In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
health=pd.read_sas(r"C:\Users\Lenovo\Downloads\health.sas7bdat")
```

In [3]:
```
health.head()
```

Out[3]:

	X1	X2	X3	X4	X5	x6
0	64.9	78.0	284.0	9.1	109.0	28.0
1	70.3	68.0	433.0	8.7	144.0	29.0
2	60.8	70.0	739.0	7.2	113.0	27.0
3	72.5	25.0	250.0	2.5	34.0	23.0
4	76.7	74.0	477.0	8.3	206.0	21.0

In [4]:
```
health.shape
```

Out[4]:
```
(53, 6)
```

In [10]:
```
sns.distplot(health["X1"])
```

```
C:\Users\Lenovo\.ipython\extensions\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[10]:
```
<AxesSubplot:xlabel='X1', ylabel='Density'>
```

In [9]:
```

```

Out[9]:
```
<AxesSubplot:xlabel='X1', ylabel='Density'>
```

In [13]:
```
import statsmodels.formula.api as sm


model=sm.ols(formula="X1~X2+X3+X4+X5+x6",data=health).fit()
model.summary()
```

Out[13]:

OLS Regression Results

Dep. Variable:	X1	**R-squared:**	0.953
Model:	OLS	**Adj. R-squared:**	0.948
Method:	Least Squares	**F-statistic:**	191.1
Date:	Thu, 02 Mar 2023	**Prob (F-statistic):**	5.06e-30
Time:	11:47:58	**Log-Likelihood:**	-102.74
No. Observations:	53	**AIC:**	217.5
Df Residuals:	47	**BIC:**	229.3
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	75.4490	2.619	28.814	0.000	70.181	80.717
X2	-0.1025	0.008	-12.644	0.000	-0.119	-0.086
X3	-0.0118	0.001	-11.498	0.000	-0.014	-0.010
X4	-1.0950	0.211	-5.202	0.000	-1.519	-0.672
X5	0.0994	0.006	17.698	0.000	0.088	0.111
x6	0.0503	0.092	0.548	0.586	-0.134	0.235

Omnibus:	4.263	**Durbin-Watson:**	2.116
Prob(Omnibus):	0.119	**Jarque-Bera (JB):**	2.609
Skew:	-0.337	**Prob(JB):**	0.271
Kurtosis:	2.148	**Cond. No.**	6.93e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 6.93e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Here i removed a one reduant variable which the p value is > 0.05

If the p value is > 0.05 assume that that variable

In [14]:
```
import statsmodels.formula.api as sm


model=sm.ols(formula="X1~X2+X3+X4+X5",data=health).fit()
model.summary()
```

Out[14]:

OLS Regression Results

Dep. Variable:	X1	**R-squared:**	0.953
Model:	OLS	**Adj. R-squared:**	0.949
Method:	Least Squares	**F-statistic:**	242.4
Date:	Thu, 02 Mar 2023	**Prob (F-statistic):**	3.53e-31
Time:	11:54:53	**Log-Likelihood:**	-102.91
No. Observations:	53	**AIC:**	215.8
Df Residuals:	48	**BIC:**	225.7
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	76.5587	1.650	46.402	0.000	73.241	79.876
X2	-0.1020	0.008	-12.757	0.000	-0.118	-0.086
X3	-0.0116	0.001	-11.890	0.000	-0.014	-0.010
X4	-1.0878	0.209	-5.215	0.000	-1.507	-0.668
X5	0.0989	0.006	17.957	0.000	0.088	0.110

Omnibus:	5.238	**Durbin-Watson:**	2.162
Prob(Omnibus):	0.073	**Jarque-Bera (JB):**	2.917
Skew:	-0.351	**Prob(JB):**	0.233
Kurtosis:	2.091	**Cond. No.**	4.42e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.42e+03. This might indicate that there are strong multicollinearity or other numerical problems.

In [15]:
```
air=pd.read_csv(r"C:\Users\Lenovo\Downloads\data sets\AirPassengers.csv")
```

In [16]:
```
air.head()
```

Out[16]:

	Week_num	Passengers	Promotion_Budget	Service_Quality_Score	Holiday_week	Delayed
0	1	37824	517356	4.00000	NO	
1	2	43936	646086	2.67466	NO	
2	3	42896	638330	3.29473	NO	
3	4	35792	506492	3.85684	NO	
4	5	38624	609658	3.90757	NO	

In [27]:
```
a=air.select_dtypes(exclude="object").copy()
```

In [28]:
```
a.head()
```

Out[28]:

	Week_num	Passengers	Promotion_Budget	Service_Quality_Score	Inter_metro_flight_ratio
0	1	37824	517356	4.00000	0.70
1	2	43936	646086	2.67466	0.80
2	3	42896	638330	3.29473	0.90
3	4	35792	506492	3.85684	0.40
4	5	38624	609658	3.90757	0.87

In [40]:
```
import statsmodels.formula.api as sm

model=sm.ols(formula="X1~X2+X3+X4+X5",data=health).fit()
model.summary()
```

Out[40]:

OLS Regression Results

Dep. Variable:	X1	**R-squared:**	0.953
Model:	OLS	**Adj. R-squared:**	0.949
Method:	Least Squares	**F-statistic:**	242.4
Date:	Thu, 02 Mar 2023	**Prob (F-statistic):**	3.53e-31
Time:	12:26:29	**Log-Likelihood:**	-102.91
No. Observations:	53	**AIC:**	215.8
Df Residuals:	48	**BIC:**	225.7
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	76.5587	1.650	46.402	0.000	73.241	79.876
X2	-0.1020	0.008	-12.757	0.000	-0.118	-0.086
X3	-0.0116	0.001	-11.890	0.000	-0.014	-0.010
X4	-1.0878	0.209	-5.215	0.000	-1.507	-0.668
X5	0.0989	0.006	17.957	0.000	0.088	0.110

Omnibus:	5.238	**Durbin-Watson:**	2.162
Prob(Omnibus):	0.073	**Jarque-Bera (JB):**	2.917
Skew:	-0.351	**Prob(JB):**	0.233
Kurtosis:	2.091	**Cond. No.**	4.42e+03

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.42e+03. This might indicate that there are strong multicollinearity or other numerical problems.

In [38]:
```
air
```

Out[38]:

	Week_num	Passengers	Promotion_Budget	Service_Quality_Score	Holiday_week	Delaye
0	1	37824	517356	4.00000	NO	
1	2	43936	646086	2.67466	NO	
2	3	42896	638330	3.29473	NO	
3	4	35792	506492	3.85684	NO	
4	5	38624	609658	3.90757	NO	
...	...	...	...	...	...	
75	76	66934	927696	1.07138	YES	
76	77	81228	1108254	0.85536	YES	
77	78	43288	638162	3.08191	NO	
78	79	43834	636636	2.75382	NO	
79	80	40852	575008	3.52768	NO	

80 rows × 9 columns

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: