

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра технологий программирования

ОТЧЕТ по лабораторной 1

«АДЭФС. Лабораторная работа №1»

Косинского Никиты Александровича
студента 3 курса, 8 группы

Преподаватель
Полузеров Тимофей Дмитриевич

Минск 2025

Содержание:

- 1. Набор данных**
- 2. Задание 1**
- 3. Вывод**
- 4. Задание 2**
- 5. Вывод**
- 6. Задание 3**
- 7. Вывод**
- 8. Общий вывод**

Набор данных: таблица Annual 2005-2011 со столбцами empl_num, k1...k20, year

Задание 1. Расчет описательных статистик:

1. Минимум, Максимум, Размах
2. Среднее
3. Дисперсия, Стандартное отклонение
4. Медиана
5. Квантили уровней 0.01, 0.05, 0.95, 0.99

```
sts = pd.DataFrame({
    'Минимум': df.min(),
    'Максимум': df.max(),
    'Размах': df.max() - df.min(),
    'Среднее': df.mean(),
    'Дисперсия': df.var(),
    'Стандартное отклонение': df.std(),
    'Медиана': df.median(),
    'Квантиль 0.01': df.quantile(0.01),
    'Квантиль 0.05': df.quantile(0.05),
    'Квантиль 0.95': df.quantile(0.95),
    'Квантиль 0.99': df.quantile(0.99)
})
```

Статистики:											
	Минимум	Максимум	Размах	Среднее	Дисперсия	Стандартное отклонение	Медиана	Квантиль 0.01	Квантиль 0.05	Квантиль 0.95	Квантиль 0.99
empl_num	10.000000	28650.000000	28640.000000	1220.773284	6.535214e+06	2556.406482	473.000000	36.940000	91.000000	5152.400000	12526.780000
k1	0.248322	18.020148	17.771826	2.002089	2.856275e+00	1.690052	1.473859	0.423381	0.663925	5.258034	9.002944
k2	0.000000	7.029135	7.029135	0.238018	2.680105e-01	0.517697	0.055551	0.000340	0.002725	1.128667	2.636539
k3	0.008329	11.187699	11.179370	0.825098	8.432594e-01	0.918292	0.538349	0.061415	0.124851	2.544703	4.917381
k4	-4.569874	0.935935	5.505809	0.038115	3.908187e-01	0.625155	0.148620	-2.450093	-0.977473	0.755809	0.881929
k5	0.009944	1.083702	1.073758	0.346330	3.920708e-02	0.198008	0.319908	0.037373	0.068282	0.712523	0.860138
k6	0.000000	1.000000	1.000000	0.238031	4.540896e-02	0.213094	0.185185	0.000000	0.000000	0.663608	0.859919
k7	0.000000	1.000000	1.000000	0.174212	4.984752e-02	0.223266	0.075676	0.000000	0.000000	0.674139	0.890017
k8	0.053766	0.990056	0.936290	0.655964	3.832761e-02	0.195774	0.681890	0.160895	0.292537	0.932723	0.962348
k9	0.059320	199.605839	199.546520	6.862756	6.309532e+01	7.943256	4.758963	0.455832	1.048860	19.642377	35.671475
k10	0.040014	11.834268	11.794255	0.831359	6.433757e-01	0.802107	0.617689	0.087824	0.167346	2.207673	4.030097
k11	0.098838	15.865194	15.766356	1.325512	2.891984e-01	0.537772	1.237491	0.578515	0.805104	2.051245	2.648635
k12	0.230067	10.809042	10.578975	1.351127	3.557425e-01	0.596442	1.180362	0.722242	0.930902	2.453133	3.685369
k13	0.033180	4.982852	4.949673	1.201444	4.904077e-01	0.700291	1.106136	0.135462	0.268949	2.520195	3.209130
k14	0.220451	68.830441	68.609990	9.944428	5.831025e+01	7.636115	8.169934	0.799687	1.577485	24.146719	36.194871
k15	0.347459	195.041667	194.694207	13.504512	1.791375e+02	13.384226	9.851228	1.429350	3.016160	34.982935	67.004107
k16	0.000000	31.559748	31.559748	2.149889	7.937106e+00	2.817287	1.235572	0.165181	0.342963	6.969065	15.233465
k17	0.447730	104151.000000	104150.552270	201.286497	5.109676e+06	2260.459197	20.099094	2.235337	4.058709	288.017943	4299.428041
k18	-0.975133	0.860892	1.836026	0.054974	1.225532e-02	0.110704	0.059466	-0.407070	-0.092951	0.185245	0.277408
k19	-0.321271	0.635036	0.956307	0.079206	8.836387e-03	0.094002	0.056022	-0.094769	-0.029419	0.255180	0.389926
k20	-2.670820	0.836596	3.507416	0.065179	2.155089e-02	0.146802	0.044154	-0.327641	-0.066867	0.274364	0.432738
year	5.000000	11.000000	6.000000	8.000000	4.001485e+00	2.000371	8.000000	5.000000	5.000000	11.000000	11.000000

Вывод: все описательные характеристики легко вычисляются с помощью методов библиотеки *pandas*.

Задание 2. Визуализация распределений:

1. Построить гистограмму для каждого столбца
2. Сделать выводы о форме распределений

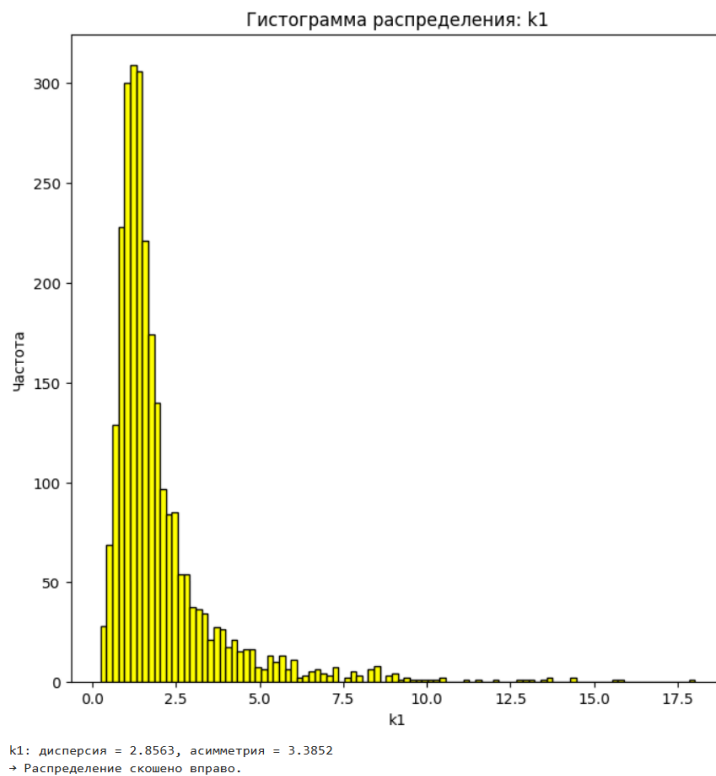
```
cols = df.select_dtypes(include=[np.number]).columns

for col in cols:
    plt.figure(figsize=(8,8))
    plt.hist(df[col].dropna(), bins=100, color='yellow', edgecolor='black')
    plt.title(f'Гистограмма распределения: {col}')
    plt.xlabel(col)
    plt.ylabel('Частота')
    plt.show()
    var = df[col].var()
    skwn = df[col].skew()

    print(f"{col}: дисперсия = {var:.4f}, асимметрия = {skwn:.4f}")

    if abs(skwn) < 0.2:
        print("→ Распределение симметричное.\n")
    elif skwn > 0.2:
        print("→ Распределение скошено вправо.\n")
    else:
        print("→ Распределение скошено влево.\n")
```

Пример для k1:



Вывод: гистограммы построены для каждого столбца таблицы с помощью библиотеки *matplotlib*.

Задание 3. Корреляционный анализ:

1. Вычислить парные корреляции Пирсона. Проверить значимость корреляций. Результаты представить в виде таблицы.
2. Для сильно коррелирующих пар построить график рассеивания (scatter plot).
3. Сделать выводы о наличии корреляций

```

corr_ma = df.corr(method='pearson')
print("\nМатрица корреляций Пирсона")
display(corr_ma)

#Проверка
p_val = df.corr(method=lambda x, y: stats.pearsonr(x, y)[1]) - np.eye(len(df.columns))
p_val_df = pd.DataFrame(p_val, columns=df.columns, index=df.columns)

print("\nr-значения для корреляций")
display(p_val_df)

a = 0.05
signf = p_val_df < a
signf_s = signf.style.map(c_bool)
display(signf_s)

edge = 0.7 # порог сильной корреляции
s_pairs = np.where((np.abs(corr_ma) > edge) & signf)

p_pairs = set()
for i, j in zip(*s_pairs):
    if i != j and (j, i) not in p_pairs:
        plt.figure(figsize=(8,8))
        sns.scatterplot(x=df.columns[i], y=df.columns[j], color='green', data=df)
        plt.title(f'Сильная корреляция: {df.columns[i]} vs {df.columns[j]} (r={corr_ma.iloc[i,j]:.2f})')
        plt.show()
        p_pairs.add((i, j))

```

Небольшая часть матрицы корреляций:

Матрица корреляций Пирсона

	empl_num	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	k11
empl_num	1.000000	0.035067	0.039541	0.057712	0.008973	-0.045174	-0.148917	-0.063621	0.044002	0.014358	0.006846	-0.002378
k1	0.035067	1.000000	0.732549	0.883032	0.452525	-0.412064	-0.113531	-0.248283	0.405975	0.557893	0.070738	-0.002682
k2	0.039541	0.732549	1.000000	0.812905	0.327937	-0.318344	-0.123917	-0.208498	0.317151	0.368025	0.078641	0.032757
k3	0.057712	0.883032	0.812905	1.000000	0.398493	-0.355593	-0.110743	-0.227242	0.350851	0.510083	0.095931	0.027845
k4	0.008973	0.452525	0.327937	0.398493	1.000000	-0.557490	-0.090429	-0.263460	0.556975	0.237392	0.209692	0.012222
k5	-0.045174	-0.412064	-0.318344	-0.355593	-0.557490	1.000000	0.036943	0.166313	-0.996378	-0.488575	0.404071	0.018430
k6	-0.148917	-0.113531	-0.123917	-0.110743	-0.090429	0.036943	1.000000	0.391975	-0.032126	-0.032287	-0.084025	-0.099150
k7	-0.063621	-0.248283	-0.208498	-0.227242	-0.263460	0.166313	0.391975	1.000000	-0.161284	-0.133519	-0.089642	-0.079915
k8	0.044002	0.405975	0.317151	0.350851	0.556975	-0.996378	-0.032126	-0.161284	1.000000	0.489973	-0.410951	-0.020897
k9	0.014358	0.557893	0.368025	0.510083	0.237392	-0.488575	-0.032287	-0.133519	0.489973	1.000000	-0.255027	-0.012213
k10	0.006846	0.070738	0.078641	0.095931	0.209692	0.404071	-0.084025	-0.089642	-0.410951	-0.255027	1.000000	0.049414
k11	-0.002378	-0.002682	0.032757	0.027845	0.012222	0.018430	-0.099150	-0.079915	-0.020897	-0.012213	0.049414	1.000000
k12	0.005375	0.038813	0.051065	0.068067	-0.077811	0.053309	-0.147884	-0.108884	-0.055908	0.014529	0.070188	0.343062
k13	-0.062109	0.048098	0.040822	0.077963	0.279167	0.025730	-0.130656	-0.245651	-0.031170	-0.225022	0.278184	0.113619
k14	0.002358	0.540829	0.374173	0.492752	0.338707	-0.319249	-0.154374	-0.325492	0.316123	0.540732	-0.031671	0.012072
k15	-0.067429	0.023106	0.092893	-0.083706	0.064928	-0.234310	-0.073267	-0.141712	0.235426	0.047573	-0.179158	0.012629

Небольшая часть матрицы р-значений:

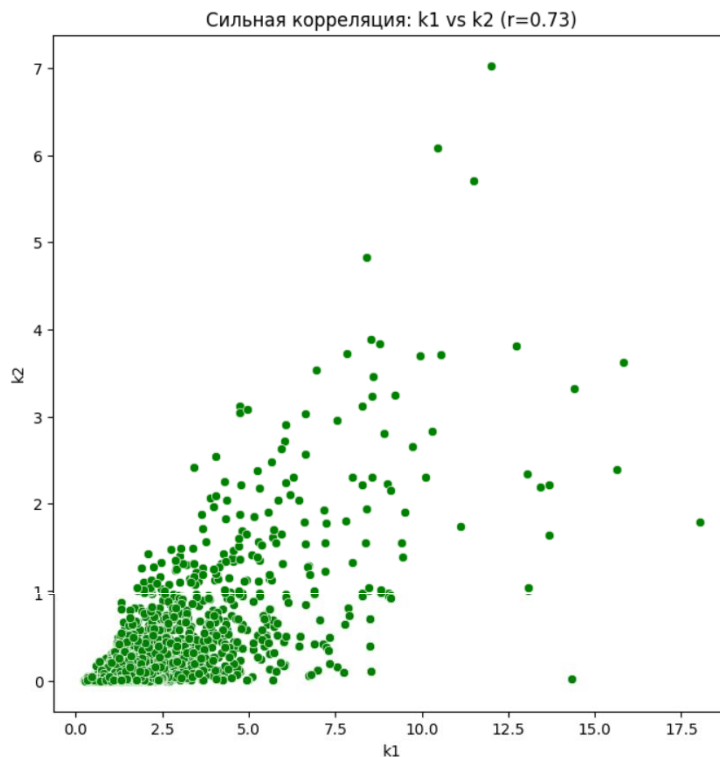
Р-значения для корреляций

	empl_num	k1	k2	k3	k4	k5	k6	k7	k8	k9
empl_num	0.000000e+00	6.873243e-02	4.011572e-02	2.725475e-03	6.414814e-01	1.901536e-02	7.813178e-15	9.511767e-04	2.235187e-02	4.562432e-01
k1	6.873243e-02	0.000000e+00	0.000000e+00	0.000000e+00	3.448569e-136	5.663318e-111	3.417791e-09	3.820532e-39	1.808754e-107	2.481508e-220
k2	4.011572e-02	0.000000e+00	0.000000e+00	0.000000e+00	1.366766e-68	1.573170e-64	1.083161e-10	7.479511e-28	4.917040e-64	3.257737e-87
k3	2.725475e-03	0.000000e+00	0.000000e+00	0.000000e+00	2.916480e-103	3.810712e-81	8.209866e-09	6.643137e-33	6.684194e-79	1.792035e-178
k4	6.414814e-01	3.448569e-136	1.366766e-68	2.916480e-103	0.000000e+00	5.974600e-220	2.576902e-06	4.970690e-44	1.832875e-219	7.734271e-36
k5	1.901536e-02	5.663318e-111	1.573170e-64	3.810712e-81	5.974600e-220	0.000000e+00	5.516271e-02	3.609331e-18	0.000000e+00	1.018998e-161
k6	7.813178e-15	3.417791e-09	1.083161e-10	8.209866e-09	2.576902e-06	5.516271e-02	0.000000e+00	1.103175e-99	9.542637e-02	9.378350e-02
k7	9.511767e-04	3.820532e-39	7.479511e-28	6.643137e-33	4.970690e-44	3.609331e-18	1.103175e-99	0.000000e+00	3.635005e-17	3.417721e-12
k8	2.235187e-02	1.808754e-107	4.917040e-64	6.684194e-79	1.832875e-219	0.000000e+00	9.542637e-02	3.635005e-17	0.000000e+00	9.023550e-163
k9	4.562432e-01	2.481508e-220	3.257737e-87	1.792035e-178	7.734271e-36	1.018998e-161	9.378350e-02	3.417721e-12	9.023550e-163	0.000000e+00
k10	7.224056e-01	2.377152e-04	4.370273e-05	6.060285e-07	3.682516e-28	2.178990e-106	1.255336e-05	3.149156e-06	2.506679e-110	2.827138e-41
k11	9.018019e-01	8.893000e-01	8.909037e-02	1.484135e-01	5.259439e-01	3.388671e-01	2.502203e-07	3.276248e-05	2.781686e-01	5.262534e-01

Р-значения, которые меньше 0,05 помечены красным и являются незначительными:

	empl_num	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	k11	k12	k13	k14	k15	k16	k17	k18	k19	k20	year
empl_num	True	False	True	True	False	True	True	True	True	False	False	False	False	True	False	True	True	True	True	False	False	False
k1	False	True	True	True	True	True	True	True	True	True	True	False	True	True	True	False	True	False	True	True	True	True
k2	True	True	True	True	True	True	True	True	True	True	True	False	True	True	True	True	True	False	True	True	True	True
k3	True	True	True	True	True	True	True	True	True	True	True	False	True	True	True	True	True	False	True	True	True	True
k4	False	True	True	True	True	True	True	True	True	True	True	False	True	True	True	True	True	False	True	True	True	False
k5	True	True	True	True	True	True	False	True	True	True	True	False	True	False	True	True	True	True	True	True	True	True
k6	True	True	True	True	True	False	True	True	False	False	True	True	True	True	True	True	True	False	True	True	True	True
k7	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True	True	False	True	True	True	True
k8	True	True	True	True	True	True	False	True	True	True	True	False	True	False	True	True	True	True	True	True	True	True
k9	False	True	True	True	True	True	False	True	True	True	True	False	False	True	True	True	True	False	True	False	False	False
k10	False	True	True	True	True	True	True	True	True	True	True	True	True	True	False	True	True	False	True	True	True	True
k11	False	False	False	False	False	False	True	True	False	False	True	True	True	True	False	False	True	False	True	True	True	True
k12	False	True	True	True	True	True	True	True	True	False	True	True	True	True	False	True	True	False	True	True	True	True
k13	True	True	True	True	True	False	True	True	False	True	True	True	True	True	True	True	True	False	True	True	True	True
k14	False	True	True	True	True	True	True	True	True	True	False	False	False	True	True	True	True	False	True	True	True	False
k15	True	False	True	True	True	True	True	True	True	True	True	False	True	True	True	True	True	False	True	True	True	True

Пример графика рассеивания для сильной корреляции между k1 и k2:



Вывод: матрица корреляции и р-значения построены с помощью метода `.corr` библиотеки *pandas*, также построена булева матрица `signf`, в которой хранятся $p < 0,05$ (для проверки надежности коэффициентов корреляции), а также для пар столбцов с коэффициентом корреляции $r > 0,7$ (сильная корреляция) и $p < 0,05$ построены графики рассеивания с помощью комбинации библиотек *matplotlib* и *seaborn* (метод `.scatterplot`).

Общий вывод: в ходе работы я хорошо познакомился с библиотеками *pandas*, *matplotlib*, *seaborn*, *numpy*, *scipy*. А также научился работать с большим количеством данных и строить графики.