

Регрессионный анализ

Используемая литература для подготовки задач к семинару:

Гусак А.А. «Теория вероятностей»

Гланц С. «Медико-биологическая статистика»



Постройте графики для приведенных наборов данных. Найдите коэффициенты для линии регрессии и коэффициенты детерминации. Что вы замечаете? Нанесите на график модель линейной регрессии.

```
X1= np.array([30,30,40, 40])
```

```
Y1= np.array([37, 47, 50, 60])
```

```
x2= np.array([30,30,40, 40, 20, 20, 50, 50])
```

```
y2= np.array([37, 47, 50, 60, 25, 35, 62, 72])
```

```
X3 = np.array([30,30,40, 40, 20, 20, 50, 50, 10, 10, 60, 60])
```

```
Y3 = np.array([37, 47, 50, 60, 25, 35, 62, 72, 13, 23, 74, 84])
```

Используемая литература для подготовки задач к семинару:

Гусак А.А. «Теория вероятностей»

Гланц С. «Медико-биологическая статистика»



```
import scipy.stats as stats  
import numpy as np  
import matplotlib.pyplot as plt
```

```
plt.scatter(x, y)  
plt.title('r= вписать значение коэффициента')  
plt.xlabel('x')  
plt.ylabel('y')  
plt.show()
```



На 8 уроке мы строили графики приведенных ниже данных.
Для какого графика можно использовать модель линейной регрессии?

```
x = np.array([10,8, 13, 9,11,14, 6,4,12, 7,5])  
y = np.array([8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68 ])
```

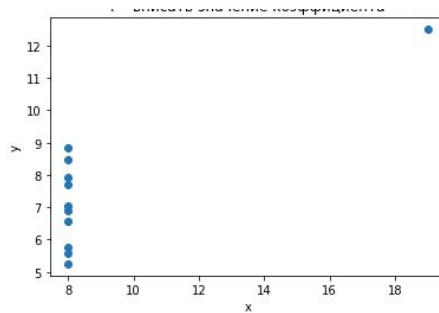
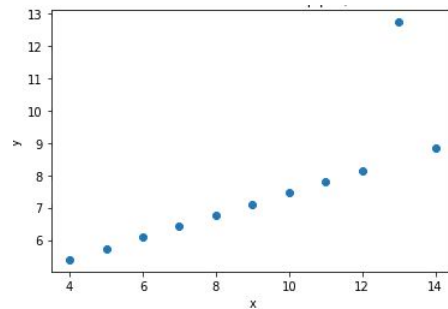
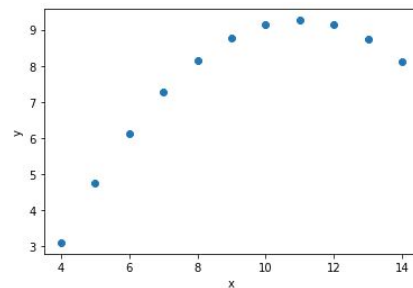
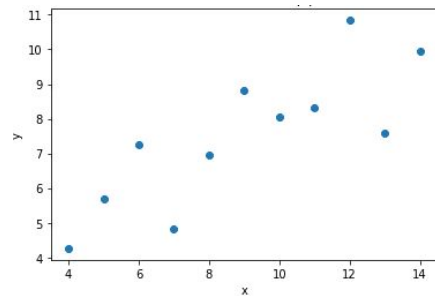
```
x= np.array([ 10,8, 13, 9,11,14, 6,4,12, 7,5 ])  
y2 = np.array([ 9.14, 8.14, 8.74,8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74])
```

```
x= np.array([ 10,8, 13, 9,11,14, 6,4,12, 7,5 ])  
y3 = np.array([7.46,6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73])
```

```
x4 = np.array([8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8])  
y4 = np.array([6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25,12.5, 5.56, 7.91, 6.89])
```

```
x0= np.array([ 10, 8, 13, 9, 11, 14, 6, 4, 12, 7,5, 15, 16, 18 ])  
y0 = np.array([ 9.14, 8.14, 8.74,8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74, 6.5, 5, 2.9])
```

Графики , которые должны были получиться





Постройте модель линейной регрессии для подходящих данных



```
from sklearn.linear_model import LinearRegression  
import matplotlib.pyplot as plt
```

```
model = LinearRegression()
```

```
model = LinearRegression()  
model.fit(X1, Y1)  
r_sq = model.score(X1, Y1)  
print('coefficient of determination:', r_sq)  
print('intercept:', model.intercept_)  
print('slope:', model.coef_)
```



```
plt.scatter(x, y)  
plt.plot(x, 3+0.5*x)  
plt.show()
```




Проверить оставшиеся условия применимости линейной регрессии



```
In [14]: y_hat= 3+0.5*x  
y_hat
```

```
Out[14]: array([[ 8. ],  
[ 7. ],  
[ 9.5],  
[ 7.5],  
[ 8.5],  
[10. ],  
[ 6. ],  
[ 5. ],  
[ 9. ],  
[ 6.5],  
[ 5.5]])
```

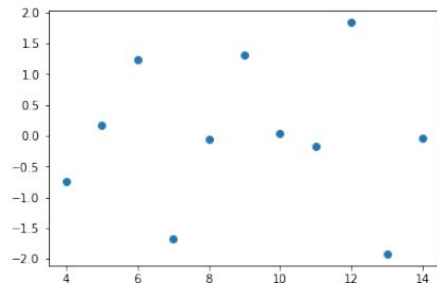
```
In [16]: y_hat1= y_hat.reshape(1,-1)  
y_hat1
```

```
Out[16]: array([[ 8. ,  7. ,  9.5,  7.5,  8.5, 10. ,  6. ,  5. ,  9. ,  6.5,  5.5]])
```

```
In [17]: res = y -y_hat1  
res
```

```
Out[17]: array([[ 0.04, -0.05, -1.92,  1.31, -0.17, -0.04,  1.24, -0.74,  1.84,  
-1.68,  0.18]])
```

```
In [18]: plt.scatter(x, res)  
plt.show()
```



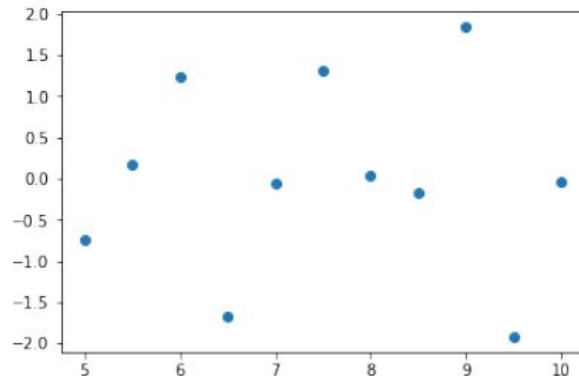
```
In [19]: import scipy.stats as stats
```

```
In [21]: stats.shapiro(res)
```

```
Out[21]: ShapiroResult(statistic=0.9420998096466064, pvalue=0.545490026473999)
```

```
In [ ]: # гомоскедастичность
```

```
In [22]: plt.scatter(y_hat, res)  
plt.show()
```





Оценить статистическую значимость полученной модели линейной регрессии



Какой должна быть численность групп, чтобы с вероятностью 90% обнаруживать снижение летальности с 90 до 30%. Уровень значимости $\alpha = 0,05$. При решении вам пригодятся табличные значения стандартного нормального распределения

Используемая литература для подготовки задач к семинару:

Гусак А.А. «Теория вероятностей»

Гланц С. «Медико-биологическая статистика



Из 10 билетов выигрышными являются 2. Чему равна вероятность того, что среди взятых наудачу 5 билетов два выигрышных?

Используемая литература для подготовки задач к семинару:

Гусак А.А. «Теория вероятностей»

Гланц С. «Медико-биологическая статистика



Повторение изученного

Число k_0 , которому при заданном n соответствует максимальная биномиальная вероятность $P_n(k_0)$, называется наивероятнейшим числом появления события A . При заданных n и p это число определяется неравенствами

$$np - q \leq k_0 \leq np + p$$

Если число $np + p$ не является целым числом, то k_0 равно целой части этого значения

Если $np + p$ – целое число, то наивероятнейшее число исходов k_0 принимает два значения $np - q$ и $np + p$

Вероятность рождения мальчиков 0.515. Найти наивероятнейшее число девочек из 600 новорожденных.

Какова вероятность наступления события B в каждом отдельном испытании, если наивероятнейшее число наступления события B в 120 испытаниях составило 32.