In [6]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import statsmodels.formula.api as smf
```

In [7]:

```
dataset=pd.read_csv("C:\\Users\\Admin\\Downloads\\assignment 3\\Salary_Data.csv")
dataset
```

Out[7]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0

	YearsExperience	Salary
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

In [8]:

```
dataset.info()
```

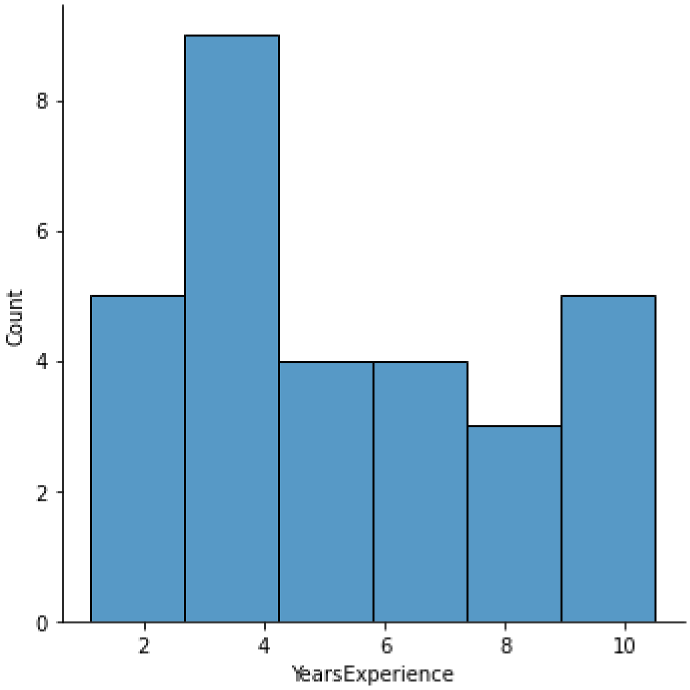
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   YearsExperience  30 non-null     float64
 1   Salary           30 non-null     float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

In [9]:

```
sns.displot(dataset['YearsExperience'])
```

Out[9]:

```
<seaborn.axisgrid.FacetGrid at 0x21677a4b880>
```

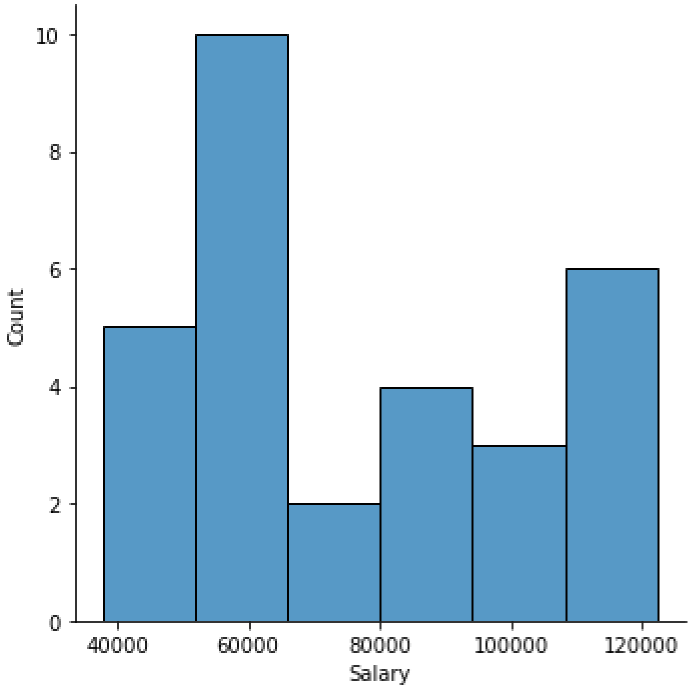


In [10]:

```
sns.displot(dataset['Salary'])
```

Out[10]:

```
<seaborn.axisgrid.FacetGrid at 0x21677a4b4f0>
```



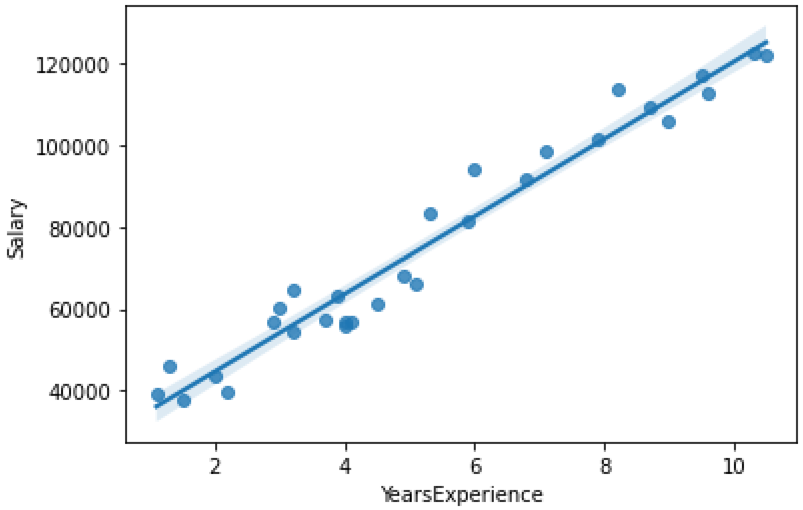
```
In [11]: dataset.corr()
```

Out[11]:

	YearsExperience	Salary
YearsExperience	1.000000	0.978242
Salary	0.978242	1.000000

```
In [12]: sns.regplot(x=dataset['YearsExperience'],y=dataset['Salary'])
```

```
Out[12]: <AxesSubplot:xlabel='YearsExperience', ylabel='Salary'>
```



```
In [13]: model=smf.ols("Salary~YearsExperience",data=dataset).fit()
```

```
In [14]: model.summary()
```

Out[14]:

OLS Regression Results

Dep. Variable:	Salary	**R-squared:**	0.957
Model:	OLS	**Adj. R-squared:**	0.955
Method:	Least Squares	**F-statistic:**	622.5
Date:	Sun, 20 Mar 2022	**Prob (F-statistic):**	1.14e-20
Time:	18:47:38	**Log-Likelihood:**	-301.44
No. Observations:	30	**AIC:**	606.9
Df Residuals:	28	**BIC:**	609.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.579e+04	2273.053	11.347	0.000	2.11e+04	3.04e+04
YearsExperience	9449.9623	378.755	24.950	0.000	8674.119	1.02e+04

Omnibus:	2.140	**Durbin-Watson:**	1.648
Prob(Omnibus):	0.343	**Jarque-Bera (JB):**	1.569
Skew:	0.363	**Prob(JB):**	0.456
Kurtosis:	2.147	**Cond. No.**	13.2

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

In [15]:

```
model.params
```

Out[15]:

```
Intercept          25792.200199
YearsExperience     9449.962321
dtype: float64
```

In [16]:

```
model.tvalues,model.pvalues
```

Out[16]:

```
(Intercept          11.346940
 YearsExperience    24.950094
 dtype: float64,
 Intercept          5.511950e-12
 YearsExperience    1.143068e-20
 dtype: float64)
```

In [17]:

```
model.rsquared,model.rsquared_adj
```

Out[17]:

```
(0.9569566641435086, 0.9554194021486339)
```

In [18]:

```
Salary=(25792.200199)+(9449.962321)*(3)
Salary
```

Out[18]:

```
54142.087162
```

In [19]:

```
new_data=pd.Series([3,5])
new_data
```

Out[19]:

```
0    3
1    5
dtype: int64
```

In [20]:

```
data_pred=pd.DataFrame(new_data,columns=['YearsExperience'])
data_pred
```

Out[20]:

	YearsExperience
0	3
1	5

In [21]:

```
model.predict(data_pred)
```

Out[21]:

```
0    54142.087163
1    73042.011806
dtype: float64
```

In [ ]: