-null | float64 |
| 5  | Smokes (years)                     | 845 non-null | float64 |
| 6  | Smokes (packs/year)                | 845 non-null | float64 |
| 7  | Hormonal Contraceptives            | 750 non-null | float64 |
| 8  | Hormonal Contraceptives (years)    | 750 non-null | float64 |
| 9  | IUD                                | 741 non-null | float64 |
| 10 | IUD (years)                        | 741 non-null | float64 |
| 11 | STDs                               | 753 non-null | float64 |
| 12 | STDs (number)                      | 753 non-null | float64 |
| 13 | STDs:condylomatosis                | 753 non-null | float64 |
| 14 | STDs:cervical condylomatosis       | 753 non-null | float64 |
| 15 | STDs:vaginal condylomatosis        | 753 non-null | float64 |
| 16 | STDs:vulvo-perineal condylomatosis | 753 non-null | float64 |
| 17 | STDs:syphilis                      | 753 non-null | float64 |
| 18 | STDs:pelvic inflammatory disease   | 753 non-null | float64 |
| 19 | STDs:genital herpes                | 753 non-null | float64 |
| 20 | STDs:molluscum contagiosum         | 753 non-null | float64 |
| 21 | STDs:AIDS                          | 753 non-null | float64 |
| 22 | STDs:HIV                           | 753 non-null | float64 |
| 23 | STDs:Hepatitis B                   | 753 non-null | float64 |
| 24 | STDs:HPV                           | 753 non-null | float64 |
| 25 | STDs: Number of diagnosis          | 858 non-null | int64   |
| 26 | STDs: Time since first diagnosis   | 71 non-null  | float64 |
| 27 | STDs: Time since last diagnosis    | 71 non-null  | float64 |
| 28 | Dx:Cancer                          | 858 non-null | int64   |
| 29 | Dx:CIN                             | 858 non-null | int64   |
| 30 | Dx:HPV                             | 858 non-null | int64   |
| 31 | Dx                                 | 858 non-null | int64   |
| 32 | Hinselmann                         | 858 non-null | int64   |
| 33 | Schiller                           | 858 non-null | int64   |
| 34 | Citology                           | 858 non-null | int64   |
| 35 | Biopsy                             | 858 non-null | int64   |

dtypes: float64(26), int64(10)  
memory usage: 241.4 KB

```
cancer.isna().sum()
```

|                                    |     |
|------------------------------------|-----|
| Age                                | 0   |
| Number of sexual partners          | 26  |
| First sexual intercourse           | 7   |
| Num of pregnancies                 | 0   |
| Smokes                             | 13  |
| Smokes (years)                     | 13  |
| Smokes (packs/year)                | 13  |
| Hormonal Contraceptives            | 108 |
| Hormonal Contraceptives (years)    | 0   |
| IUD                                | 117 |
| IUD (years)                        | 0   |
| STDs                               | 105 |
| STDs (number)                      | 0   |
| STDs:condylomatosis                | 105 |
| STDs:cervical condylomatosis       | 105 |
| STDs:vaginal condylomatosis        | 105 |
| STDs:vulvo-perineal condylomatosis | 105 |
| STDs:syphilis                      | 105 |
| STDs:pelvic inflammatory disease   | 105 |
| STDs:genital herpes                | 105 |
| STDs:molluscum contagiosum         | 105 |
| STDs:AIDS                          | 105 |
| STDs:HIV                           | 105 |
| STDs:Hepatitis B                   | 105 |
| STDs:HPV                           | 105 |
| STDs: Number of diagnosis          | 0   |
| Dx:Cancer                          | 0   |
| Dx:CIN                             | 0   |
| Dx:HPV                             | 0   |
| Dx                                 | 0   |
| Hinselmann                         | 0   |
| Schiller                           | 0   |
| Citology                           | 0   |
| Biopsy                             | 0   |

dtype: int64

```
cancer.dropna(subset=['First sexual intercourse','Number of sexual partners',
                      'Smokes (years)'],inplace = True)
```

```
def pdf(x):
    mean = np.mean(x)
    print(mean)
    std = np.std(x)
    print(std)
    y_out = 1/(std * np.sqrt(2 * np.pi)) * np.exp( - (x - mean)**2 / (2 * std**2))
    return y_out
```

```
coltoint = {'Number of sexual partners': int,
            'Num of pregnancies' : int}
cancer = cancer.astype(coltoint)
```

```
cancer.isna().sum()
```

|                                 |     |
|---------------------------------|-----|
| Age                             | 0   |
| Number of sexual partners       | 0   |
| First sexual intercourse        | 0   |
| Num of pregnancies              | 0   |
| Smokes                          | 0   |
| Smokes (years)                  | 0   |
| Smokes (packs/year)             | 0   |
| Hormonal Contraceptives         | 91  |
| Hormonal Contraceptives (years) | 0   |
| IUD                             | 101 |
| IUD (years)                     | 0   |
| STDs                            | 90  |
| STDs (number)                   | 0   |
| STDs:condylomatosis             | 90  |
| STDs:cervical condylomatosis    | 90  |

```
STDs:vaginal condylomatosis      90
STDs:vulvo-perineal condylomatosis 90
STDs:syphilis                    90
STDs:pelvic inflammatory disease 90
STDs:genital herpes              90
STDs:molluscum contagiosum       90
STDs:AIDS                        90
STDs:HIV                        90
STDs:Hepatitis B                 90
STDs:HPV                        90
STDs: Number of diagnosis        0
Dx:Cancer                        0
Dx:CIN                           0
Dx:HPV                           0
Dx                               0
Hinselmann                      0
Schiller                        0
Citology                        0
Biopsy                          0
dtype: int64
```

```
# removing unnecessary columns for easy access data
cancer.drop(columns= ['STDs:condylomatosis', 'STDs:cervical condylomatosis',
                      'STDs:vaginal condylomatosis', 'STDs:vulvo-perineal condylomatosis',
                      'STDs:syphilis','STDs:pelvic inflammatory disease',
                      'STDs:genital herpes','STDs:molluscum contagiosum',
                      'STDs:AIDS','STDs:HIV','STDs:Hepatitis B',
                      'STDs:HPV', 'First sexual intercourse' ],inplace = True)
```

```
-----
KeyError                                Traceback (most recent call last)
<ipython-input-48-d3fed2252223> in <cell line: 2>()
      1 # removing unnecessary columns for easy access data
----> 2 cancer.drop(columns= ['STDs:condylomatosis', 'STDs:cervical condylomatosis',
      3                      'STDs:vaginal condylomatosis', 'STDs:vulvo-perineal condylomatosis',
      4                      'STDs:syphilis','STDs:pelvic inflammatory disease',
      5                      'STDs:genital herpes','STDs:molluscum contagiosum',

      ^ 3 frames
/usr/local/lib/python3.10/dist-packages/pandas/core/indexes/base.py in drop(self, labels, errors)
    6697         if mask.any():
    6698             if errors != "ignore":
-> 6699                 raise KeyError(f"{list(labels[mask])} not found in axis")
    6700             indexer = indexer[~mask]
    6701         return self.delete(indexer)

KeyError: "[ 'STDs:condylomatosis', 'STDs:cervical condylomatosis', 'STDs:vaginal condylomatosis', 'STDs:vulvo-perineal
condylomatosis', 'STDs:syphilis', 'STDs:pelvic inflammatory disease', 'STDs:genital herpes', 'STDs:molluscum
contagiosum', 'STDs:AIDS', 'STDs:HIV', 'STDs:Hepatitis B', 'STDs:HPV'] not found in axis"
```

```
# because of the column "STDs: Time since first diagnosis" and "STDs: Time since last diagnosis"
# has a great amount of non-null values I will remove these columns
cancer = cancer.drop(columns=['STDs: Time since first diagnosis','STDs: Time since last diagnosis'])
cancer
```

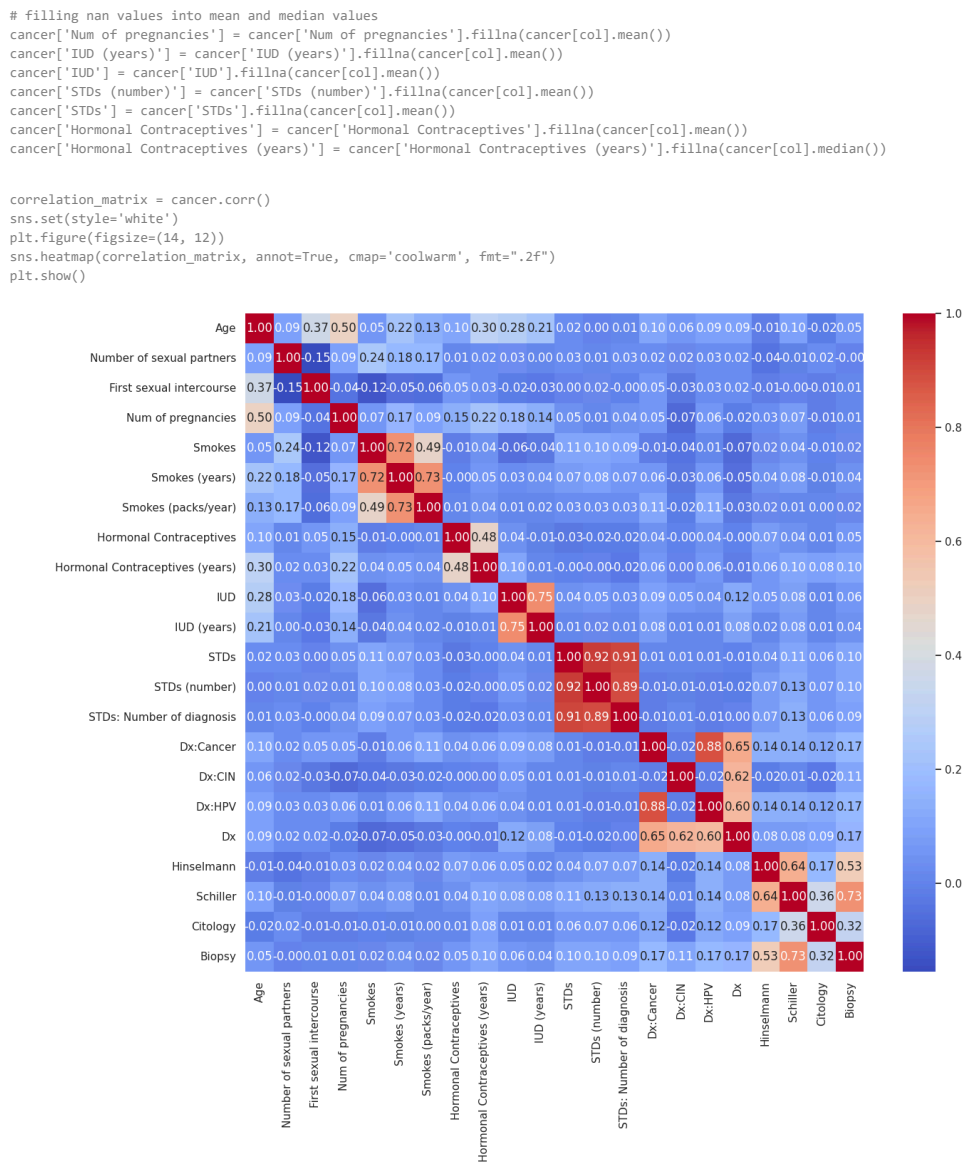
|     | Age | Number<br>of<br>sexual<br>partners | First<br>sexual<br>intercourse | Num of<br>pregnancies | Smokes | Smokes<br>(years) | Smokes<br>(packs/year) | Hormonal<br>Contraceptives | Hormonal<br>Contraceptives<br>(years) | IUD | ... | ST  |
|-----|-----|------------------------------------|--------------------------------|-----------------------|--------|-------------------|------------------------|----------------------------|---------------------------------------|-----|-----|-----|
| 0   | 18  | 4.0                                | 15.0                           | 1.0                   | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... |     |
| 1   | 15  | 1.0                                | 14.0                           | 1.0                   | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... |     |
| 2   | 34  | 1.0                                | NaN                            | 1.0                   | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... |     |
| 3   | 52  | 5.0                                | 16.0                           | 4.0                   | 1.0    | 37.0              | 37.0                   | 1.0                        | 3.00                                  | 0.0 | ... |     |
| 4   | 46  | 3.0                                | 21.0                           | 4.0                   | 0.0    | 0.0               | 0.0                    | 1.0                        | 15.00                                 | 0.0 | ... |     |
| ... | ... | ...                                | ...                            | ...                   | ...    | ...               | ...                    | ...                        | ...                                   | ... | ... | ... |
| 853 | 34  | 3.0                                | 18.0                           | 0.0                   | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... |     |
| 854 | 32  | 2.0                                | 19.0                           | 1.0                   | 0.0    | 0.0               | 0.0                    | 1.0                        | 8.00                                  | 0.0 | ... |     |
| 855 | 25  | 2.0                                | 17.0                           | 0.0                   | 0.0    | 0.0               | 0.0                    | 1.0                        | 0.08                                  | 0.0 | ... |     |
| 856 | 33  | 2.0                                | 24.0                           | 2.0                   | 0.0    | 0.0               | 0.0                    | 1.0                        | 0.08                                  | 0.0 | ... |     |
| 857 | 29  | 2.0                                | 20.0                           | 1.0                   | 0.0    | 0.0               | 0.0                    | 1.0                        | 0.50                                  | 0.0 | ... |     |

```
for col in cancer.columns:
    if cancer[col].isna().sum() > 0:
        print(f'{col}\n'
              f'mean: {cancer[col].mean()}\n'
              f'median:{cancer[col].median()}\n'
              f'mode: {stats.mode(cancer[col])}\n')

Hormonal Contraceptives
mean: 0.6459197786998617
median:1.0
mode: ModeResult(mode=1.0, count=467)

IUD
mean: 0.11360448807854137
median:0.0
mode: ModeResult(mode=0.0, count=632)

STDs
mean: 0.10359116022099447
median:0.0
mode: ModeResult(mode=0.0, count=649)
```



## Training

```
X = cancer.drop(columns = 'Biopsy')
y = cancer['Biopsy']

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size = 0.3, random_state = 21)

X_train
```

|     | Age | Number<br>of<br>sexual<br>partners | First<br>sexual<br>intercourse | Num of<br>pregnancies | Smokes | Smokes<br>(years) | Smokes<br>(packs/year) | Hormonal<br>Contraceptives | Hormonal<br>Contraceptives<br>(years) | IUD | ... | ST  |
|-----|-----|------------------------------------|--------------------------------|-----------------------|--------|-------------------|------------------------|----------------------------|---------------------------------------|-----|-----|-----|
| 580 | 17  | 2                                  | 15.0                           | 1                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 0.33                                  | 0.0 | ... | I   |
| 733 | 31  | 2                                  | 18.0                           | 4                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 3.00                                  | 0.0 | ... | I   |
| 831 | 40  | 3                                  | 23.0                           | 2                     | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... | I   |
| 604 | 23  | 3                                  | 17.0                           | 2                     | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... | I   |
| 59  | 35  | 3                                  | 17.0                           | 4                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 7.00                                  | 1.0 | ... | I   |
| ... | ... | ...                                | ...                            | ...                   | ...    | ...               | ...                    | ...                        | ...                                   | ... | ... | ... |
| 126 | 31  | 2                                  | 19.0                           | 2                     | 0.0    | 0.0               | 0.0                    | 0.0                        | 0.00                                  | 0.0 | ... | I   |
| 390 | 20  | 2                                  | 17.0                           | 1                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 0.08                                  | 0.0 | ... | I   |
| 52  | 36  | 5                                  | 17.0                           | 3                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 0.50                                  | 0.0 | ... | I   |
| 813 | 20  | 1                                  | 17.0                           | 0                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 1.00                                  | 0.0 | ... | I   |
| 218 | 28  | 3                                  | 17.0                           | 4                     | 0.0    | 0.0               | 0.0                    | 1.0                        | 1.00                                  | 0.0 | ... | I   |

X\_test

|     | Age | Number<br>of<br>sexual<br>partners | First<br>sexual<br>intercourse | Num of<br>pregnancies | Smokes | Smokes<br>(years) | Smokes<br>(packs/year) | Hormonal<br>Contraceptives | Hormonal<br>Contraceptives<br>(years) | IUD      | ..  |
|-----|-----|------------------------------------|--------------------------------|-----------------------|--------|-------------------|------------------------|----------------------------|---------------------------------------|----------|-----|
| 69  | 35  | 5                                  | 23.0                           | 2                     | 0.0    | 0.0               | 0.00                   | 1.000000                   | 0.33                                  | 0.000000 | .   |
| 701 | 29  | 2                                  | 17.0                           | 0                     | 0.0    | 0.0               | 0.00                   | 0.065111                   | 0.00                                  | 0.065111 | .   |
| 538 | 24  | 3                                  | 18.0                           | 2                     | 0.0    | 0.0               | 0.00                   | 1.000000                   | 0.16                                  | 0.000000 | .   |
| 676 | 70  | 4                                  | 27.0                           | 3                     | 1.0    | 3.0               | 0.75                   | 0.000000                   | 0.00                                  | 0.000000 | .   |
| 448 | 17  | 3                                  | 16.0                           | 1                     | 0.0    | 0.0               | 0.00                   | 1.000000                   | 0.25                                  | 0.000000 | .   |
| ... | ... | ...                                | ...                            | ...                   | ...    | ...               | ...                    | ...                        | ...                                   | ...      | ... |
| 406 | 17  | 1                                  | 17.0                           | 1                     | 0.0    | 0.0               | 0.00                   | 0.000000                   | 0.00                                  | 0.000000 | .   |
| 429 | 18  | 1                                  | 15.0                           | 2                     | 0.0    | 0.0               | 0.00                   | 1.000000                   | 0.08                                  | 0.000000 | .   |
| 455 | 18  | 1                                  | 16.0                           | 2                     | 0.0    | 0.0               | 0.00                   | 0.000000                   | 0.00                                  | 0.000000 | .   |
| 486 | 28  | 2                                  | 20.0                           | 2                     | 0.0    | 0.0               | 0.00                   | 1.000000                   | 1.00                                  | 0.000000 | .   |
| 183 | 28  | 2                                  | 18.0                           | 3                     | 0.0    | 0.0               | 0.00                   | 0.000000                   | 0.00                                  | 0.000000 | .   |

y\_train

```
580    0
733    0
831    0
604    0
59     0
..
126    0
390    0
52     0
813    0
218    0
Name: Biopsy, Length: 569, dtype: int64
```

y\_test

```
69     0
701    0
538    0
676    0
448    0
..
406    0
429    0
455    0
486    0
183    0
Name: Biopsy, Length: 245, dtype: int64
```

## Linear Regression

```
model = LinearRegression()
```

```
model.fit(X_train, y_train)
```

```
LinearRegression
LinearRegression()
```

## Logistic Regression

```
model = LogisticRegression(random_state = 1,
                           C=0.1, fit_intercept = True).fit(X_train, y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

Increase the number of iterations (`max_iter`) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

[https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
model.predict(X_train)
```

[illegible]

```
model.predict(X_test)
```

[illegible]

```
model.score(X_train, y_train)
```

0.9384885764499121

```
model.score(X_test, y_test)
```

0.9346938775510204