

ГУАП

КАФЕДРА № 41

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

канд.техн. наук, доцент

должность, уч. степень, звание

подпись, дата

Т.А. Волкова

инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №1

СТАТИЧЕСКОЕ ИССЛЕДОВАНИЕ ДАННЫХ

по курсу: СТАТИЧЕСКАЯ ОБРАБОТКА ИНФОРМАЦИИ

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ гр. № 4315

подпись, дата

А.С. Алексеева

инициалы, фамилия

Санкт-Петербург 2025

1 Цель работы

Провести статистический анализ данных о численности врачей в регионах России, изучить основные характеристики распределения и выявить региональные различия в обеспеченности медицинскими кадрами.

2 Задание

Исследование должно содержать:

- Описание данных, с указанием их источника (какие данные взяты, из какого раздела, что фактически означают, количество данных, в чем измеряются и т.п.);
- Формальное представление данных (таблицей, если данных слишком много, то обрезать - оставить начало/конец);
- Наглядное представление данных (в виде гистограммы, столбчатой или круговой диаграммы с подписями осей, легендой и словесным описанием);
- Числовые оценки числовых характеристик данных с пояснением полученных значений (найти основные числовые характеристики - описать формулы + полученные значения + объяснение, что они означают относительно тематики данных);
- Оценку плотности распределения. Проверку подтверждения гипотезы о характере распределения (построить гистограмму, сделать предположение о функции плотности, проверить гипотезу, сделать вывод о характере распределения);
- Интервальные оценки числовых характеристик данных с пояснением полученных значений (найти доверительные интервалы);
- Сравнительная оценка числовых показателей с выводами;
- Оценка дополнительных данных, связанных с изучаемыми данными (убрать нерелевантные данные, провести сравнение со статистикой за другой год, сравнить между собой данные внутри выборки – по округам, областям и т.п. в зависимости от данных);
- Расчет дополнительных числовых показателей и их интерпретация;
- Получение и вывод формул необходимых для расчета;
- Подробные выводы по всем пунктам (о полученных характеристиках, наличии выбросов, форме распределения и т.п.).

3 Описание данных

Источником данных является Федеральная служба государственной статистики (Росстат), раздел «Здравоохранение» (<https://www.rosstat.gov.ru/folder/13721>).
Наименование показателя: «Численность врачей всех специальностей (физических лиц) в организациях, оказывающих медицинские услуги населению, на конец отчетного года».

Данные представляют собой ежегодную статистику по количеству врачей (в физических лицах), работающих в медицинских организациях системы здравоохранения Российской Федерации. Учитываются врачи всех специальностей, осуществляющие медицинскую деятельность на конец каждого отчетного года. Данные представлены в разрезе всех субъектов Российской Федерации с агрегацией по федеральным округам, а также в целом по Российской Федерации. Данные охватывают 20-летний период с 2004 по 2023 год включительно. Единицей измерения является человек (чел).

Объем данных:

- 85 строк (включая Российскую Федерацию, 8 федеральных округов и 76 субъектов РФ);
- 20 столбцов (годы с 2004 по 2023);
- Общее количество значений: 1700 данныхых точек.

Особенности данных:

- Данные за 2022 и 2023 годы приведены без учета статистической информации по Донецкой Народной Республике (ДНР), Луганской Народной Республике (ЛНР), Запорожской и Херсонской областям;
- Некоторые регионы (Республика Крым, г. Севастополь, Чеченская Республика) имеют данные не за весь рассматриваемый период.

Для статистического анализа выбраны данные за 2023 год по федеральным округам Российской Федерации (82 наблюдений), что позволяет получить репрезентативную выборку для изучения региональных различий в обеспеченности медицинскими кадрами.

4 Формальное представление данных

Для статистического анализа были выбраны данные за 2023 год по федеральным округам Российской Федерации. Исходные данные были получены из официального файла Росстата в формате XLS, содержащего обширную таблицу с динамикой численности врачей по всем субъектам РФ за период с 2004 по 2023 годы.

В целях наглядности и демонстрации структуры исходных данных в Таблице 1 представлена репрезентативная часть массива за 2023 год. Для экономии пространства приведены первые пять и последние пять записей из общего числа 82 наблюдений (субъектов РФ), что позволяет оценить состав и характер данных. Полная версия таблицы использовалась для всех последующих вычислений и визуализаций.

Таблица 1 — Фрагмент данных о численности врачей по регионам РФ на 2023 год

№ п/п	Субъект Российской Федерации	Численность врачей, чел.
1	Белгородская область	6371
2	Брянская область	4771
3	Владимирская область	5828
...
78	г. Санкт-Петербург	51166
79	Хабаровский край	8156
80	Чеченская Республика	5438
81	Чувашская Республика	5241
82	г. Москва	98955

Процесс подготовки данных включал преобразование исходной таблицы из формата XLS в CSV. Исходный файл содержал обширный массив данных с детализацией по всем субъектам Российской Федерации за многолетний период (рис. 1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Численность врачей всех специальностей (физических лиц) в организациях, оказывающих медицинские услуги, на конец года																				
2		ВСЕГО, чел																			
3		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022 ¹⁾	2023 ¹⁾
4	Российская Федерация	688221	690286	702167	707318	703849	711271	715801	732825	703220	702577	709397	673040	680916	697088	703748	714630	737371	741858	744074	758775
5	Центральный федеральный округ	194218	193873	198648	201134	200502	202267	205628	211668	199436	198137	195500	180432	181718	183888	187786	195706	210336	214379	217424	224880
6	Белгородская область	5925	5990	6167	6241	6193	6221	6189	6165	6276	6283	6307	6400	6457	6365	6263	6269	6231	6426	6469	6705
7	Брянская область	4991	4838	4831	4790	4700	4748	4620	4570	4600	4561	4569	4653	4778	4931	4861	4792	4793	4568	4458	4615
8	Владимирская область	5151	5002	5016	4985	4905	4844	4767	4977	4792	4790	4752	4727	4706	4736	4769	4774	4896	4965	5018	4976
9	Воронежская область	12375	12529	12455	12538	11931	12244	12274	12581	12683	12614	12603	12060	11885	12064	12164	12292	12060	12259	12271	12754
10	Ивановская область	5771	5621	5657	5563	5472	5449	5504	5749	5533	4929	4996	4506	4479	4445	4444	4402	4483	4349	4308	4318
11	Калужская область	4035	3967	3977	3910	3953	4069	3992	4025	4066	4093	4111	3892	4064	4101	4078	4266	4317	4484	4281	4327
12	Костромская область	2602	2568	2583	2608	2476	2427	2307	2364	2355	2333	2304	2346	2407	2409	2391	2415	2316	2313	2284	2265
13	Курская область	5782	5891	5905	6112	6267	6198	6529	6405	6360	5732	5789	5490	5448	5529	5676	5668	5778	5722	5554	5678
14	Липецкая область	4808	4756	4916	4987	4949	4901	4730	4835	4849	4992	4943	4911	4887	4827	4732	4821	4649	4561	4513	4490
15	Московская область	23874	23538	24866	26116	25337	25316	25465	29346	27404	27830	27523	27789	28208	28682	29064	32747	35828	35996	37862	38668
16	Орловская область	3157	3095	3225	3239	3177	3192	3136	3157	3182	3212	3293	3382	3369	3355	3309	3284	3277	3248	3148	3207
17	Рязанская область	6419	6331	6404	6449	6714	6386	6542	6580	5989	5882	5914	5408	5831	5683	5738	5837	5933	5965	6053	6187
18	Смоленская область	6412	6122	6041	5966	5735	5814	5863	5882	5351	5303	5259	4803	4850	4946	4856	4821	4698	4613	4504	4672
19	Тамбовская область	3939	3842	3907	3832	3808	3804	3708	3672	3679	3689	3675	3763	3849	3892	3892	4061	4221	4426	4435	4408
20	Тверская область	7364	7307	7367	7243	6926	7025	7169	7331	7168	6296	6272	5660	5722	5861	5797	5718	5717	5755	5846	5901
21	Тульская область	5505	5474	5526	5463	5376	5304	5137	5223	5165	5110	5244	5296	5430	5517	5454	5650	5782	5890	6007	6167
22	Ярославская область	7585	7630	7761	7763	7725	7881	7996	8367	7409	7372	7341	6708	6741	6681	6758	6308	6653	6674	6449	6587
23	г. Москва	78523	79372	82044	83329	84858	86444	89700	90439	82595	83116	80605	68640	68607	69864	73540	77581	88504	92165	93064	98955
24	Северо-Западный федеральный округ	73827	74458	75051	75488	75929	77816	78772	80522	78241	79522	79655	75451	78358	81121	82051	83664	85835	87309	87664	89375
25	Республика Карелия	3472	3412	3420	3363	3361	3358	3241	3224	3290	3291	3210	3050	3135	3210	3170	3177	3222	3227	3114	3138
26	Республика Коми	4376	4338	4492	4399	4380	4422	4272	4251	4263	4245	4209	4144	4272	4324	4310	4221	4150	4024	3924	3872
27	Архангельская область	6842	6817	6808	7078	6842	6818	6928	6868	6587	6495	6604	6249	6285	6354	6403	6306	6182	6182	6142	6182
28	Ненецкий автономный округ	152	164	165	163	157	163	187	199	212	210	214	204	212	200	209	227	227	237	230	223
29	Архангельская область (кроме Ненецкого автономного округа)	6690	6653	6643	6915	6685	6655	6741	6669	6375	6285	6390	6045	6073	6154	6194	6079	5955	5945	5912	5959
30	Вологодская область	4257	4357	4359	4344	4395	4307	4155	4170	4226	4137	4187	4181	4121	4187	4127	4119	4161	4065	4067	4085
31	Калининградская область	3398	3339	3456	3420	3316	3260	3248	3354	3965	4234	4141	4354	4401	4480	4545	4705	4742	4977	5112	5185
32	Ленинградская область	5132	5159	5105	5923	5801	5928	5909	5933	6024	6059	5930	5982	6313	6552	7104	6951	7014	7277	7244	7406
33	Мурманская область	4279	4209	4143	4186	4040	4572	4555	4416	4397	4260	4865	4026	4010	3921	3833	3747	3694	3620	3714	3744
34	Ненецкая область	2618	2609	2654	2621	2707	2640	2612	2591	2607	2643	2632	2650	2649	2625	2600	2691	2519	2531	2484	2622
35	Псковская область	2527	2498	2456	2448	2273	2285	2323	2370	2354	2331	2185	2206	2234	2210	2138	1939	2011	2058	2057	2005

Рисунок 1 – Исходный файл с данными

Структура данных была экспортирована в CSV формат, что обеспечило оптимальное соотношение читаемости и удобства обработки средствами программирования, а также позволило сохранить целостность и непротиворечивость данных для последующего статистического анализа (рис. 2).

region	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Российская Федерация	688221	690286	702167	707318	703849	711271	715801	732825	703220	702577	709397	673040	680916	697088	703748	714630	737371	741858	744074	758775
Центральный федеральный округ	194218	193873	198648	201134	200502	202267	205628	211668	199436	198137	195500	180432	181718	183888	187786	195706	210336	214379	217424	224880
Белгородская область	5925	5990	6167	6241	6193	6221	6189	6165	6276	6283	6307	6400	6457	6365	6263	6269	6231	6426	6469	6705
Брянская область	4991	4838	4831	4790	4700	4748	4620	4570	4600	4561	4569	4653	4778	4931	4861	4792	4793	4568	4458	4615
Владимирская область	5151	5002	5016	4985	4905	4844	4767	4977	4792	4790	4752	4727	4706	4736	4769	4774	4896	4965	5018	4976
Воронежская область	12375	12529	12455	12538	11931	12244	12274	12581	12663	12614	12603	12060	11885	12064	12164	12292	12060	12259	12271	12754
Ивановская область	5771	5621	5657	5563	5472	5449	5504	5749	5533	4929	4996	4506	4479	4445	4444	4402	4483	4349	4308	4318
Калужская область	4035	3967	3977	3910	3953	4069	3992	4025	4066	4093	4111	3892	4064	4101	4078	4266	4317	4484	4281	4327
Костромская область	2602	2568	2583	2608	2476	2427	2307	2364	2355	2333	2304	2346	2407	2409	2391	2415	2316	2313	2284	2265
Курская область	5782	5891	5905	6112	6267	6198	6529	6405	6360	5732	5789	5490	5448	5529	5676	5668	5778	5722	5554	5678
Липецкая область	4808	4756	4916	4987	4949	4901	4730	4835	4849	4992	4943	4911	4887	4827	4732	4821	4649	4561	4513	4490
Московская область	23874	23538	24866	26116	25337	25316	25465	29346	27404	27830	27523	27789	28208	28682	29064	32747	35828	35996	37862	38668
Орловская область	3157	3095	3225	3239	3177	3192	3136	3157	3182	3212	3293	3382	3369	3355	3309	3284	3277	3248	3148	3207
Рязанская область	6419	6331	6404	6449	6714	6386	6542	6580	5989	5882	5914	5408	5831	5683	5738	5837	5933	5965	6053	6187
Смоленская область	6412	6122	6041	5966	5735	5814	5863	5882	5351	5303	5259	4803	4850	4946	4856	4821	4698	4613	4504	4672
Тамбовская область	3939	3842	3907	3832	3808	3804	3708	3672	3679	3689	3675	3763	3849	3892	3892	4061	4221	4426	4435	4408
Тверская область	7364	7307	7367	7243	6926	7025	7169	7331	7168	6296	6272	5660	5722	5861	5797	5718	5717	5755	5846	5901
Тульская область	5505	5474	5526	5463	5376	5304	5137	5223	5165	5110	5244	5296	5430	5517	5454	5650	5782	5890	6007	6167
Ярославская область	7585	7630	7761	7763	7725	7881	7996	8367	7409	7372	7341	6706	6741	6681	6758	6308	6853	6674	6449	6587
г. Москва	78523	79372	82044	83329	84858	86444	89700	90435	82595	83116	80605	68640	68607	69864	73540	77581	88504	92165	93964	98955
Северо-Западный федеральный округ	73827	74458	75051	75488	75929	77816	78772	80522	78241	79522	79655	75451	78358	81121	82051	83664	85835	87309	87664	89375
Республика Карелия	3472	3412	3420	3363	3361	3358	3241	3224	3290	3291	3210	3050	3135	3210	3170	3177	3222	3227	3114	3138

Рисунок 2 – Представление данных в CSV формате

Для последующего статистического анализа исходные данные за 2023 год были подготовлены в двух основных формах:

- Вариационный ряд — исходные значения численности врачей по 82 регионам, упорядоченные по возрастанию:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

где $x_{(1)}$ - минимальное значение,

$x_{(n)}$ - максимальное значение,

n - объем выборки.

Статистический ряд (ряд распределения) — данные, сгруппированные по интервалам значений. Для группировки использовались:

- Формула Старджесса для определения оптимального количества интервалов:

$$k = 1 + 3,3 \cdot \lg(n)$$

- Расчет длины интервала:

$$h = \frac{R}{k} = \frac{x_{\max} - x_{\min}}{k},$$

где R — размах вариации, ед.;

x_{\max} — максимальное значение в выборке, ед.;

x_{\min} — минимальное значение в выборке, ед.;

k — количество интервалов группировки, ед.

5 Наглядное представление данных

На рисунке 3 представлена столбчатая диаграмма по численности врачей в федеральных округах в 2023 году.

Описание диаграммы:

- Ось X: Федеральные округа Российской Федерации (топ-20 регионов по

численности врачей);

- Ось Y: Численность врачей, человек;
- Легенда: Цветовые различия столбцов визуально выделяют различные федеральные округа;
- Масштаб: Линейная шкала от 0 до 100000 человек с шагом 10000.

На диаграмме видно, что наибольшее число врачей сосредоточено в городе Москве (98955 человек), что более чем в 2,5 раза превышает показатель следующего за ним Санкт-Петербурга (51166 человек). Наименьшая численность врачей среди топ-20 регионов наблюдается в Красноярском крае (16428 человек), что в 6 раз меньше показателя Москвы.

Диаграмма наглядно демонстрирует значительную региональную диспропорцию в распределении медицинских кадров по территории России. Особенно выделяются два города федерального значения (Москва и Санкт-Петербург), имеющие аномально высокую численность врачей по сравнению с другими регионами. Также можно отметить, что первые 5 регионов имеют численность врачей выше 25000 человек, в то время как остальные регионы из топ-20 находятся в диапазоне от 16000 до 25000 врачей.

Разница между максимальным и минимальным значениями по всем регионам составляет 98732 человека (от 223 человек в Ненецком автономном округе до 98955 человек в Москве), что подчеркивает экстремальный разброс в обеспеченности медицинскими кадрами между регионами.

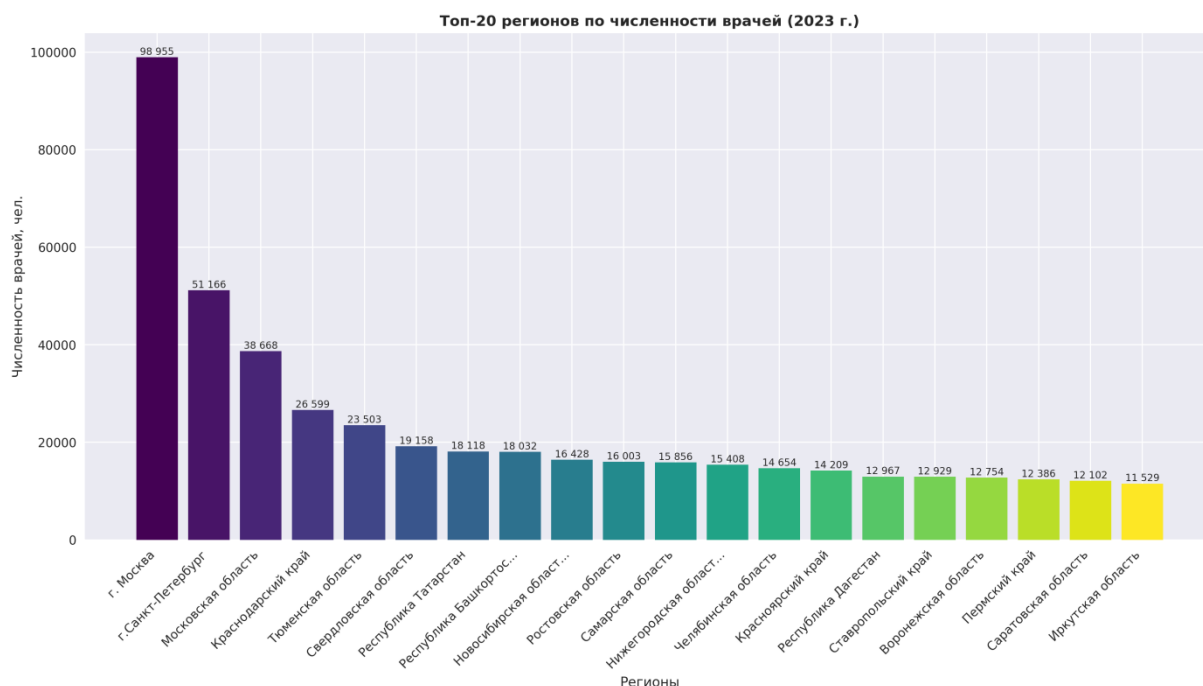


Рисунок 3 – Столбчатая диаграмма численности врачей по федеральным округам за 2023 год

На рисунке 4 представлена гистограмма распределения численности врачей по всем 82 регионам России за 2023 год.

Описание диаграммы:

- Ось X: Численность врачей, человек
- Ось Y: Количество регионов

Гистограмма показывает правостороннее асимметричное распределение:

- Большинство регионов сосредоточены в первом интервале (примерно 0-20000 врачей);
- Распределение имеет длинный «хвост» в правой части с регионами, имеющими очень высокую численность врачей;
- Только 6 регионов имеют численность врачей более 60000 человек.

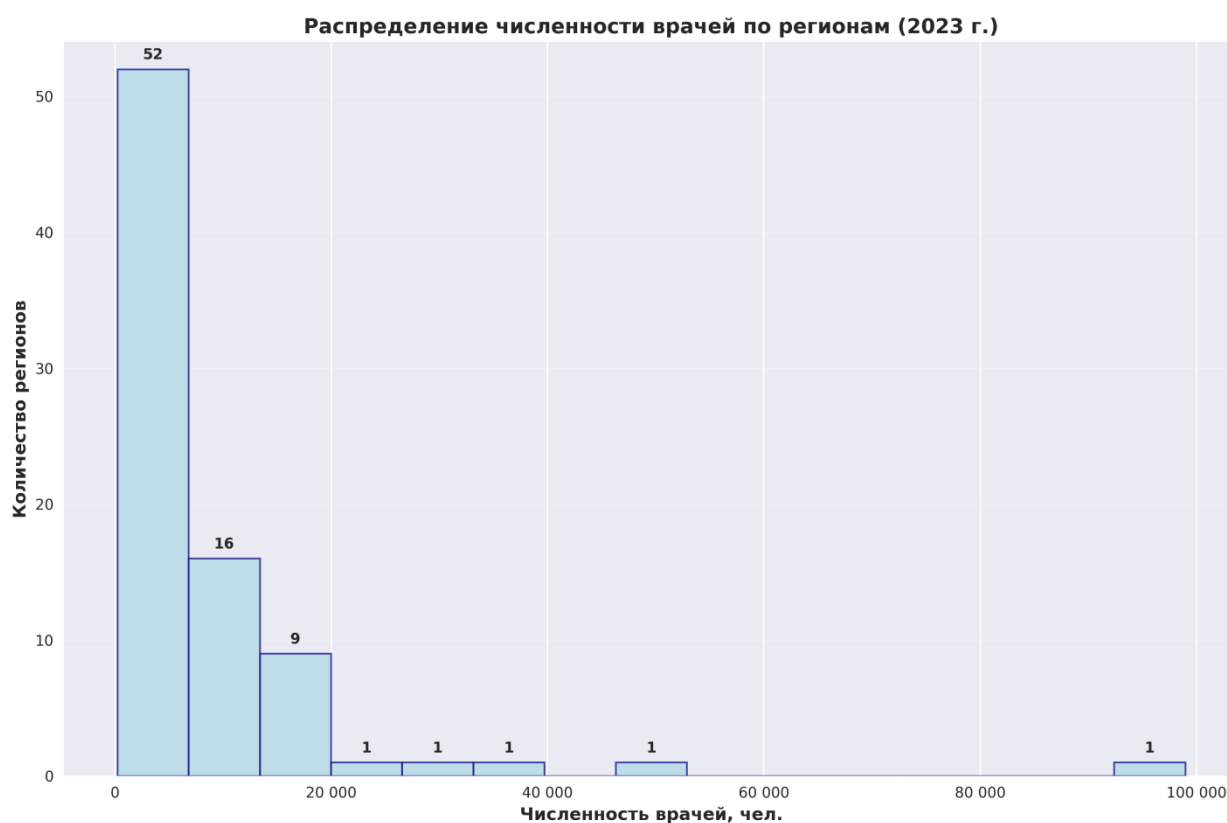


Рисунок 4 – Гистограмма распределения по всем регионам

На рисунке 5 представлена диаграмма средней численности врачей по федеральным округам.

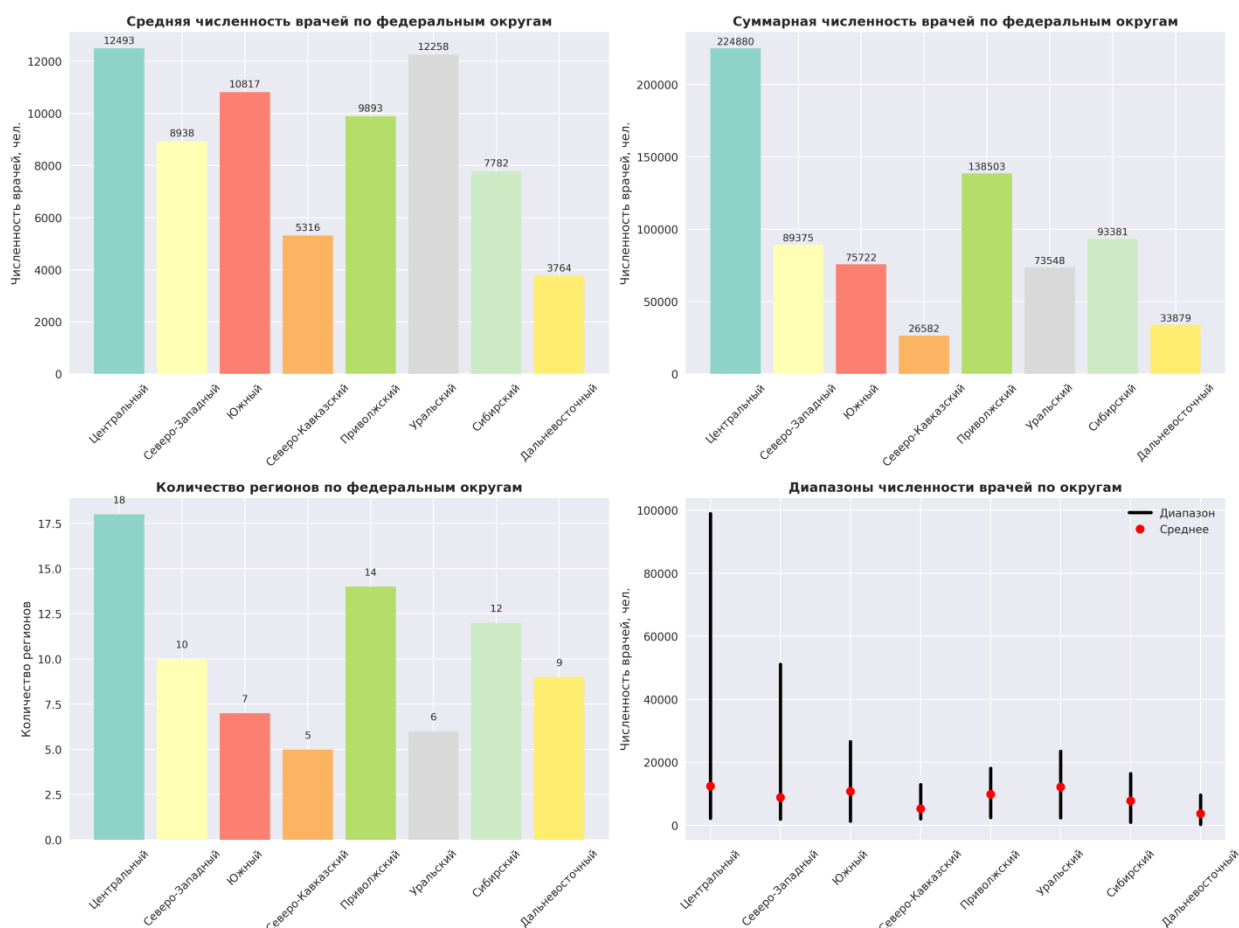


Рисунок 5 – Диаграмма средней численности врачей по федеральным округам

6 Настройка программного окружения и установка зависимостей

Перед проведением статистического анализа была выполнена предварительная настройка программного окружения. Для изоляции зависимостей и обеспечения воспроизводимости результатов было создано виртуальное окружение Python с использованием модуля `venv`. Виртуальное окружение позволяет управлять пакетами для конкретного проекта без влияния на системные установки.

Были установлены следующие библиотеки, необходимые для проведения анализа:

- `pandas` - для работы с табличными данными и их обработки;
- `numpy` - для выполнения математических операций и расчетов;
- `scipy` - для проведения статистических тестов и анализа распределений;
- `matplotlib` - для построения графиков и визуализации результатов.

Установка зависимостей проводилась с помощью менеджера пакетов `pip` на основе файла `requirements.txt` (рис. 6), который содержал перечень необходимых библиотек с указанием версий. Данный подход обеспечивает согласованность версий пакетов и предотвращает потенциальные конфликты зависимостей.


```
≡ requirements.txt ×
≡ requirements.txt
1 pandas==2.1.4
2 numpy==1.26.0
3 scipy==1.11.3
4 matplotlib==3.8.0
```

Рисунок 6 – Файл requirements.txt

На рисунке 7 представлен процесс настройки программного окружения, включающий создание виртуального окружения, активацию и установку пакетов через pip.

```
(.env) anya@anya-zenbook:~/Documents/ГУАП/Learning/4315 Большой ГУАП/5сем/Статистическая обработка информации/ЛР_1/ЛР_1$ python3 -m venv .venv
(.env) anya@anya-zenbook:~/Documents/ГУАП/Learning/4315 Большой ГУАП/5сем/Статистическая обработка информации/ЛР_1/ЛР_1$ source .venv/bin/activate
(.env) anya@anya-zenbook:~/Documents/ГУАП/Learning/4315 Большой ГУАП/5сем/Статистическая обработка информации/ЛР_1/ЛР_1$ pip install -r requirements.txt
^CERROR: Operation cancelled by user
(.env) anya@anya-zenbook:~/Documents/ГУАП/Learning/4315 Большой ГУАП/5сем/Статистическая обработка информации/ЛР_1/ЛР_1$ pip install -r requirements.txt
Collecting pandas==2.1.4 (from -r requirements.txt (line 1))
  Downloading pandas-2.1.4-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (18 kB)
Collecting numpy==1.26.0 (from -r requirements.txt (line 2))
  Downloading numpy-1.26.0-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (58 kB)
    58.5/58.5 kB 684.7 kB/s eta 0:00:00
Collecting scipy==1.11.3 (from -r requirements.txt (line 3))
  Downloading scipy-1.11.3-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (60 kB)
    60.4/60.4 kB 1.4 MB/s eta 0:00:00
Collecting matplotlib==3.8.0 (from -r requirements.txt (line 4))
  Downloading matplotlib-3.8.0-cp312-cp312-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.8 kB)
```

Рисунок 7 – Настройка окружения и установка зависимостей

7 Числовые оценки числовых характеристики данных

В листинге А представлен код функции, для расчета числовых характеристик данных. На рисунке 8 представлен вывод расчета основных числовых характеристик в терминале.

```
ОСНОВНЫЕ ВЫБОРОЧНЫЕ ОЦЕНКИ
=====
Объем выборки: 82 регионов
Среднее выборочное: 9220.65 чел.
Выборочные начальные моменты:
  a2: 248577652.21
  a3: 1.53e+13
  a4: 1.30e+18
Выборочные центральные моменты:
  m2: 163557333.25
  m3: 1.00e+13
  m4: 8.41e+17
Смещенная выборочная дисперсия: 163557333.25
Несмещенная выборочная дисперсия: 165576559.59
Выборочное среднее квадратическое отклонение: 12788.95 чел.
```

Рисунок 8 - Вывод расчета основных числовых характеристик в терминале

7.1 Объем выборки

Объем выборки (n) рассчитывается как количество наблюдений в совокупности. Объем выборки определяет надежность статистических выводов. При $n = 82$ имеется достаточный объем данных для проведения статистического анализа и построения доверительных интервалов.

7.2 Среднее выборочное

Среднее выборочное рассчитывается по формуле:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

где x_i - значение i -го элемента выборки,

n - объем выборки.

Средняя численность врачей в расчете на один регион составляет 9221 человек. Данный показатель представляет точечную оценку математического ожидания генеральной совокупности, однако существенное отличие от медианы указывает на асимметрию распределения.

7.3 Выборочные начальные моменты

Выборочные начальные моменты k -го порядка (a_k) рассчитываются по формуле:

$$a_k = \frac{\sum_{i=1}^n x_i^k}{n}$$

Выборочные начальные моменты равняются:

- $a_1 : 248577552.21$ - совпадает со средним значением;
- $a_2 : 1.53 \times 10^{13}$;
- $a_3 : 1.30 \times 10^{18}$.

Начальные моменты используются для расчета центральных моментов и анализа формы распределения.

7.4 Выборочные центральные моменты

Выборочные центральные моменты k -го порядка (m_k) рассчитываются по формуле:

$$m_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^k}{n}$$

Выборочные центральные моменты равняются:

- $m_2 : 163557333.25$ - второй центральный момент (дисперсия);
- $m_3 : 1.00 \times 10^{13}$ - третий центральный момент, используется для расчета

коэффициента асимметрии;

– $m_4 : 8.41 \times 10^{17}$ - четвертый центральный момент, используется для расчета коэффициента эксцесса.

Центральные моменты характеризуют форму распределения данных: его рассеивание, асимметрию и остроту вершины.

7.5 Выборочная дисперсия

Выборочная дисперсия (смещенная) рассчитывается по формуле:

$$S_{\text{смещ}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Выборочная дисперсия (смещенная) составляет 163557333

Выборочная дисперсия (несмещенная) рассчитывается по формуле:

$$S_{\text{несмещ}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Выборочная дисперсия (несмещенная) составляет 165576560.

Дисперсия характеризует средний квадрат отклонений значений от среднего. Большая величина дисперсии свидетельствует о высокой вариативности данных между регионами. Несмещенная оценка (s^2) является более точной при работе с выборкой и используется для корректного статистического анализа генеральной совокупности.

7.6 Выборочное среднее квадратичное отклонение

Среднее квадратичное отклонение рассчитывается по формуле:

$$S = \sqrt{S_{\text{несмещ}}^2}$$

СКО показывает среднее отклонение значений численности врачей от среднего значения. Величина в 12 788.95 человек (что составляет примерно 138.7% от среднего значения) указывает на чрезвычайно высокую вариативность данных и сильно скошенное распределение, при котором среднее значение плохо описывает типичный регион.

8 Оценка плотности распределения

Для анализа характера распределения численности врачей по регионам Российской Федерации был проведен комплексный статистический анализ, включающий построение гистограммы, оценку плотности распределения и формальную проверку статистических гипотез.

В листинге Б представлен код функции, для оценки плотности распределения.

8.1 Визуальный анализ распределения

На рисунке 9 представлена гистограмма эмпирического распределения данных с наложенными теоретическими кривыми плотности.



Рисунок 9 - Сравнение эмпирической и теоретической плотностей распределения

Визуальный анализ позволяет сделать следующие наблюдения:

- Эмпирическое распределение (гистограмма) демонстрирует выраженную правостороннюю асимметрию;
- Теоретическая нормальная плотность (красная линия) плохо соответствует эмпирическим данным;
- KDE-оценка плотности (зеленая пунктирная линия) более точно отражает форму реального распределения, показывая концентрацию данных в левой части с длинным правым хвостом.

8.2 Сравнение различных оценок центра распределения

На рисунке 10 представлено сравнение различных точечных оценок среднего значения.

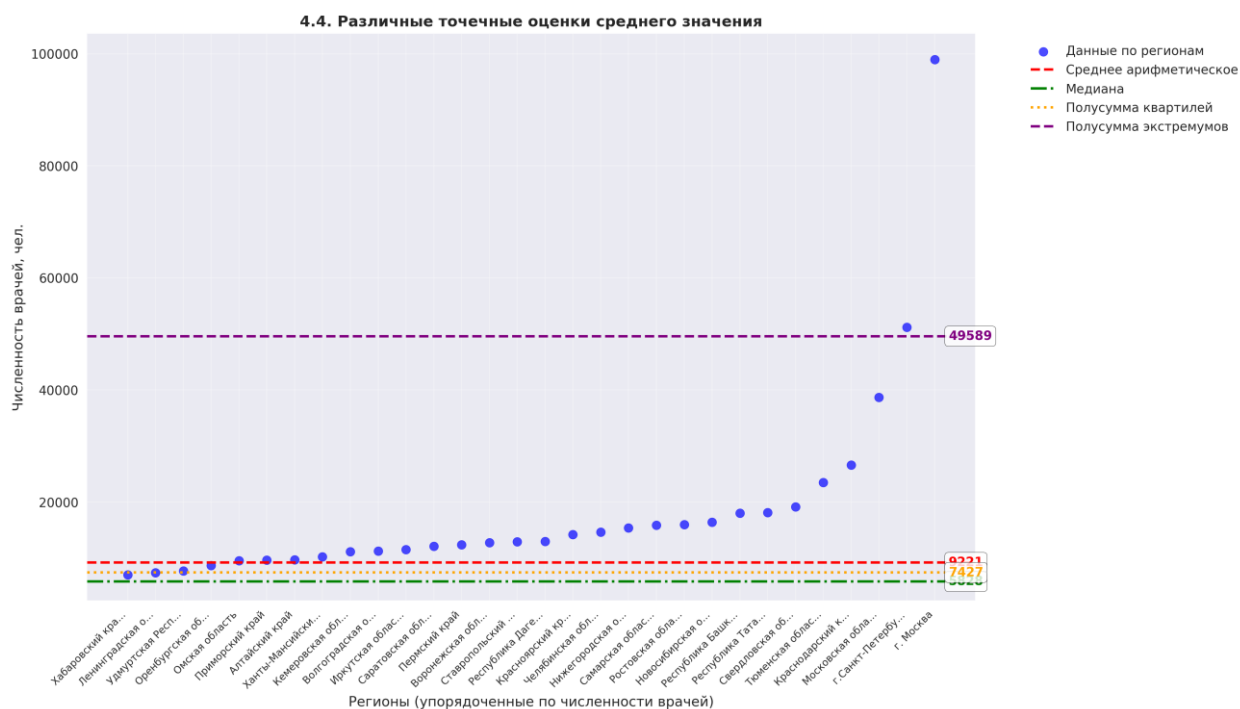


Рисунок 10 - Различные точечные оценки среднего значения

Анализ различных оценок центра распределения показывает:

- Среднее арифметическое: 9 221 чел. - сильно смещено вправо из-за влияния выбросов (Москва, Санкт-Петербург);
- Медиана: 5 828 чел. - более устойчивая оценка, менее чувствительная к выбросам;
- Полусумма квартилей: 7 427 чел. - близка к медиане, отражает центр основной массы данных;
- Полусумма экстремумов: 49 589 чел. - значительно завышена из-за большого размаха данных.

Значительный разброс между различными оценками среднего (от 5 828 до 49 589 чел.) дополнительно подтверждает асимметричность распределения.

8.3 Формальная проверка гипотезы о нормальности

Для количественной оценки соответствия распределения нормальному закону был применен критерий Шапиро-Уилка:

- Нулевая гипотеза (H_0): Распределение численности врачей является нормальным;
- Альтернативная гипотеза (H_1): Распределение значимо отличается от нормального.

Для теста Шапиро-Уилка использовалась формула:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

где a_i - константы,

$x_{(i)}$ - упорядоченное значение выборки.

Результаты:

- Статистика $W = 0.5234$;
- $p\text{-value} = 0$.

Поскольку полученное $p\text{-value}$ (0.0000) меньше стандартного уровня значимости $\alpha = 0.05$, отвергаем нулевую гипотезу H_0 . Распределение численности врачей по регионам статистически значимо отличается от нормального распределения.

8.4 Вывод о характере распределения:

Характер распределения можно охарактеризовать как:

- Правостороннее асимметричное - наличие регионов с аномально высокой численностью врачей (Москва, Санкт-Петербург);
- С выраженной положительной асимметрией - коэффициент асимметрии $+4,7128$ подтверждает сильное смещение вправо;
- Мультимодальное - наличие нескольких локальных максимумов в распределении, что свидетельствует о кластеризации регионов по уровням обеспеченности врачами.

9 Интервальные оценки числовых характеристик данных

Для получения более надежных оценок параметров распределения численности врачей по регионам Российской Федерации были построены доверительные интервалы для основных числовых характеристик. Поскольку истинное распределение данных неизвестно, использовались подходы для нормального распределения при уровне доверия 95%.

В листинге В представлен код функции для расчета интервальных оценок числовых характеристик данных. На рисунке 11 представлен вывод расчета интервальных оценок в терминале.

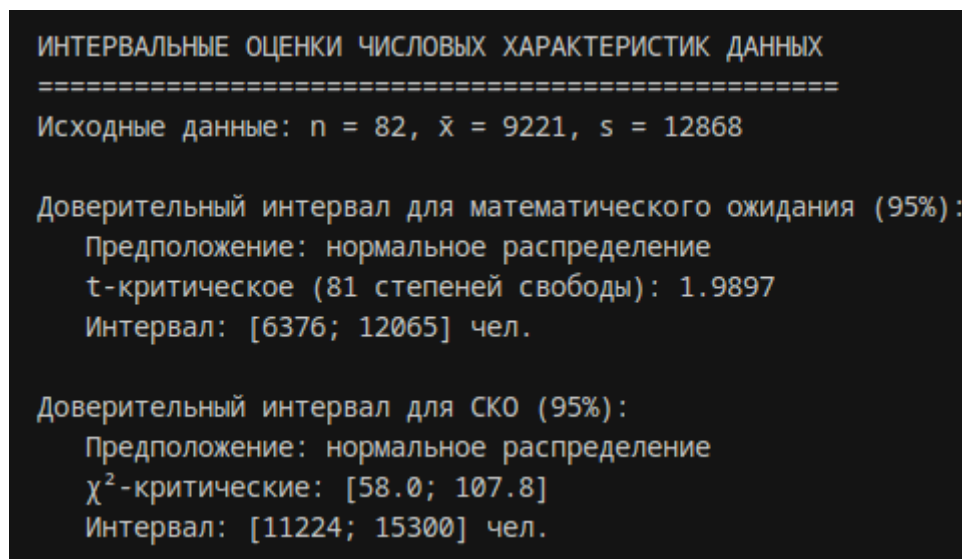


Рисунок 11 - Интервальные оценки числовых характеристик данных

9.1 Доверительный интервал для математического ожидания

В предположении нормальности распределения был построен доверительный интервал с использованием t-распределения Стьюдента:

$$P\left(\bar{x} - \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}} < \alpha < \bar{x} + \frac{st_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right) = 1 - \alpha,$$

где $\bar{x}=9221$ чел. - выборочное среднее;

$S=12868$ чел. - выборочное стандартное отклонение;

$n=82$ - объем выборки;

$st_{1-\alpha/2}(n-1) = 1,9897$ — квантиль t-распределения с 81 степенью свободы.

Полученный интервал: [6376; 12065] человек.

С надежностью 95% можно утверждать, что математическое ожидание численности врачей по всем регионам России находится в пределах от 6376 до 12065 человек. Ширина интервала составляет 5689 человек, что свидетельствует о значительной неопределенности в оценке среднего значения. Такая большая ширина интервала объясняется высокой вариативностью данных между регионами.

9.2 Доверительный интервал для среднеквадратичного отклонения

Для построения доверительного интервала для параметра σ использовалось распределение хи-квадрат:

$$P\left(\frac{s\sqrt{n}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \frac{s\sqrt{n}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right) = 1 - \alpha,$$

где $\chi^2_{1-\alpha/2}(n-1)$ и $\chi^2_{\alpha/2}(n-1)$ - критические значения распределения хи-квадрат (равняются 107.8 и 58.0 соответственно).

Полученный интервал равен [11224; 15300] человек.

С надежностью 95% можно утверждать, что стандартное отклонение численности врачей по регионам России находится в пределах от 11224 до 15300 человек. Интервал не включает малые значения, что подтверждает высокую вариативность данных. Точечная оценка стандартного отклонения (12868 чел.) близка к центру интервала, что указывает на хорошую точность оценки.

10 Сравнительная оценка числовых показателей с выводами

Проведенный статистический анализ позволяет провести сравнительную оценку расчетных числовых показателей и сделать выводы о статическом исследовании данных о численности врачей.

10.1 Сравнение показателей центра распределения

Числовой показатель: Среднее арифметическое (9221 чел.) существенно превышает медиану (5828 чел.) и полусумму квартилей (7427 чел.).

Столь значительный разрыв (разница между средним и медианой составляет ~3400 чел.) опровергает гипотезу о том, что среднее значение адекватно описывает «типичный» регион. Медиана является более репрезентативной оценкой центра для данной выборки, что подтверждает вывод о сильном влиянии регионов-выбросов (Москва, Санкт-Петербург) на итоговые характеристики.

10.2 Сравнение показателей вариативности и вывод о степени неравномерности

Числовые показатели: Стандартное отклонение (12868 чел.) больше самого среднего значения, а размах выборки достигает 98732 чел.

Коэффициент вариации (>100%) и экстремальный размах являются числовым подтверждением ключевого вывода работы о наличии экстремальной региональной диспропорции. Вывод о «критическом уровне различий» между регионами напрямую следует из этих расчетных показателей.

10.3 Сравнение показателей формы распределения и вывод о нормальности:

Числовые показатели: Коэффициент асимметрии (+4.71), p-value критерия Шапиро-Уилка (≈ 0.0).

Сравнительная оценка и вывод: Полученные значения формально опровергают нулевую гипотезу (Н₀) о нормальности распределения. Это позволяет сделать качественный вывод о том, что распределение не является случайным и подчиняется специфическим закономерностям (например, социально-экономическим), что подтверждает тезис о системном характере диспропорций.

10.4 Сравнение интервальных и точечных оценок и вывод о надежности результатов:

Числовые показатели: Доверительный интервал для математического ожидания [6376; 12065] очень широк (ширина ~5700 чел.).

Большая ширина интервала, захватывающая как значения, близкие к медиане, так и к среднему, подтверждает вывод о высокой неопределенности и нестабильности оценок из-за огромного разброса данных. Это свидетельствует о необходимости осторожного использования точечных оценок без учета их доверительных интервалов.

Сравнение всех ключевых числовых показателей – среднего, медианы, мер вариативности, коэффициентов формы распределения и доверительных интервалов приводит к выводу, что в России существует выраженная, статистически подтвержденная региональная асимметрия в распределении врачебных кадров, носящая системный характер.

11 Оценка дополнительных данных, связанных с изучаемыми данными

Проведен сравнительный анализ данных за 2013 и 2023 годы, а также анализ распределения численности врачей по типам регионов и выявление регионов-лидеров. На рисунке 12 представлена оценка дополнительных данных, выведенных через терминал. В листинге Д представлены функции для оценки дополнительных данных.

ОЦЕНКА ДОПОЛНИТЕЛЬНЫХ ДАННЫХ
=====
СРАВНЕНИЕ 2013-2023:

Средняя численность:
- 2013: 8662 чел.
- 2023: 9221 чел.
- Изменение: +559 чел. (+6.5%)
Рост в 39 регионах (47.6%)
ТИПЫ РЕГИОНОВ:

• Республики (20): 5438 чел.
• Края (9): 11051 чел.
• Области (47): 8156 чел.
• Города (2): 75060 чел.
• Автономные округа (4): 3610 чел.
ТОП-5 РЕГИОНОВ:

• г. Москва: 98955 чел. (13.1%)
• г. Санкт-Петербург: 51166 чел. (6.8%)
• Московская область: 38668 чел. (5.1%)
• Краснодарский край: 26599 чел. (3.5%)
• Тюменская область: 23503 чел. (3.1%)

Рисунок 12 - Оценка дополнительных данных в терминале

11.1 Сравнительный анализ динамики за 2013-2023 годы

За 10-летний период наблюдается положительная динамика в обеспеченности регионов медицинскими кадрами:

- Средняя численность врачей выросла с 8662 до 9221 человека, что составляет прирост +559 человек (+6.5%)
- Рост численности отмечен в 39 регионах (47.6% от общего числа), в то время как в остальных регионах наблюдалось снижение или стагнация показателей

Положительная динамика свидетельствует о системных мерах по увеличению численности медицинских кадров, однако неравномерность роста по регионам указывает на сохранение территориальных диспропорций.

11.2 Анализ по типам регионов

Наблюдается значительная дифференциация в обеспеченности врачами в зависимости от типа региона:

- Города федерального значения (Москва, Санкт-Петербург) - 75060 чел. (в среднем);
- Края - 11051 чел.;
- Области - 8156 чел.;
- Республики - 5438 чел.;
- Автономные округа - 3610 чел.

Выявлена четкая иерархия: города федерального значения значительно превосходят все другие типы регионов по средней численности врачей (в 6.8 раза больше, чем в среднем по областям).

11.3 Анализ регионов-лидеров

Топ-5 регионов по численности врачей концентрируют значительную долю медицинских кадров страны:

- г. Москва - 98955 чел. (13.1% от общей численности);
- г. Санкт-Петербург - 51166 чел. (6.8%);
- Московская область - 38668 чел. (5.1%);
- Краснодарский край - 26599 чел. (3.5%);
- Тюменская область - 23503 чел. (3.1%).

Суммарно первые 5 регионов обеспечивают 31.6% от общей численности врачей в России, что свидетельствует о высокой концентрации медицинских кадров в ограниченном числе регионов.

12 Расчет дополнительных числовых показателей

В листинге В представлен код функции, для расчета дополнительных числовых характеристик данных. На рисунке 13 представлен вывод расчета дополнительных числовых характеристик в терминале.

```

ДОПОЛНИТЕЛЬНЫЕ ВЫБОРОЧНЫЕ ОЦЕНКИ
=====
Выборочная медиана: 5827.50 чел.
Выборочное среднее абсолютное отклонение: 5963.87 чел.
Выборочные квартили:
    Q1 (25%): 3631.50 чел.
    Q2 (50%, медиана): 5827.50 чел.
    Q3 (75%): 11222.75 чел.
Интерквартильная широта: 7591.25 чел.
Полусумма выборочных квартилей: 7427.12 чел.
Разность выборочных квартилей: 7591.25 чел.
Экстремальные элементы:
    Минимум: 223 чел. (Ненецкий автономный округ)
    Максимум: 98955 чел. (г. Москва)
Размах выборки: 98732.00 чел.
Полусумма экстремальных элементов: 49589.00 чел.
Выборочная оценка асимметрии: 4.8004
Выборочная оценка эксцесса: 28.4220

```

Рисунок 13 - Вывод расчета дополнительных числовых характеристик в терминале

11.1 Выборочная медиана

Медиана рассчитывается как значение, разделяющее упорядоченную выборку пополам:

$$\text{med} = \begin{cases} x_{(l+1)}, & n=2l+1 \\ \frac{x_{(l)}+x_{(l+1)}}{2}, & n=2l \end{cases}$$

где l - целая часть от деления объема выборки n на 2.

Половина регионов имеет численность врачей менее 5828 человек. Значительное отличие медианы от среднего арифметического (3393 человека) подтверждает правостороннюю асимметрию распределения.

11.2 Выборочное среднее абсолютное отклонение

Выборочное абсолютное отклонение рассчитывается по формуле:

$$d = \frac{\sum_{i=1}^n |x_i - \text{med}|}{n}$$

Выборочное абсолютное отклонение составляет 5964 человек.

Данный показатель характеризует среднее абсолютное отклонение значений от медианы. В отличие от дисперсии и стандартного отклонения, абсолютное отклонение менее чувствительно к экстремальным значениям и выбросам в данных. Величина в 5964

человека показывает, что в среднем регионы отклоняются от медианного значения численности врачей, что дополнительно подтверждает высокую вариативность данных.

11.7 Выборочные квартили

Квартили разделяют упорядоченную выборку на четыре равные части:

- Q_1 (25-й процентиль): 3632 человека;
- Q_2 (50-й процентиль): 5828 человек;
- Q_3 (75-й процентиль): 11223 человека;

25% регионов имеют численность врачей ниже 3632 человек, 75% - ниже 11223 человек. Центральные 50% данных расположены в интервале от 3632 до 11223 врачей.

11.8 Интерквартильная широта, полусумма и разность выборочных квартилей

Интерквартильная широта (разность выборочных квартилей) рассчитывается по формуле:

$$IQR = Q_3 - Q_1$$

Полусумма выборочных квартилей рассчитывается по формуле:

$$t_q = \frac{(Q_1 + Q_3)}{2}$$

Интерквартильная широта характеризует разброс средних 50% данных. Значение в 7591 человек показывает значительную вариативность даже в центральной части распределения, исключая влияние экстремальных значений.

Полусумма квартилей представляет альтернативную оценку центра распределения, менее чувствительную к выбросам (составляет 7427 человек).

11.9 Экстремальные элементы и размах выборки

Минимум составляет 223 человек (Немецкий автономный округ). Максимум составляет 98955 человек (г. Москва). Размах выборки рассчитывается как разница между максимумом и минимумом:

$$R = X_{\max} - X_{\min}$$

Размах выборки составляет 98732 человек.

Экстремальные значения показывают границы вариации данных. Разница между самым и наименее обеспеченным врачами регионами составляет 98 732 человека (в 444 раза), что свидетельствует о чрезвычайно высокой региональной дифференциации в обеспеченности медицинскими кадрами.

11.10 Полусумма экстремальных выборочных элементов

Полусумма экстремальных элементов рассчитывается по формуле:

$$t_r = \frac{(X_{\max} + X_{\min})}{2}$$

Полусумма экстремальных элементов составляет 49589 человек. Данный показатель представляет еще одну оценку центра распределения, однако в случае сильно асимметричных данных с выбросами (как в данной выборке) он может существенно смещаться в сторону экстремальных значений.

11.11 Выборочная оценка асимметрии

Выборочная оценка асимметрии рассчитывается по формуле:

$$A_s = \frac{\mu_3}{\sigma^3} = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}}{S^3},$$

где μ_3 - третий центральный момент,

σ^3 - стандартное отклонение.

Коэффициент асимметрии составляет 4.7128. Положительное значение указывает на выраженную правостороннюю асимметрию распределения. «Хвост» распределения вытянут в сторону больших значений, что соответствует наличию регионов с аномально высокой численностью врачей.

11.12 Выборочная оценка эксцесса

Выборочная оценка эксцесса рассчитывается по формуле:

$$E_k = \frac{\mu_4}{\sigma^4} - 3 = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}}{S^4} - 3,$$

где μ_4 - четвертый центральный момент.

Коэффициент эксцесса составляет 27.6603. Положительное значение свидетельствует о островершинной форме распределения (лептокуртическое распределение), что характерно для данных с концентрацией значений в хвостах распределения и наличием выбросов.

Выводы

На основе комплексного статистического анализа данных о численности врачей в регионах Российской Федерации за 2023 год были получены следующие основные выводы:

– Установлена экстремальная региональная дифференциация в распределении врачебных кадров. Размах выборки составляет 98732 человека - от 223 чел. в Ненецком

автономном округе до 98955 чел. в Москве, что превышает разницу в 444 раза.

- Выявлено правостороннее асимметричное распределение с коэффициентом асимметрии +4,71. Распределение значительно отличается от нормального (p -value критерия Шапиро-Уилка = 0,0000), что свидетельствует о системном характере наблюдаемых диспропорций.

- Обнаружено сильное влияние выбросов на статистические характеристики. Среднее арифметическое (9221 чел.) существенно превышает медиану (5828 чел.), что указывает на нерепрезентативность среднего значения для описания «типичного» региона.

- Подтверждена высокая вариативность данных: стандартное отклонение (12868 чел.) составляет 139.6% от среднего значения, а доверительный интервал для математического ожидания [6376; 12065] имеет ширину 5689 чел., что свидетельствует о значительной неопределенности оценок.

- Выявлена концентрация медицинских кадров: 25% наиболее обеспеченных регионов сосредотачивают более 31% от общей численности врачей, при этом первые 5 регионов (Москва, Санкт-Петербург, Московская область, Краснодарский край, Тюменская область) обеспечивают 31.6% всех врачей страны.

- Установлена четкая иерархия по типам регионов: города федерального значения (75060 чел. в среднем) превосходят другие типы регионов в 6.8 раза по сравнению с областями (8156 чел.).

- Зафиксирована положительная динамика: за 10-летний период средняя численность врачей выросла на 6.5%, однако рост носил неравномерный характер - лишь 47.6% регионов показали увеличение показателя.

- Проведенное исследование статистически подтвердило наличие выраженных региональных диспропорций в распределении врачебных кадров, что обосновывает необходимость разработки дифференцированных подходов к кадровой политике в здравоохранении с учетом территориальной специфики.

Материалы исследования, включая исходный код и дополнительные визуализации, размещены в открытом репозитории:

<https://github.com/AnyAleks/rosstat-doctors-statistical-research>

ЛИСТИНГ А

Листинг расчета числовых характеристик

```
def calculate_basic_characteristics(self):
    print("ОСНОВНЫЕ ВЫБОРОЧНЫЕ ОЦЕНКИ")
    print("=" * 50)
    chars = {}
    n = len(self.doctors_data)
    chars['n'] = n
    # Объем выборки
    print(f"Объем выборки: {chars['n']} регионов")
    # Среднее выборочное
    chars['mean'] = np.mean(self.doctors_data)
    print(f"Среднее выборочное: {chars['mean']:.2f} чел.")
    # Выборочные начальные моменты
    chars['a2'] = np.mean(self.doctors_data**2)
    chars['a3'] = np.mean(self.doctors_data**3)
    chars['a4'] = np.mean(self.doctors_data**4)
    print(f"Выборочные начальные моменты:")
    print(f"  a2: {chars['a2']:.2f}")
    print(f"  a3: {chars['a3']:.2e}")
    print(f"  a4: {chars['a4']:.2e}")
    # Выборочные центральные моменты
    chars['m2'] = np.mean((self.doctors_data - chars['mean'])**2)
    chars['m3'] = np.mean((self.doctors_data - chars['mean'])**3)
    chars['m4'] = np.mean((self.doctors_data - chars['mean'])**4)
    print(f"Выборочные центральные моменты:")
    print(f"  m2: {chars['m2']:.2f}")
    print(f"  m3: {chars['m3']:.2e}")
    print(f"  m4: {chars['m4']:.2e}")
    # Смещенная выборочная дисперсия
    chars['variance_biased'] = chars['m2']
    print(f"Смещенная выборочная дисперсия: {chars['variance_biased']:.2f}")
    # Несмещенная выборочная дисперсия (ДОБАВЛЕНО)
    chars['variance_unbiased'] = np.var(self.doctors_data, ddof=1)
    print(f"Несмещенная выборочная дисперсия: {chars['variance_unbiased']:.2f}")
    # Выборочное среднее квадратическое отклонение (смещенное)
    chars['std_biased'] = np.sqrt(chars['variance_biased'])
    print(f"Выборочное среднее квадратическое отклонение: {chars['std_biased']:.2f} чел.")
    # Несмещенное СКО (ДОБАВЛЕНО для совместимости)
    chars['std_dev'] = np.sqrt(chars['variance_unbiased'])
    print()
    return chars
```


ЛИСТИНГ Б

Листинг оценки плотности распределения

```
def section_1_5_distribution_density(self):
    print("ОЦЕНКА ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ")
    print("=" * 50)
    # Проверка гипотезы о нормальности
    shapiro_stat, shapiro_p = stats.shapiro(self.doctors_data)
    print("Гипотезы:")
    print(" H0: Распределение является нормальным")
    print(" H1: Распределение не является нормальным")
    print(f" Критерий Шапиро-Уилка: W = {shapiro_stat:.4f}, p-value = {shapiro_p:.4f}")
    if shapiro_p > 0.05:
        print(" Вывод: Нет оснований отвергать H0 - распределение можно считать нормальным")
    else:
        print(" Вывод: Отвергаем H0 - распределение значимо отличается от нормального")
    print()
```

ЛИСТИНГ В

Листинг интервальных оценок числовых характеристик данных

```
def section_1_6_interval_estimates(self):
    print("ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ЧИСЛОВЫХ ХАРАКТЕРИСТИК ДАННЫХ")
    print("=" * 50)
    mean = np.mean(self.doctors_data)
    std = np.std(self.doctors_data, ddof=1)
    n = len(self.doctors_data)
    print(f"Исходные данные: n = {n},  $\bar{x}$  = {mean:.0f}, s = {std:.0f}")
    # Доверительный интервал для математического ожидания (нормальное распределение)
    t_critical = stats.t.ppf(0.975, df=n-1)
    margin_error = t_critical * (std / np.sqrt(n-1))
    ci_lower = mean - margin_error
    ci_upper = mean + margin_error
    print("\nДоверительный интервал для математического ожидания (95%):")
    print(f"Предположение: нормальное распределение")
    print(f"t-критическое ({n-1} степеней свободы): {t_critical:.4f}")
    print(f"Интервал: [{ci_lower:.0f}; {ci_upper:.0f}] чел.")
    # Доверительный интервал для СКО (нормальное распределение)
    chi2_lower = stats.chi2.ppf(0.025, df=n-1)
    chi2_upper = stats.chi2.ppf(0.975, df=n-1)
    std_lower = std * np.sqrt(n) / np.sqrt(chi2_upper)
    std_upper = std * np.sqrt(n) / np.sqrt(chi2_lower)
    print("\nДоверительный интервал для СКО (95%):")
    print(f"Предположение: нормальное распределение")
    print(f" $\chi^2$ -критические: [{chi2_lower:.1f}; {chi2_upper:.1f}]")
    print(f"Интервал: [{std_lower:.0f}; {std_upper:.0f}] чел.")
```

ЛИСТИНГ Г

Листинг с оценкой дополнительных данных, связанными с изучаемыми данными

```
def section_2_2_additional_data_analysis(self, chars: Dict):
    """Оценка дополнительных данных"""
    print("ОЦЕНКА ДОПОЛНИТЕЛЬНЫХ ДАННЫХ")
    print("=" * 50)
    # 1. Сравнение 2013 и 2023 годов
    print("СРАВНЕНИЕ 2013-2023:")
    print("-" * 30)
    self.simple_year_comparison()
    # 2. Анализ по типам регионов
    print("\nТИПЫ РЕГИОНОВ:")
    print("-" * 30)
    self.simple_region_types()
    # 3. Топ-5 регионов
    print("\nТОП-5 РЕГИОНОВ:")
    print("-" * 30)
    self.simple_top_regions()
    print()

def simple_year_comparison(self):
    """Простое сравнение 2013 и 2023 годов"""
    if '2013' not in self.data.columns:
        print(" Нет данных за 2013 год")
        return
    try:
        # Берем только общие регионы
        common_data = self.regions_data[['region', '2013', '2023']].dropna()
        if len(common_data) == 0:
            print(" Нет данных для сравнения")
            return
        data_2013 = common_data['2013'].values
        data_2023 = common_data['2023'].values
        mean_2013 = np.mean(data_2013)
        mean_2023 = np.mean(data_2023)
        growth = mean_2023 - mean_2013
        growth_percent = (growth / mean_2013) * 100
        print(f" Средняя численность:")
        print(f" - 2013: {mean_2013:.0f} чел.")
        print(f" - 2023: {mean_2023:.0f} чел.")
        print(f" - Изменение: {growth:+.0f} чел. ({growth_percent:+.1f}%)")
        # Считаем сколько регионов показали рост
        growth_count = np.sum(data_2023 > data_2013)
        growth_percent_regions = (growth_count / len(common_data)) * 100
        print(f" Рост в {growth_count} регионах ({growth_percent_regions:.1f}%)")
    except Exception as e:
        print(f" Ошибка: {e}")

def simple_region_types(self):
```

```

"""Простой анализ по типам регионов"""
types = {
    'Республики': 'Республика',
    'Края': 'край',
    'Области': 'область',
    'Города': 'Москва|Санкт-Петербург|Севастополь',
    'Автономные округа': 'автономный'
}
for type_name, pattern in types.items():
    regions = self.regions_data[
        self.regions_data['region'].str.contains(pattern, na=False)
    ]
    if len(regions) > 0:
        values = regions[self.year_column].values
        mean = np.mean(values)
        count = len(values)
        print(f" • {type_name} ({count}): {mean:.0f} чел.")

def simple_top_regions(self):
    """Простой анализ топ-5 регионов"""
    top_5 = self.regions_data.nlargest(5, self.year_column)
    total = np.sum(self.doctors_data)
    for _, row in top_5.iterrows():
        share = (row[self.year_column] / total) * 100
        print(f" • {row['region']}: {row[self.year_column]:.0f} чел. ({share:.1f}%)")

```

ЛИСТИНГ Д

Листинг расчета дополнительных числовых показателей

```
def calculate_additional_characteristics(self, basic_chars=None):
    print("ДОПОЛНИТЕЛЬНЫЕ ВЫБОРОЧНЫЕ ОЦЕНКИ")
    print("=" * 50)
    chars = {}
    if basic_chars is None:
        basic_chars = self.calculate_basic_characteristics()
    # Выборочная медиана
    chars['median'] = np.median(self.doctors_data)
    print(f"Выборочная медиана: {chars['median']:.2f} чел.")
    # Выборочное абсолютное отклонение
    chars['mad'] = np.mean(np.abs(self.doctors_data - chars['median']))
    print(f"Выборочное среднее абсолютное отклонение: {chars['mad']:.2f} чел.")
    # Квартили
    chars['q1'] = np.percentile(self.doctors_data, 25)
    chars['q2'] = np.percentile(self.doctors_data, 50)
    chars['q3'] = np.percentile(self.doctors_data, 75)
    print(f"Выборочные квартили:")
    print(f" Q1 (25%): {chars['q1']:.2f} чел.")
    print(f" Q2 (50%, медиана): {chars['q2']:.2f} чел.")
    print(f" Q3 (75%): {chars['q3']:.2f} чел.")
    # Интерквартильная широта
    chars['iqr'] = chars['q3'] - chars['q1']
    print(f"Интерквартильная широта: {chars['iqr']:.2f} чел.")
    # Полусумма выборочных квартилей
    chars['midquartile'] = (chars['q1'] + chars['q3']) / 2
    print(f"Полусумма выборочных квартилей: {chars['midquartile']:.2f} чел.")

    # Разность выборочных квартилей
    chars['quartile_diff'] = chars['q3'] - chars['q1']
    print(f"Разность выборочных квартилей: {chars['quartile_diff']:.2f} чел.")
    # Экстремальные элементы
    chars['min'] = np.min(self.doctors_data)
    chars['max'] = np.max(self.doctors_data)
    min_region = self.regions_data[self.regions_data[self.year_column] ==
    chars['min']][['region']].iloc[0]
    max_region = self.regions_data[self.regions_data[self.year_column] ==
    chars['max']][['region']].iloc[0]
    print(f"Экстремальные элементы:")
    print(f" Минимум: {chars['min']:.0f} чел. ({min_region})")
    print(f" Максимум: {chars['max']:.0f} чел. ({max_region})")
    # Размах выборки
    chars['range'] = chars['max'] - chars['min']
    print(f"Размах выборки: {chars['range']:.2f} чел.")

    # Полусумма экстремальных элементов
    chars['midrange'] = (chars['min'] + chars['max']) / 2
    print(f"Полусумма экстремальных элементов: {chars['midrange']:.2f} чел.")
    # Выборочная оценка асимметрии
    chars['skewness'] = basic_chars['m3'] / (basic_chars['std_biased']**3)
```

```
print(f"Выборочная оценка асимметрии: {chars['skewness']:.4f}")
# Выборочная оценка эксцесса
chars['kurtosis'] = (basic_chars['m4'] / (basic_chars['std_biased']**4)) - 3
print(f"Выборочная оценка эксцесса: {chars['kurtosis']:.4f} ")
print()
return char
```