

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
Направление: 01.03.02 Прикладная математика и информатика  
ООП: Прикладная математика, фундаментальная информатика и  
программирование и

**ОТЧЕТ ПО УЧЕБНОЙ ПРАКТИКЕ**  
**(НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ)**

**Тема задания:** Регрессионный анализ совершения самоубийств по регионам России.

**Выполнил:** Сыдыгалиева Бегаим Нурбековна, студент группы 21.Б03

**Научный Руководитель:** Буре Владимир Мансурович

Санкт-Петербург  
2024

## СОДЕРЖАНИЕ

<b>ПОСТАНОВКА ЗАДАЧИ.....</b>	<b>3</b>
<b>ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ.....</b>	<b>4</b>
<b>СТАТИСТИЧЕСКИЕ ДАННЫЕ.....</b>	<b>5</b>
<b>ОБРАБОТКА И АНАЛИЗ ДАННЫХ.....</b>	<b>6</b>
Построение регрессионной модели для 2019 года.....	6
Проверка значимости уравнения в целом.....	10
Анализ значимости коэффициентов регрессии:.....	12
Анализ остатков модели:.....	12
Построение регрессионной модели на значимых коэффициентах.....	14
Построение регрессионной модели для 2021 года.....	16
Проверка значимости модели F-критерием:.....	17
Анализ значимости коэффициентов регрессии:.....	18
Анализ остатков:.....	19
Построение регрессионной модели на значимых коэффициентах.....	19
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>20</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....</b>	<b>21</b>

## ПОСТАНОВКА ЗАДАЧИ

- Тема: Регрессионный анализ совершения самоубийств по регионам России.
- Объект исследования: Социально-экономические факторы, воздействующие на состояние человека
- Цель работы: Выявление статистически значимых факторов

В процессе работы были построены линейные регрессионные модели, проанализированы характеристики и проверены гипотезы с помощью языка программирования `python` с использованием библиотек для анализа данных и машинного обучения: `sklearn`, `statsmodels`. [4]

В основе суицидального поведения лежит множество причин. Факторы, приводящие к самоубийству, сложны и взаимосвязаны. По данным ВОЗ к таким факторам относят возрастные особенности, половые различия, психические заболевания, биологические, социальные факторы окружающей среды и факторы, связанные с историей жизни индивида. В структуру социальных факторов и факторов окружающей среды входят: доступность средств самоубийства, место жительства человека, трудовая занятость, иммиграционные процессы, принадлежность определенной религии и экономические условия.

Оценка и выявление причин “смертей от отчаяния” имеет большое значение для обеспечения устойчивого развития общества

Для анализа влияния социально-экономических факторов на девиантное поведения граждан можно использовать метод множественной линейной регрессии.

# ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ

**Линейный регрессионный анализ (Linear regression)** - Это метод аппроксимации зависимостей между предикторными переменными (или регрессорами) и критериальными на основе линейной модели. [5]

Данный метод позволяет предсказывать значения зависимой переменной  $y$  по значениям независимой переменной  $x$

В качестве функции потерь (меры количества ошибок, которые наша линейная регрессия делает на наборе данных будет рассмотрена **Среднеквадратичная ошибка (MSE)**.

Вычисление MSE: все отдельные остатки регрессии возводятся в квадрат, суммируются, сумма делится на общее число ошибок

В случае множественной линейной регрессии модель будет описываться гиперплоскостью. Коэффициенты уравнения множественной линейной регрессии подбираются так, чтобы минимизировать сумму квадратов отклонения реальных точек данных от этой гиперплоскости.

# СТАТИСТИЧЕСКИЕ ДАННЫЕ

**Источник:** Данные собраны из Единой межведомственной информационно-статистической системы (ЕМИСС) и из Федеральной службы государственной статистики.

## Используемые инструменты:

- Для предобработки и анализа данных были использованы библиотеки *numpy* и *pandas*
- Для построения модели линейной регрессии: *sklearn* (библиотека Python для машинного обучения. Она включает все алгоритмы и инструменты, которые нужны для задач классификации, регрессии и кластеризации. Она также включает все методы оценки производительности модели машинного обучения)

## Исследуемый датасет:

Целевой столбец:

- 'num\_suicide' - количество совершенных самоубийств на 10000 тысяч человек

Признаки:

- 'alco' - Розничные продажи алкогольной продукции (в литрах этанола)
- 'cash\_income' - Денежные доходы (в среднем на душу)
- 'mental\_disorder' - Число зарегистрированных больных с диагнозом психического расстройства и расстройствами поведения
- 'num\_crimes' - Количество совершенных преступлений
- 'poverty\_level' - Уровень бедности
- 'employee\_ratio' - Отношение числа занятых в экономике региона к численности населения региона в трудоспособном возрасте
- 'incomes\_below\_the\_poverty' - Количество людей за чертой бедности

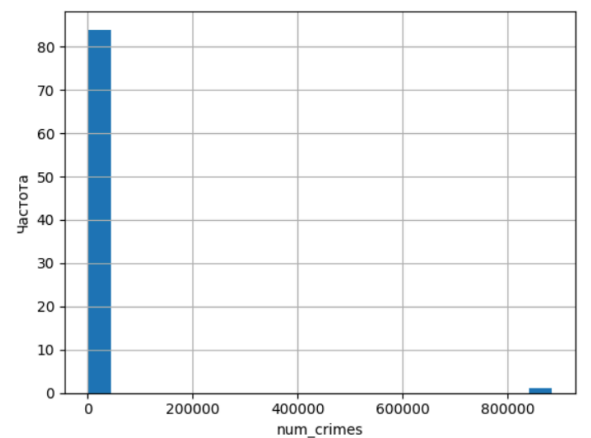
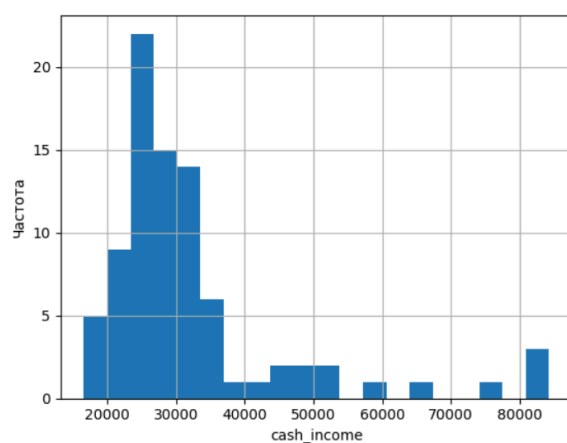
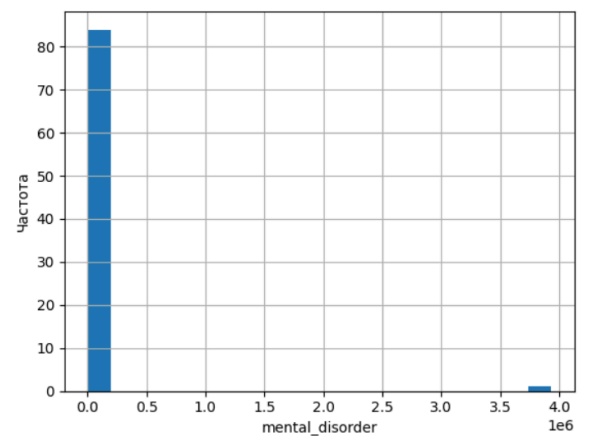
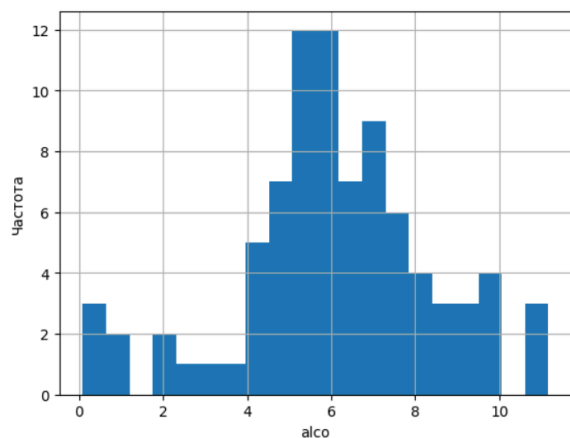
	district	num_suicide	alco	cash_income	incomes_below_the_poverty_line	mental_disorder	num_crimes	poverty_level	rape_num	employee_ratio
0	Российская Федерация	11.7	6.0336	10.477457	2.509599	15.185182	13.692960	2.509599	3177	4.461300
1	Белгородская область	11.0	4.9824	10.385852	2.054124	10.518619	8.772765	2.054124	5	4.476200
2	Брянская область	11.3	5.5272	10.254919	2.624669	10.020426	8.860499	2.624669	17	4.347694
3	Владимирская область	14.3	7.6111	10.142150	2.533697	10.690034	8.870663	2.533697	18	4.458988
4	Воронежская область	11.1	4.9702	10.374178	2.186051	11.093873	9.343734	2.186051	38	4.454347
...	...	...	...	...	...	...	...	...	...	...
80	Амурская область	30.2	7.2232	10.413433	2.753661	10.165583	8.978787	2.753661	33	4.457830
81	Магаданская область	16.3	10.6641	11.087773	2.230014	8.191186	7.349231	2.230014	5	4.681205
82	Сахалинская область	0.4	11.1699	10.985564	2.104134	9.416948	8.344980	2.104134	3	4.601162
83	Еврейская автономная область	32.7	8.4527	10.188742	3.173878	8.709630	7.221836	3.173878	6	4.265493
84	Чукотский автономный округ	18.0	8.3082	11.331224	2.140066	7.935587	6.331502	2.140066	3	4.675629

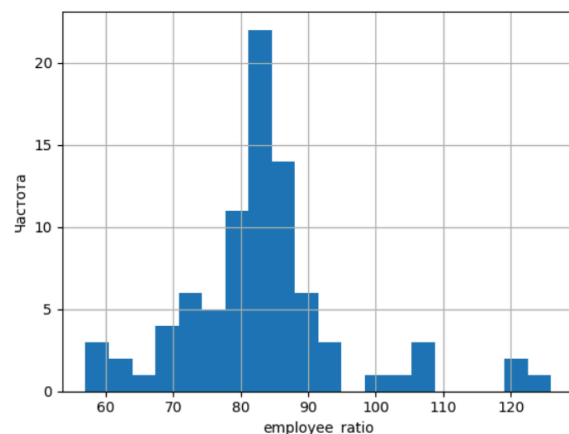
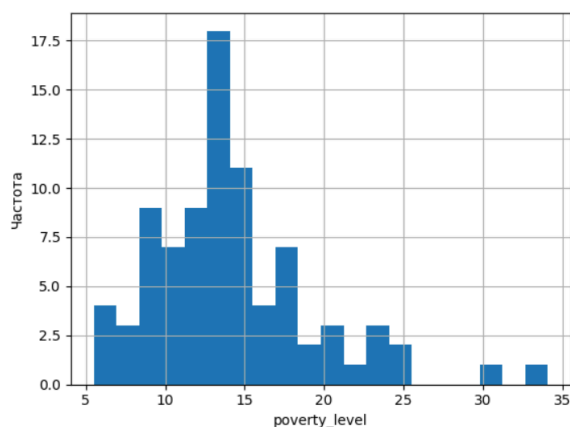
85 rows × 10 columns

## ОБРАБОТКА И АНАЛИЗ ДАННЫХ

### Построение регрессионной модели для 2019 года

Визуализация разброса значений на гистограммах



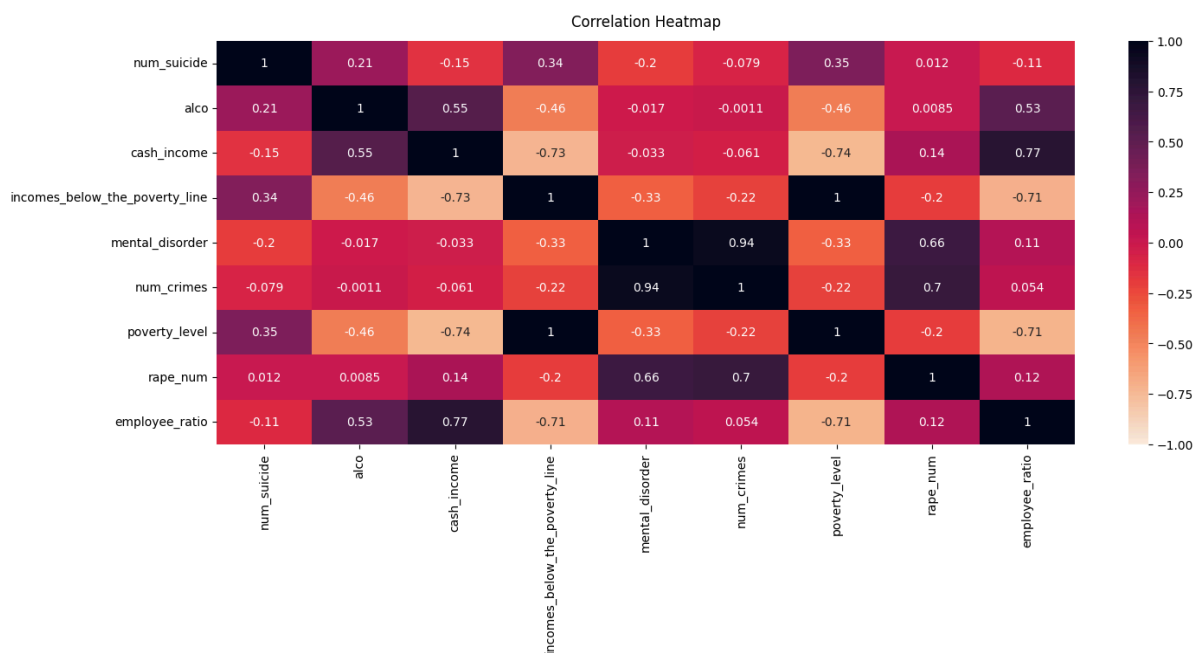


При расчете коэффициентов асимметрии значений у данных выявлена правосторонняя асимметрия у последних пяти признаков:

```
data['num_crimes'].skew()
✓ 0.0s
9.123644847909572
```

данные рассматриваемых признаков имеют логнормальное распределение

## Корреляционная матрица:

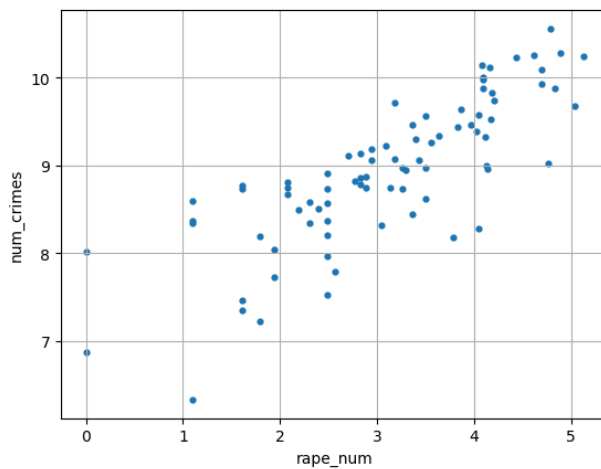


Для проверки связи между факторами рассматривается коэффициент корреляции r-Пирсона и графики рассеяния

```
x = data['rape_num']
print(x.corr(data['num_crimes']))
```

✓ 0.0s

... 0.7011446801844059



Проверка существования корреляции проводится двусторонним t-тестом

```
def t_statistic(xs, ys):
    '''Вычисление t-статистики'''
    r = xs.corr(ys)
    df = xs.count() - 2
    return r * np.sqrt(df / 1 - r ** 2)

def t_test(xs, ys):
    '''Выполнение двустороннего t-теста'''
    t_value = t_statistic(xs, ys)
    df = xs.count() - 2
    p = 2 * stats.t.sf(t_value, df) # функция выживания
    return {'t-значение':t_value, 'p-значение':p}
```

✓ 0.0s

```
t_test(data['incomes_below_the_poverty_line'],data['poverty_level'])
```

✓ 0.0s

```
{'t-значение': 8.969080126971338, 'p-значение': 8.292094798620893e-14}
```

При отклонении нулевой гипотезы были удалены факторы, имеющие меньшую связь с целевой переменной



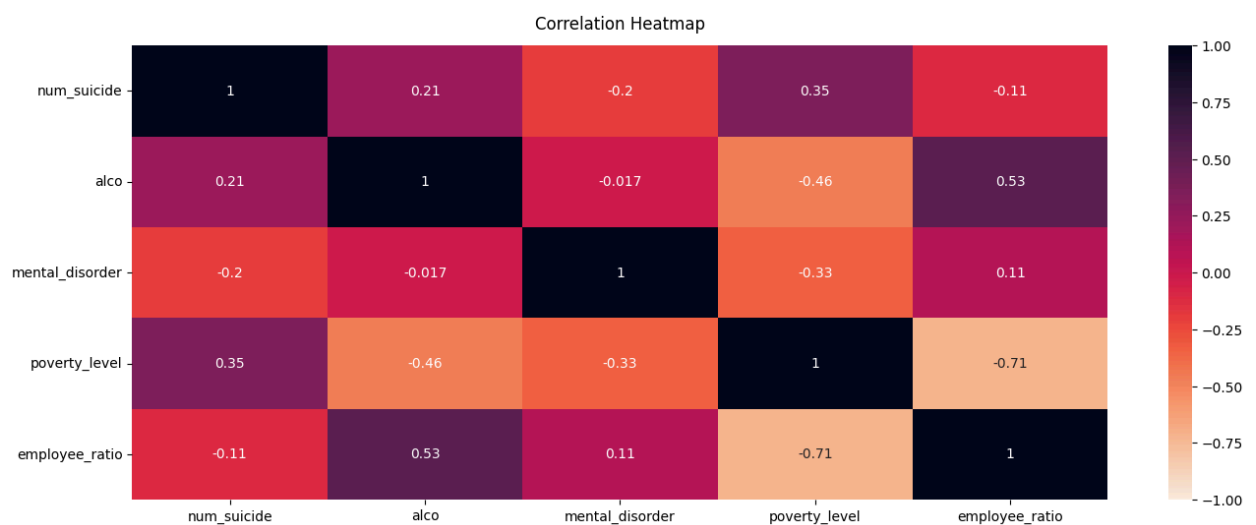
```
t_test(data['rape_num'],data['num_crimes'])
✓ 0.0s
{'t-значение': 6.330074436489531, 'p-значение': 1.2331414173069111e-08}
```

```
t_test(data['incomes_below_the_poverty_line'],data['poverty_level'])
✓ 0.0s
{'t-значение': 8.969080126971338, 'p-значение': 8.292094798620893e-14}
```

```
t_test(data['mental_disorder'],data['num_crimes'])
✓ 0.0s
{'t-значение': 8.427291235284113, 'p-значение': 9.917283607862887e-13}
```

```
t_test(data['poverty_level'],data['employee_ratio'])
✓ 0.0s
{'t-значение': -6.445230992257486, 'p-значение': 1.999999925387852}
```

Линейная множественная регрессионная модель построена по оставшимся четырем признакам: 'alco', 'mental\_disorder', 'poverty\_level', 'employee\_ratio'



### Коэффициенты модели:

alco 1.68

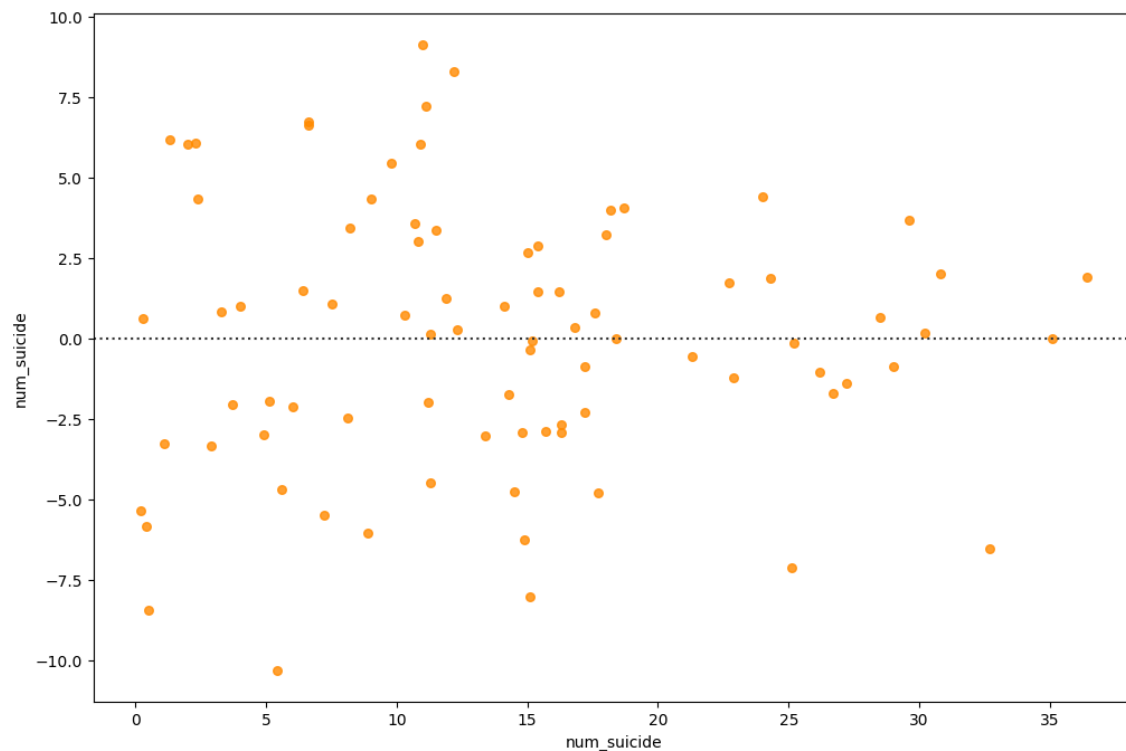
mental\_disorder 0.06

incomes\_below\_the\_poverty\_line 15.72

employee\_ratio 5.91

Intercept: -63.74

## График остатков модели:



MSE: 55.553

Величина средней ошибки аппроксимации: 281.45722302798134

Множественный коэффициент детерминации R2: 0.2954427464594547

## Проверка значимости уравнения в целом

Применим F-критерий:

Значение F-критерия =  $3.945880658551282e-05$

Нулевая гипотеза отвергается, модель признается статистически значимой на 5% уровне

	notation	coef_value (R)	coef_value_squared (R2)	p_level	a_level	F_calc	df1	df2	F_table	F_calc >= F_table	a_calc	a_calc <= a_level	significance_check
Coef. of determination	R2	0.543546	0.295443	0.95	0.05	8.281789	4	79	2.487366	True	0.000012	True	significance

Общие сведения построенной модели (пакет Statmodels)

OLS Regression Results						
=====						
Dep. Variable:	num_suicide	R-squared:	0.295			
Model:	OLS	Adj. R-squared:	0.260			
Method:	Least Squares	F-statistic:	8.282			
Date:	Sun, 26 May 2024	Prob (F-statistic):	1.25e-05			
Time:	01:35:53	Log-Likelihood:	-287.92			
No. Observations:	84	AIC:	585.8			
Df Residuals:	79	BIC:	598.0			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
Intercept	-63.7441	48.890	-1.304	0.196	-161.058	33.570
alco	1.6821	0.430	3.910	0.000	0.826	2.538
mental_disorder	0.0590	0.939	0.063	0.950	-1.810	1.928
poverty_level	15.7151	3.743	4.198	0.000	8.264	23.166
employee_ratio	5.9055	9.003	0.656	0.514	-12.015	23.826
=====						
Omnibus:	0.192	Durbin-Watson:	1.864			
Prob(Omnibus):	0.909	Jarque-Bera (JB):	0.296			
Skew:	-0.107	Prob(JB):	0.863			
Kurtosis:	2.802	Cond. No.	780.			
=====						

регрессионная модель объясняет 29.5% вариации переменной Y.

расчетный уровень значимости Prob (F-statistic) = 1.25e-05

так как значение Prob (F-statistic) < 0.05, то нулевая гипотеза R-squared = 0 НЕ ПРИНИМАЕТСЯ, т.е. коэффициент детерминации ЗНАЧИМ

Проверка адекватности модели:

	SST	SSE	dft	dfe	MST	MSE	p_level	F_calc	F_table	F_calc >= F_table	a_calc	adequacy_check
linear_ols	6623.298095	4666.492715	83	79	79.798772	59.069528	0.95	1.35093	1.445939	False	0.08964	adequacy

## Анализ значимости коэффициентов регрессии:

Проверка гипотезы о значимости и доверительные интервалы

	notation	coef_value	std_err	p_level	a_level	t_calc	df	t_table	t_calc >= t_table	a_calc	a_calc <= a_level	significance_check
Intercept	Intercept	-63.744056	48.890262	0.95	0.05	1.303819	79	1.99045	False	0.196082	False	not significance
alco	alco	1.682105	0.430248	0.95	0.05	3.909618	79	1.99045	True	0.000194	True	significance
mental_disorder	mental_disorder	0.059005	0.938811	0.95	0.05	0.062851	79	1.99045	False	0.950044	False	not significance
poverty_level	poverty_level	15.715145	3.743382	0.95	0.05	4.198114	79	1.99045	True	0.000070	True	significance
employee_ratio	employee_ratio	5.905464	9.003320	0.95	0.05	0.655921	79	1.99045	False	0.513781	False	not significance

conf_int_low	conf_int_high
-161.057688	33.569577
0.825718	2.538492
-1.809652	1.927662
8.264130	23.166160
-12.015196	23.826124

Расчетный уровень значимости  $P > |t|$  не превышает 0.05 у коэффициента при alco и poverty\_level - оба коэффициента регрессии значимы, остальные - нет

## Анализ остатков модели:

Skew = -0.107 и Kurtosis = 2.802 - показатели асимметрии и эксцесса остатков

p-value коэффициентов:

```
result_linear_ols.pvalues =
```

```
Intercept          0.196082
alco                0.000194
mental_disorder    0.950044
poverty_level      0.000070
employee_ratio     0.513781
```

Доверительные интервалы:

```

                                0          1
Intercept          -161.057688  33.569577
alco                0.825718    2.538492
mental_disorder    -1.809652    1.927662
poverty_level      8.264130    23.166160
employee_ratio     -12.015196  23.826124
```

Obs	Dep Var	Predicted	Std Error	Mean ci	Mean ci	Predict ci	Predict ci	Residual	Std Error	Student	Cook's
	Population	Value	Mean Predict	95% low	95% upp	95% low	95% upp		Residual	Residual	D
1	11.000	3.972	2.055	-0.117	8.062	-11.863	19.808	7.028	7.406	0.949	0.014
2	11.300	13.067	1.039	10.998	15.135	-2.370	28.504	-1.767	7.615	-0.232	0.000
3	14.300	15.839	1.078	13.693	17.986	0.391	31.287	-1.539	7.610	-0.202	0.000
4	11.100	5.930	1.655	2.637	9.223	-9.718	21.578	5.170	7.505	0.689	0.005
5	17.200	15.819	0.956	13.917	17.722	0.404	31.235	1.381	7.626	0.181	0.000
6	11.900	12.144	1.101	9.954	14.335	-3.310	27.598	-0.244	7.606	-0.032	0.000
7	18.400	15.314	1.187	12.952	17.677	-0.165	30.794	3.086	7.593	0.406	0.001
8	10.900	7.042	1.553	3.952	10.133	-8.565	22.649	3.858	7.527	0.513	0.002
9	2.400	6.229	1.685	2.876	9.582	-9.432	21.890	-3.829	7.499	-0.511	0.003
10	11.500	9.892	3.022	3.877	15.907	-6.546	26.330	1.608	7.067	0.228	0.002
11	10.300	12.170	1.178	9.825	14.515	-3.306	27.647	-1.870	7.595	-0.246	0.000
12	3.700	13.011	0.882	11.255	14.768	-2.387	28.410	-9.311	7.635	-1.220	0.004
13	8.900	18.557	1.203	16.162	20.952	3.072	34.041	-9.657	7.591	-1.272	0.008
14	9.800	7.308	1.570	4.183	10.433	-8.306	22.922	2.492	7.524	0.331	0.001
15	16.800	14.500	1.065	12.380	16.620	-0.944	29.944	2.300	7.612	0.302	0.000
16	10.700	9.448	1.176	7.108	11.788	-6.028	24.924	1.252	7.595	0.165	0.000
17	5.100	13.332	1.132	11.078	15.585	-2.131	28.795	-8.232	7.602	-1.083	0.005
18	2.300	4.468	3.134	-1.770	10.706	-12.053	20.989	-2.168	7.018	-0.309	0.004
19	15.100	22.345	1.959	18.445	26.244	6.557	38.132	-7.245	7.432	-0.975	0.013
20	25.100	24.418	2.298	19.844	28.992	8.451	40.385	0.682	7.334	0.093	0.000
21	29.600	14.931	3.949	7.072	22.791	-2.267	32.130	14.669	6.594	2.225	0.355
83	32.700	26.056	2.462	21.155	30.957	9.992	42.120	6.644	7.281	0.913	0.019
84	18.000	11.943	3.019	5.934	17.951	-4.493	28.378	6.057	7.068	0.857	0.027

## Проверка гипотезы о нормальном распределении остатков

Для исследования применяем критерий Жака-Бера:

SignificanceResult:

(statistic=0.2955573116880474, pvalue=0.8626220301574741)

```
residuals = y - y_pred
std = np.std(residuals)* np.sqrt((n - 1)/ n)
K = 1/n * sum(((residuals - residuals.mean()) / std)**4)
S = 1/n * sum(((residuals - residuals.mean()) / std)**3)
zh_b_statistic = n/6 * (S**2 + ((K - 3)**2)/4)
hi_critical = 5.9916
if (zh_b_statistic > hi_critical):
    print("Нулевая гипотеза о нормальности распределения остатков ряда отвергается на уровне значимости 0.05")
else: print("Нулевая гипотеза о нормальности распределения остатков ряда не отвергается")
```

Нулевая гипотеза о нормальности распределения остатков ряда не отвергается

### Анализ независимости случайных отклонений :

Критерий Дарбина-Уотсона

d = 1.8789636574595088

гипотеза о независимости случайных отклонений не отвергается

### Построение регрессионной модели на значимых коэффициентах

OLS Regression Results						
Dep. Variable:	num_suicide	R-squared:	0.292			
Model:	OLS	Adj. R-squared:	0.274			
Method:	Least Squares	F-statistic:	16.67			
Date:	Sun, 26 May 2024	Prob (F-statistic):	8.64e-07			
Time:	04:06:47	Log-Likelihood:	-288.15			
No. Observations:	84	AIC:	582.3			
Df Residuals:	81	BIC:	589.6			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-33.6943	8.379	-4.021	0.000	-50.367	-17.022
alco	1.7678	0.399	4.434	0.000	0.974	2.561
poverty_level	14.2105	2.673	5.317	0.000	8.893	19.528
Omnibus:	0.111	Durbin-Watson:		1.862		
Prob(Omnibus):	0.946	Jarque-Bera (JB):		0.190		
Skew:	-0.083	Prob(JB):		0.909		
Kurtosis:	2.836	Cond. No.		74.3		

result\_linear\_ols.pvalues =

Intercept 1.292115e-04

alco 2.884002e-05

poverty\_level 9.118108e-07

### Доверительный интервал

	0	1
Intercept	-50.366587	-17.022110
alco	0.974441	2.561062
poverty_level	8.892995	19.528019

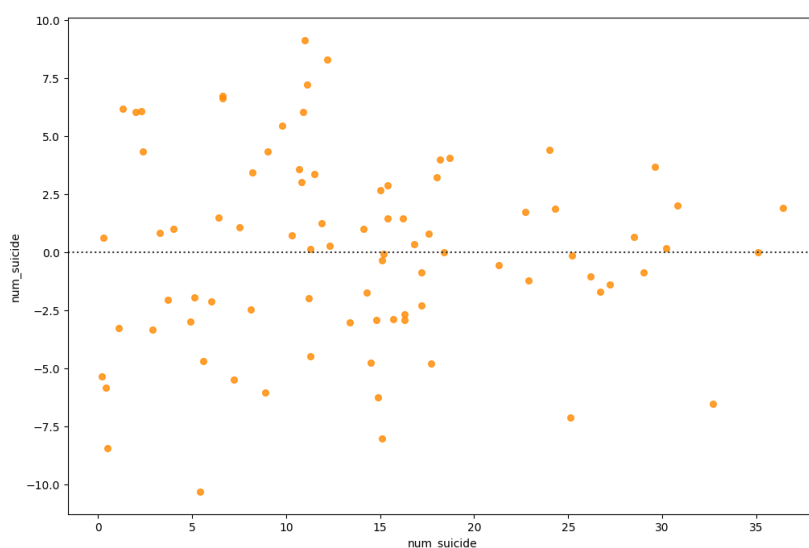
$R^2 = 0.292$ , незначительно меньше значения со всеми коэффициентами

### Анализ остатков:

#### Критерий Жака-Бера

Значение статистики: 0.22358939412175294

Нулевая гипотеза о нормальности распределения остатков ряда не отвергается



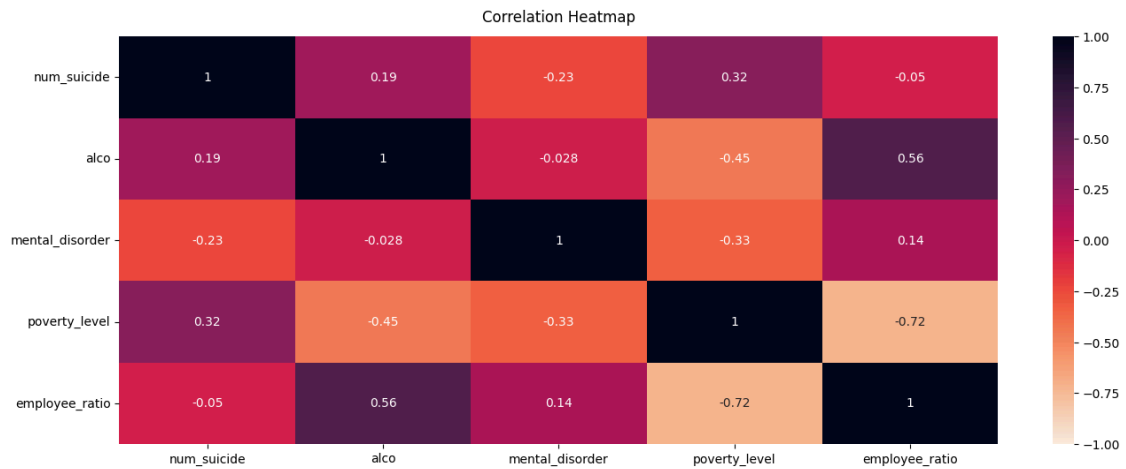
#### Критерий Дарбина-Уотсона

$d = 1.8789636574595088$

гипотеза о независимости случайных отклонений не отвергается

## Построение регрессионной модели для 2021 года

Аналогично исследуем факторы, выявляя несильно коррелируемые



### Коэффициенты:

alco 1.31

mental\_disorder -0.55

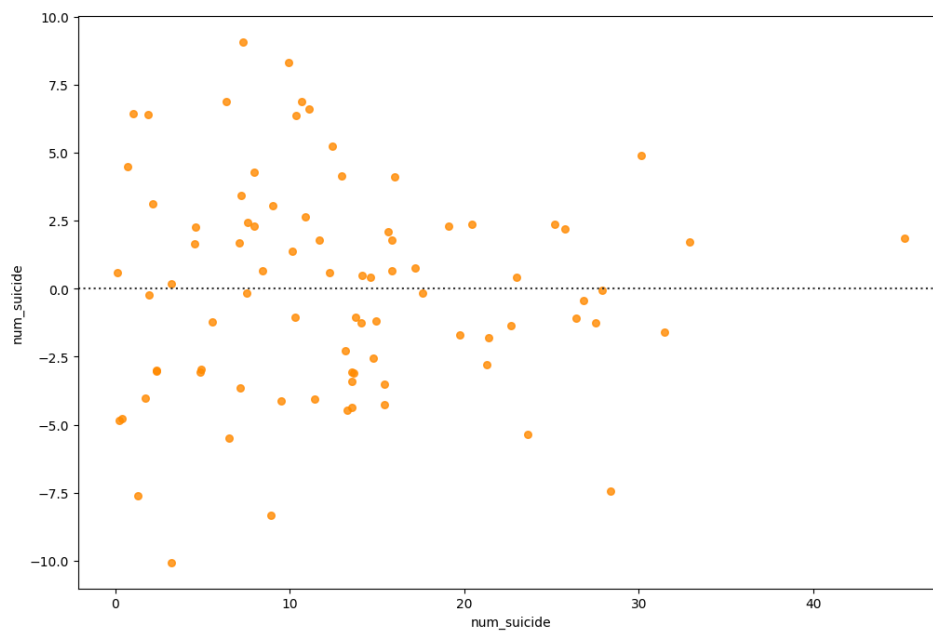
incomes\_below\_the\_poverty\_line 14.25

employee\_ratio 11.01

Intercept: -74.04

### График остатков





Ошибка аппроксимации: 302.6840681979601

MSE: 59.649

Множественный коэффициент детерминации R2: 0.2527374521860234

### Проверка значимости модели F-критерием:

F-critical = 0.00031632028131289847

rejected, модель признается статистически значимой

	notation	coef_value (R)	coef_value_squared (R2)	p_level	a_level	F_calc	df1	df2	F_table	F_calc >= F_table	a_calc	a_calc <= a_level	significance_check
Coef. of determination	R2	0.50273	0.252737	0.95	0.05	6.6798	4	79	2.487366	True	0.00011	True	significance

OLS Regression Results						
Dep. Variable:	num_suicide	R-squared:	0.253			
Model:	OLS	Adj. R-squared:	0.215			
Method:	Least Squares	F-statistic:	6.680			
Date:	Sun, 26 May 2024	Prob (F-statistic):	0.000110			
Time:	03:43:30	Log-Likelihood:	-290.91			
No. Observations:	84	AIC:	591.8			
Df Residuals:	79	BIC:	604.0			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-74.0422	48.235	-1.535	0.129	-170.052	21.968
alco	1.3114	0.461	2.843	0.006	0.393	2.229
mental_disorder	-0.5484	0.969	-0.566	0.573	-2.478	1.381
poverty_level	14.2511	3.648	3.907	0.000	6.991	21.511
employee_ratio	11.0141	9.201	1.197	0.235	-7.301	29.329
Omnibus:	3.922	Durbin-Watson:	1.785			
Prob(Omnibus):	0.141	Jarque-Bera (JB):	3.337			
Skew:	0.326	Prob(JB):	0.189			
Kurtosis:	3.726	Cond. No.	750.			

Регрессионная модель объясняет 25.3% вариации переменной num\_suicide

Проверка адекватности:

	SST	SSE	dft	dfe	MST	MSE	p_level	F_calc	F_table	F_calc >= F_table	a_calc	adequacy_check
linear_ols	6705.172689	5010.524427	83	79	80.785213	63.42436	0.95	1.273725	1.445939	False	0.139882	adequacy

## Анализ значимости коэффициентов регрессии:

result\_linear\_ols.pvalues =

Intercept 0.128774

alco 0.005686

mental\_disorder 0.573185

poverty\_level 0.000196

employee\_ratio 0.234881

Доверительные интервалы:

Intercept -170.052202 21.967809

alco 0.393255 2.229466

mental\_disorder -2.478087 1.381204

poverty\_level 6.990867 21.511258

employee\_ratio -7.300665 29.328908

	notation	coef_value	std_err	p_level	a_level	t_calc	df	t_table	t_calc >= t_table	a_calc	a_calc <= a_level	significance_check
Intercept	Intercept	-74.042196	48.235321	0.95	0.05	1.535020	79	1.99045	False	0.128774	False	not significance
alco	alco	1.311361	0.461255	0.95	0.05	2.843026	79	1.99045	True	0.005686	True	significance
mental_disorder	mental_disorder	-0.548441	0.969452	0.95	0.05	0.565723	79	1.99045	False	0.573185	False	not significance
poverty_level	poverty_level	14.251063	3.647514	0.95	0.05	3.907062	79	1.99045	True	0.000196	True	significance
employee_ratio	employee_ratio	11.014121	9.201329	0.95	0.05	1.197014	79	1.99045	False	0.234881	False	not significance

Значимы коэффициенты: alco и poverty\_level

## Анализ остатков:

**Критерий Жака-Бера** - проверка гипотезы о нормальности распределения остатков

Значение статистики = 3.8787816804764272

Нулевая гипотеза о нормальности распределения остатков ряда не отвергается

**Критерий Дарбина-Уотсона**

d = 1.865232128745912

гипотеза о независимости случайных отклонений не отвергается

## Построение регрессионной модели на значимых коэффициентах

OLS Regression Results						
=====						
Dep. Variable:	num_suicide	R-squared:	0.235			
Model:	OLS	Adj. R-squared:	0.216			
Method:	Least Squares	F-statistic:	12.46			
Date:	Sun, 26 May 2024	Prob (F-statistic):	1.92e-05			
Time:	13:35:21	Log-Likelihood:	-291.88			
No. Observations:	84	AIC:	589.8			
Df Residuals:	81	BIC:	597.0			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
Intercept	-28.0151	8.286	-3.381	0.001	-44.501	-11.529
alco	1.5756	0.418	3.771	0.000	0.744	2.407
poverty_level	12.2905	2.666	4.610	0.000	6.985	17.596
=====						
Omnibus:	7.240	Durbin-Watson:	1.764			
Prob(Omnibus):	0.027	Jarque-Bera (JB):	7.185			
Skew:	0.517	Prob(JB):	0.0275			
Kurtosis:	3.992	Cond. No.	72.6			
=====						

**p-values коэффициентов**

result\_linear\_ols.pvalues =

Intercept	0.001114
alco	0.000308
poverty_level	0.000015

#### Доверительные интервалы:

	0	1
Intercept	-44.501475	-11.528693
alco	0.744155	2.407006
poverty_level	6.985432	17.595614

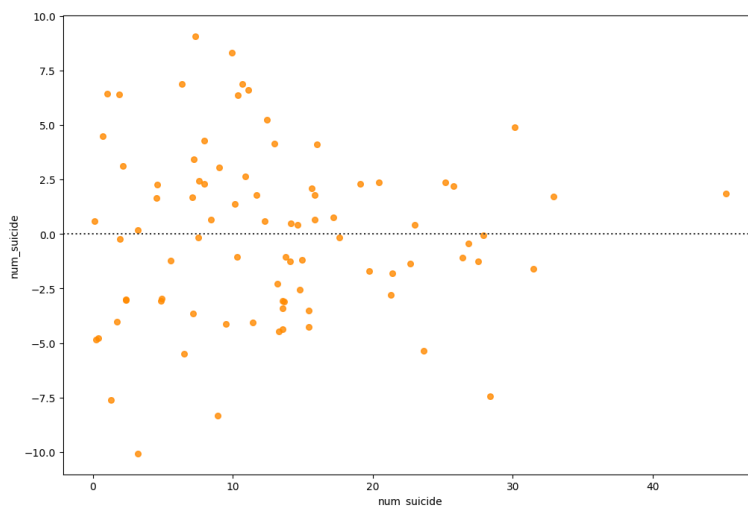
#### Анализ остатков:

##### Критерий Жака-Бера

Значение статистики: 3.8787816804764272

Нулевая гипотеза о нормальности распределения остатков ряда не отвергается

#### График остатков:



##### Критерий Дарбина-Уотсона

$d = 1.865232128745912$

гипотеза о независимости случайных отклонений не отвергается

## ЗАКЛЮЧЕНИЕ

В данной работе были выявлены статистически значимые коэффициенты: количество алкоголя потребляемое в области на 10000 т. человек и уровень бедности. Также проведено сравнение результатов, получаемых при построении моделей для различных исследуемых годов - значимые коэффициенты сохранялись

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Буре В. М., Парилина Е. М. Методы прикладной статистики в R и Excel. СПб.: Лань, 2013.
- [2] Носко В. П. Эконометрика. М.: Издат. дом «Дело», 2011. 672 с.
- [3] Федеральная служба государственной статистики. [Электронный ресурс]: URL: <https://rosstat.gov.ru/>
- [4] Единая межведомственная информационно-статистическая система (ЕМИСС). [Электронный ресурс]: URL: <https://fedstat.ru/>
- [5] В. Ф. Войцех. Динамика суицидов в регионах России. ФГУ “Московский НИИ психиатрии Росздрава”. [Электронный ресурс]: URL: [https://mniip.serbssky.ru/assets/uploads/2017/12/scp\\_2008-1-14.pdf](https://mniip.serbssky.ru/assets/uploads/2017/12/scp_2008-1-14.pdf)
- [6] Ю. Н. Тюрин А. А. Макаров Теория вероятностей и математическая статистика «Московские учебники» Москва, 2004
- [7] Scikit-Learn. [Электронный ресурс]: URL: <https://scikit-learn.org/stable/>
- [8] Статистический анализ с использованием пакетов Python. [Электронный ресурс]: URL: <https://habr.com/ru/articles/557424/>
- [9] Учебник по машинному обучению. ШАД [Электронный ресурс]: URL: <https://education.yandex.ru/handbook/ml>
- [10] Computer Science Center : Лекции по анализу данных. [Электронный ресурс]: URL: [https://youtube.com/playlist?list=PLlb7e2G7aSpRb95\\_Wi7IZ-zA6fQjV3\\_l7&si=\\_P1as40xkkHJySW4](https://youtube.com/playlist?list=PLlb7e2G7aSpRb95_Wi7IZ-zA6fQjV3_l7&si=_P1as40xkkHJySW4)