In [1]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as sf
import numpy as np
```

In [2]:

```
titanic = pd.read_csv(r'C:\Users\Admin\OneDrive\Рабочий стол\DataScience
```

In [3]:

```
titanic.head()
```

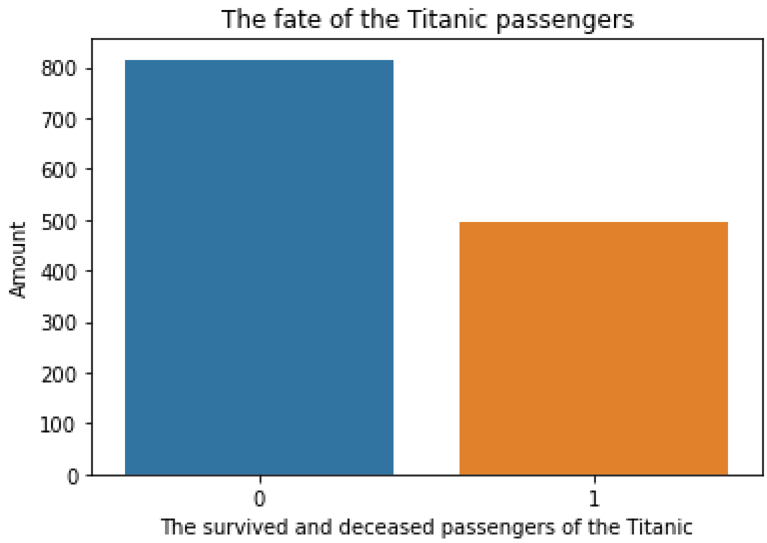
Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [6]:

```
sns.countplot(x='Survived', data=titanic)
plt.xlabel('The survived and deceased passengers of the Titanic')
plt.ylabel('Amount')
plt.title('The fate of the Titanic passengers')
```

Out[6]: Text(0.5, 1.0, 'The fate of the Titanic passengers')



*0 - perished and 1 -survived

*Afterward there are featued the application of the logistic analysis

In [6]:

```
logistic_res=sf.glm('Survived ~ C(Pclass) + C(Sex) + Age', titanic, fami
logistic_res.summary()
```

Out[6]:

Generalized Linear Model Regression Results

Dep. Variable:	Survived	**No. Observations:**	1046
Model:	GLM	**Df Residuals:**	1041
Model Family:	Binomial	**Df Model:**	4
Link Function:	logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-398.21
Date:	Sat, 23 Oct 2021	**Deviance:**	796.42
Time:	13:39:01	**Pearson chi2:**	1.10e+03
No. Iterations:	5		
Covariance Type:	nonrobust		

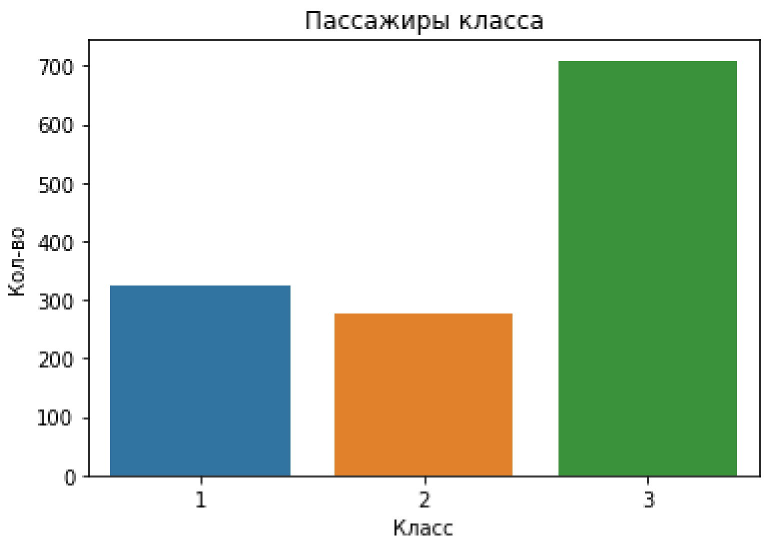
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	3.9568	0.372	10.641	0.000	3.228	4.686
C(Pclass)[T.2]	-1.2370	0.256	-4.833	0.000	-1.739	-0.735
C(Pclass)[T.3]	-2.2390	0.256	-8.759	0.000	-2.740	-1.738
C(Sex)[T.male]	-3.5601	0.196	-18.158	0.000	-3.944	-3.176
Age	-0.0313	0.007	-4.407	0.000	-0.045	-0.017

According to the data summary we are able to emphasize that the given endogenous variables are statically significant relative to the exogenous one that so-called 'survived' because of p-value is less than 0.05. Moreover, it is seen that only passengers of both first-class and women had highly possibility to survive compared to the others.

In [8]:

```
sns.countplot(x="Pclass", data=titanic)
plt.xlabel("Класс")
plt.ylabel("Кол-во")
plt.title('Пассажиры класса')
```

Out[8]: Text(0.5, 1.0, 'Пассажиры класса')



Multinomial logistic regression is the second method of logistic regression. There are presented 3 types of classes where we will put in the formula as an exogenous variable 'Pclass' and 'Sex' + 'Age' will be as endogenous variables. This model allows conducting analysis employing several endogenous variables that affect one exogenous variable simultaneously.

In [9]:

```
multi_res = sf.mnlogit('Pclass ~ C(Sex) + Age', titanic).fit()
multi_res.summary()
```

```
Optimization terminated successfully.
         Current function value: 0.944697
         Iterations 5
```

Out[9]:

MNLogit Regression Results

Dep. Variable:	Pclass	**No. Observations:**	1046
Model:	MNLogit	**Df Residuals:**	1040
Method:	MLE	**Df Model:**	4
Date:	Wed, 27 Oct 2021	**Pseudo R-squ.:**	0.1028
Time:	23:32:08	**Log-Likelihood:**	-988.15
converged:	True	**LL-Null:**	-1101.4
Covariance Type:	nonrobust	**LLR p-value:**	7.608e-48

Pclass=2	**coef**	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	1.4181	0.253	5.611	0.000	0.923	1.913
C(Sex)[T.male]	0.4934	0.183	2.691	0.007	0.134	0.853
Age	-0.0522	0.007	-7.728	0.000	-0.065	-0.039
Pclass=3	**coef**	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	2.6021	0.239	10.900	0.000	2.134	3.070
C(Sex)[T.male]	1.0119	0.174	5.824	0.000	0.671	1.352
Age	-0.0843	0.007	-12.514	0.000	-0.098	-0.071

To sum up we can conclude that 'Pclass-2' had more chances to survive that night than

'Pclass3' but we must remember that it occurs only relative to the 'Class 1'.

```
In [ ]:
```