

# 1. Проблема

## Задача 5

### Регрессия

По экспериментальным данным ( $n = 30$ ) построено уравнение регрессии:  $y = 3,1 + 3,5x$

$$S^2_{\text{общ}} = 2,4; r_{xy} = 0,8; S^2_{\text{ост}} = 0,6; t_{0,95}(29) = 2,8; F_{0,95}(29) = 14,5$$

Что можно сказать по полученным данным?

## 2. Объяснение обозначений

$S^2_{\text{общ}} = ESS$  - Общая сумма квадратов отклонений. Характеризует величину разброса значений зависимой переменной. Этот разброс может быть вызван, с одной стороны, изменениями входных факторов, а с другой — случайными воздействиями или неучтенными в модели факторами. Если неучтенных факторов нет и случайные воздействия отсутствуют, то все изменения отклика должны объясняться моделью.

$S^2_{\text{ост}} = TSS$  - Остаточная сумма квадратов, называемая иногда суммой квадратов остатков.

$r_{xy}$  - Коэффициент корреляции Пирсона (коэффициент связи переменных)

$t_{0,95}(29)$  - Значение t-критерия Стьюдента для уровня значимости  $\alpha = 0.05$  и 30 наблюдений

$F_{0,95}(29)$  - Значение F-критерия Фишера при уровне значимости  $\alpha = 0.05$ , 30 наблюдений и 1 степени свободы

## 3. Анализ

3.1. Анализ, исходя из данных значений критериев и коэффициента корреляции Пирсона.

$R^2 = r_{xy}^2 = 0.8^2 = 0.64$  - из условия, что больше 0.5 и меньше 0.8 показывает, что регрессионную модель можно считать приемлемой.

$t_{0.95}(29) = 2.8 > 2.05 \approx t_{0.95, \text{кр}}(29)$  говорит нам о значимости коэффициента регрессии с вероятностью 95% при построении уравнения регрессии по 30 наблюдениям.

$F_{0.95}(29) = 14.5 > 4.18 \approx F_{0.95, \text{кр}}(29)$  говорит нам о том, что регрессионная модель хорошо объясняет общую дисперсию зависимой переменной с вероятностью 95%. Т.е. объясненная дисперсия существенно больше, чем необъясненная.

### 3.2. Анализ, исходя только из данных $S^2_{\text{общ}}$ и $S^2_{\text{ост}}$

Посчитаем объясненную сумму квадратов:

$$RSS = TSS - ESS = 2.4 - 0.6 = 1.8$$

Получаем:

TSS	RSS	ESS
2.4	1.8	0.6

Теперь посчитаем  $R^2$ :

$$R^2 = \frac{RSS}{TSS} = \frac{1.8}{2.4} = \frac{3}{4} = 0.75$$

В моем пересчете получается, что  $R^2$  ближе к 0.8. А значит, что модель на самом деле еще более приемлемая.

Посчитаем  $F$ -критерий Фишера:

$$F = \frac{S^2_{\hat{y}}}{S^2} = \frac{RSS}{ESS} \times (N - 2) = (N - 2) \frac{R^2}{1 - R^2} = 28 \times 3 = 84.$$

Тоже не совпадает. И т.к. значение еще выше, мы можем говорить о том, что объясненная дисперсия еще больше, чем предполагалось в предыдущем пункте, относительно необъясненной

Соответственно  $t$ -критерий Стьюдента:

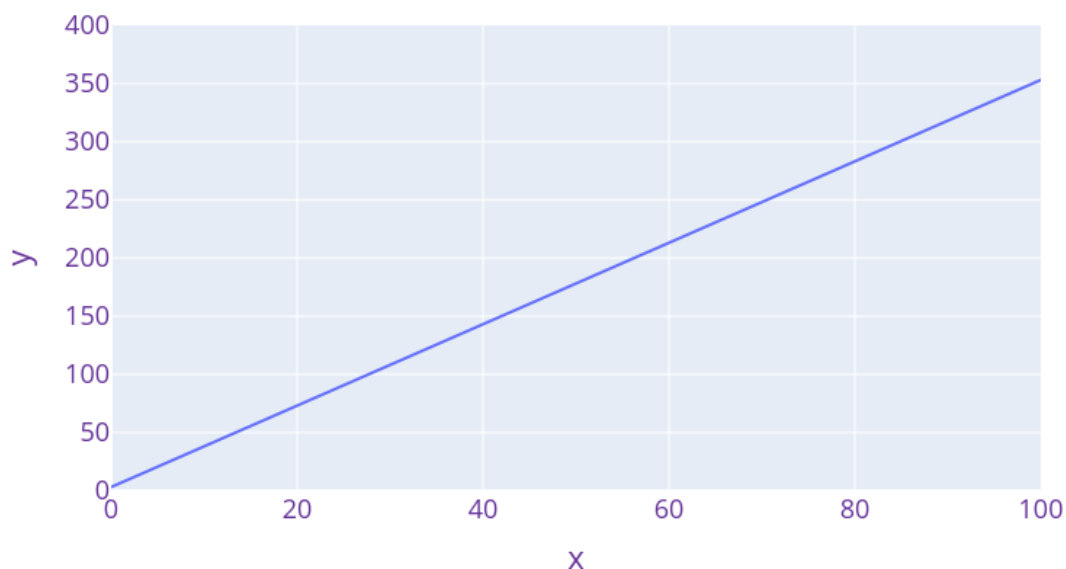
$$t = \sqrt{F} \approx 9.2$$

Также больше, и так же говорит о еще большей значимости коэффициента регрессии.

### 3.3 График

Наконец, построим график для красоты. Просто так.

## График полученной регрессии



Код для получения графика:

```
In [ ]: import plotly.graph_objects as go
import numpy as np

x = np.arange(-1000, 1000, 1999)

fig = go.Figure(data=go.Scatter(x=x, y=3.1 + 3.5*x))
fig.update_layout(
    width = 800,
    height = 500
)
fig.update_layout(
    title="График полученной регрессии",
    xaxis_title="x",
    yaxis_title="y",
    legend_title="Legend Title",
    font=dict(
        family="Open Sans",
        size=18,
        color="RebeccaPurple"
    )
)
fig.update_xaxes(range=[0, 100])
fig.update_yaxes(range=[0, 400])
fig.show()
```