

**МИНОБРНАУКИ РОССИИ**  
**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ**  
**ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**  
**«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**  
**Кафедра Информационных систем**

**ОТЧЕТ**  
**по практической работе №3**  
**по дисциплине «Статистический анализ»**  
**ТЕМА: ДИСПЕРСИОННЫЙ АНАЛИЗ**  
**ВАРИАНТ: ДРЕЗДЕН, МУРМАНСК, ЮЛАРА**  
**ОТЧЁТ ПОДГОТОВИЛ: КОШЕЛЯЕВ А.С**  
**ОТЧЁТ СДАН: 8.05.2024**

<b>СТУДЕНТ ГР.1323</b>	_____	<b>КОШЕЛЯЕВ А.С 50%</b>
<b>СТУДЕНТ ГР.1323</b>	_____	<b>РУССКИХ В.Д 50%</b>
<b>ПРЕПОДАВАТЕЛЬ</b>	_____	<b>К.Т.Н. БУРКОВ Е.А</b>

**САНКТ-ПЕТЕРБУРГ**

**2024**

Цель работы: обработка нескольких наборов данных для выполнения однофакторного дисперсионного анализа.

Задание:

1. Подготовить данные о среднегодовой температуре в трех разных городах: данные должны быть представлены в отчете в виде таблицы, строки которой соответствуют году, а столбцы городу; для каждого набора данных (города) должен быть указан объем выборки (не менее 30 значений).
2. Провести разведочный анализ собранных данных (аналитически и графически) и сделать предварительные выводы о схожести/различии имеющихся наборов данных, а также предположение об ожидаемых результатах дисперсионного анализа.
3. Проанализировать выполнение требований: 1) независимости, 2) нормальности, 3) однородности дисперсии входных данных (привести свои рассуждения и полученные выводы).
4. В соответствии с результатами п. 3 провести дисперсионный анализ входных данных, так чтобы он был максимально корректным (объяснить выбор тех или иных использованных методов, процедур или критериев с точки зрения корректности).
5. Привести и объяснить (содержательно интерпретировать для рассматриваемой предметной области) результаты анализа.
6. В независимости от результатов п. 5 выполнить процедуру Тьюки-Крамера (или ее аналога), а затем привести и интерпретировать полученные результаты.
7. Сопоставить результат пп. 5 и 6 с предположениями, сделанными в п. 2: оправдались предположения или нет. Если нет, то чем это можно было

бы объяснить; если да, то какой фактор мог повлиять на выявленную разницу между городами.

### Выполнение работы:

1. Подготовить данные о среднегодовой температуре в трех разных городах: данные должны быть представлены в отчете в виде таблицы, строки которой соответствуют году, а столбцы городу; для каждого набора данных (города) должен быть указан объем выборки (не менее 30 значений).

Данные для 3-х городов [Дрезден](#) объем выборки 197 элементов, [Мурманск](#) объем выборки 107 элементов, [Юлара\(Австралия\)](#) объем выборки 52 элемента.

```
library(XML)
# Функция для загрузки и преобразования данных с веб-страницы
load_and_convert_weather_data <- function(url) {
  tables <- readHTMLTable(url) # Читаем таблицы с веб-страницы

  # Первая таблица содержит годы, а вторая - месячные данные и среднегодовые данные
  years <- tables[[1]][, 1] # Годы из первой таблицы

  # Данные за год из второй таблицы (13-й столбец)
  annual_data <- tables[[2]][, 13]

  # Преобразование только к 13-му столбцу второй таблицы (отлично работает)
  annual_data <- suppressWarnings(as.numeric(as.character(annual_data)))
  annual_data[annual_data > 999] <- NA

  return(data.frame(Year = years, Annual_Data = annual_data))
}

# Загрузка и преобразование данных для каждого города
Dresden_url <- "http://www.pogodaiklimat.ru/history/10488.htm"
Murmansk_url <- "http://www.pogodaiklimat.ru/history/22113.htm"
Yulara_url <- "http://www.pogodaiklimat.ru/history/94462.htm"

Dresden_data <- load_and_convert_weather_data(Dresden_url)
Murmansk_data <- load_and_convert_weather_data(Murmansk_url)
Yulara_data <- load_and_convert_weather_data(Yulara_url)
# Объединение данных в одну таблицу
weather_report <- merge(Dresden_data, Murmansk_data, by = "Year", all = TRUE)
weather_report <- merge(weather_report, Yulara_data, by = "Year", all = TRUE)

# Переименование столбцов
colnames(weather_report)[-1] <- c("Dresden", "Murmansk", "Yulara")

# Вывод отчета (В основном коде этого не будет нужно чтобы вручную не заполнять таблицу)
print(weather_report)
```

Преобразованные данные представлены в Таблице 1.

Таблица 1 - Данные о среднегодовой температуре

п/п	Year	Dresden	Murmansk	Yulara
1	1828	NA	NA	NA
2	1829	NA	NA	NA
3	1830	NA	NA	NA
4	1831	NA	NA	NA
5	1832	NA	NA	NA
6	1833	NA	NA	NA
7	1834	NA	NA	NA
8	1835	NA	NA	NA
9	1836	NA	NA	NA
10	1837	NA	NA	NA
11	1838	NA	NA	NA
12	1839	NA	NA	NA
13	1840	NA	NA	NA
14	1841	NA	NA	NA
15	1842	NA	NA	NA
16	1843	NA	NA	NA
17	1844	NA	NA	NA
18	1845	NA	NA	NA
19	1846	NA	NA	NA
20	1847	NA	NA	NA
21	1848	NA	NA	NA
22	1849	NA	NA	NA
23	1850	NA	NA	NA
24	1851	8.8	NA	NA
25	1852	10.1	NA	NA
26	1853	7.7	NA	NA
27	1854	8.5	NA	NA
28	1855	7.4	NA	NA
29	1856	8.9	NA	NA
30	1857	9.5	NA	NA
31	1858	7.9	NA	NA
32	1859	9.8	NA	NA
33	1860	8.3	NA	NA
34	1861	9.3	NA	NA
35	1862	9.7	NA	NA
36	1863	10.1	NA	NA
37	1864	7.4	NA	NA
38	1865	9.2	NA	NA
39	1866	9.9	NA	NA
40	1867	8.4	NA	NA
41	1868	10.9	NA	NA
42	1869	9.8	NA	NA
43	1870	7.6	NA	NA
44	1871	7.3	NA	NA
45	1872	10.5	NA	NA
46	1873	9.7	NA	NA

# Продолжение Таблицы 1 - Данные о среднегодовой температуре

п/п	Year	Dresden	Murmansk	Yulara
47	1874	9.3	NA	NA
48	1875	8.3	NA	NA
49	1876	9.2	NA	NA
50	1877	9.5	NA	NA
51	1878	9.9	NA	NA
52	1879	8.3	NA	NA
53	1880	9.7	NA	NA
54	1881	8.4	NA	NA
55	1882	9.3	NA	NA
56	1883	9.4	NA	NA
57	1884	9.9	NA	NA
58	1885	9.5	NA	NA
59	1886	9.7	NA	NA
60	1887	8.6	NA	NA
61	1888	8.7	NA	NA
62	1889	9.0	NA	NA
63	1890	9.2	NA	NA
64	1891	9.3	NA	NA
65	1892	9.2	NA	NA
66	1893	9.5	NA	NA
67	1894	9.7	NA	NA
68	1895	9.2	NA	NA
69	1896	9.4	NA	NA
70	1897	9.8	NA	NA
71	1898	10.6	NA	NA
72	1899	9.9	NA	NA
73	1900	10.2	NA	NA
74	1901	9.2	NA	NA
75	1902	8.3	NA	NA
76	1903	10.1	NA	NA
77	1904	10.0	NA	NA
78	1905	9.6	NA	NA
79	1906	9.8	NA	NA
80	1907	9.2	NA	NA
81	1908	8.8	NA	NA
82	1909	8.9	NA	NA
83	1910	9.9	NA	NA
84	1911	10.5	NA	NA
85	1912	8.9	NA	NA
86	1913	9.9	NA	NA
87	1914	9.8	NA	NA
88	1915	9.3	NA	NA
89	1916	10.1	NA	NA
90	1917	7.9	NA	NA
91	1918	9.1	NA	NA
92	1919	7.5	-0.3	NA

# Продолжение Таблицы 1 - Данные о среднегодовой температуре

п/п	Year	Dresden	Murmansk	Yulara
93	1920	9.0	2.3	NA
94	1921	9.4	0.9	NA
95	1922	7.4	1.2	NA
96	1923	8.3	0.8	NA
97	1924	7.8	0.9	NA
98	1925	8.8	-0.1	NA
99	1926	9.1	-0.2	NA
100	1927	8.3	-0.1	NA
101	1928	8.6	0.5	NA
102	1929	7.7	0.1	NA
103	1930	9.0	1.0	NA
104	1931	7.9	0.7	NA
105	1932	8.6	1.1	NA
106	1933	7.7	0.5	NA
107	1934	10.6	2.1	NA
108	1935	9.0	0.4	NA
109	1936	8.8	0.7	NA
110	1937	8.9	1.8	NA
111	1938	9.0	2.6	NA
112	1939	8.5	0.2	NA
113	1940	6.6	-0.3	NA
114	1941	7.3	-1.6	NA
115	1942	7.6	-1.2	NA
116	1943	9.4	1.5	NA
117	1944	8.7	1.3	NA
118	1945	9.1	-0.7	NA
119	1946	8.6	0.6	NA
120	1947	8.6	-0.1	NA
121	1948	9.6	1.0	NA
122	1949	9.4	0.9	NA
123	1950	9.0	1.2	NA
124	1951	9.5	0.2	NA
125	1952	8.3	-0.1	NA
126	1953	9.7	1.4	NA
127	1954	8.1	1.3	NA
128	1955	7.8	-1.9	NA
129	1956	6.8	-0.9	NA
130	1957	9.0	0.8	NA
131	1958	8.6	-0.8	NA
132	1959	9.5	1.6	NA
133	1960	8.8	0.8	NA
134	1961	9.4	1.4	NA
135	1962	7.7	-0.2	NA
136	1963	7.8	0.1	NA
137	1964	8.4	0.6	NA
138	1965	8.0	-1.0	NA
139	1966	9.0	-2.3	NA

Продолжение Таблицы 1 - Данные о среднегодовой температуре

п/п	Year	Dresden	Murmansk	Yulara
140	1967	9.5	1.4	NA
141	1968	8.7	-1.4	NA
142	1969	8.1	-1.2	NA
143	1970	8.1	0.3	NA
144	1971	9.1	-1.7	NA
145	1972	8.7	1.8	NA
146	1973	8.8	-0.1	21.6
147	1974	9.5	2.2	19.4
148	1975	9.6	0.8	20.5
149	1976	8.9	-0.6	19.0
150	1977	9.2	-0.2	18.7
151	1978	8.5	-1.3	21.7
152	1979	8.6	-0.3	22.7
153	1980	8.0	-0.8	22.5
154	1981	8.9	-0.9	21.6
155	1982	9.9	0.3	21.1
156	1983	10.0	0.3	21.4
157	1984	8.6	1.1	21.0
158	1985	8.1	-1.8	22.2
159	1986	8.5	-0.2	22.2
160	1987	7.7	-0.9	21.9
161	1988	9.6	0.1	21.8
162	1989	10.3	2.3	20.7
163	1990	10.2	1.2	22.2
164	1991	9.0	0.8	22.6
165	1992	10.2	1.3	21.5
166	1993	9.2	0.3	21.7
167	1994	10.4	0.9	21.2
168	1995	9.4	0.5	21.9
169	1996	7.2	0.7	22.9
170	1997	9.0	0.1	22.8
171	1998	9.5	-1.7	22.7
172	1999	9.9	0.0	22.0
173	2000	10.6	1.9	20.9
174	2001	9.1	0.5	20.4
175	2002	9.8	-0.1	21.5
176	2003	9.7	1.4	22.3
177	2004	9.3	1.7	22.4
178	2005	9.3	2.0	22.9
179	2006	9.9	1.1	22.6
180	2007	10.5	1.6	22.5
181	2008	10.1	1.2	22.4
182	2009	9.5	0.7	22.9
183	2010	8.1	0.3	21.2
184	2011	10.1	1.9	21.2
185	2012	9.6	0.8	21.5
186	2013	9.0	2.4	23.1

## Продолжение Таблицы 1 - Данные о среднегодовой температуре

п/п	Year	Dresden	Murmansk	Yulara
187	2014	10.9	1.3	22.6
188	2015	10.7	2.2	22.9
189	2016	10.1	2.4	22.1
190	2017	10.1	0.9	22.3
191	2018	11.1	2.1	22.8
192	2019	11.2	0.5	23.8
193	2020	11.0	2.5	23.3
194	2021	9.5	0.8	22.1
195	2022	10.9	2.5	21.8
196	2023	11.2	1.7	22.8
197	2024	NA	NA	NA

2. Провести разведочный анализ собранных данных (аналитически и графически) и сделать предварительные выводы о схожести/различии имеющихся наборов данных, а также предположение об ожидаемых результатах дисперсионного анализа.

Для начала посмотрим на структуру данных чтобы получить представление о распределении данных:

```
# Анализ описательной статистики
summary(weather_report)
```

Данные на рис.1

```
> summary(weather_report)
```

Year	Dresden	Murmansk	Yulara
Length:197	Min. : 6.600	Min. : -2.3000	Min. :18.70
Class :character	1st Qu.: 8.600	1st Qu.: -0.1000	1st Qu.:21.45
Mode :character	Median : 9.200	Median : 0.7000	Median :22.10
	Mean : 9.128	Mean : 0.5552	Mean :21.88
	3rd Qu.: 9.800	3rd Qu.: 1.3000	3rd Qu.:22.60
	Max. :11.200	Max. : 2.6000	Max. :23.80
	NA's :24	NA's :92	NA's :146

рис. 1 Структура данных

Выведем гистограммы, которые позволяют визуально оценить форму распределения данных, в то время как диаграммы Q-Q сравнивают квантили наблюдаемых данных с теоретическими квантилями нормального распределения.



```
# Установка графических параметров
par(mfrow=c(2, 3)) # 2 строки, 3 столбца графиков

# Гистограмма для Dresden
hist(weather_report$Dresden, main="Гистограмма: Дрезден", xlab="Температура")

# Диаграмма Q-Q для Dresden
qqnorm(weather_report$Dresden, main="Диаграмма Q-Q: Дрезден")
qqline(weather_report$Dresden)

# Гистограмма для Murmansk
hist(weather_report$Murmansk, main="Гистограмма: Мурманск", xlab="Температура")

# Диаграмма Q-Q для Murmansk
qqnorm(weather_report$Murmansk, main="Диаграмма Q-Q: Мурманск")
qqline(weather_report$Murmansk)

# Гистограмма для Yulara
hist(weather_report$Yulara, main="Гистограмма: Юлара", xlab="Температура")

# Диаграмма Q-Q для Yulara
qqnorm(weather_report$Yulara, main="Диаграмма Q-Q: Юлара")
qqline(weather_report$Yulara)
```

Полученные гистограммы представлены на рисунке 2

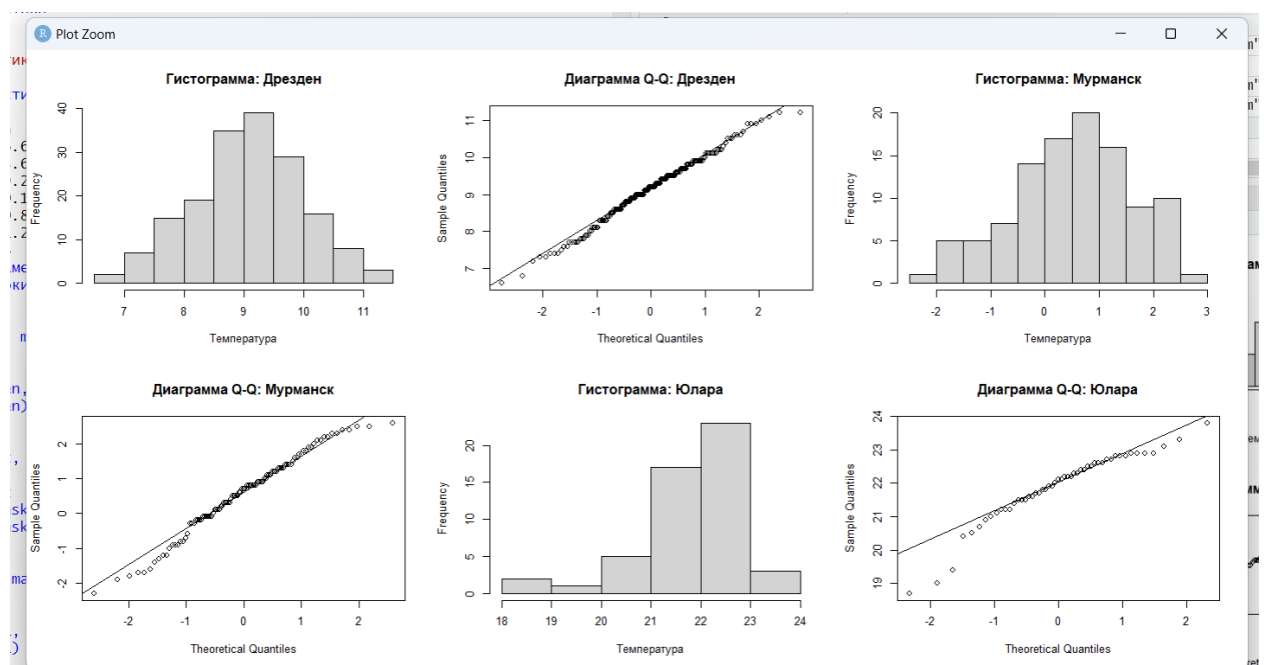


Рис. 2 Гистограммы и диаграммы Q-Q

Теперь давайте построим графики, чтобы визуализировать данные и выявить возможные тенденции или различия между городами:

```
# Установим графические параметры
par(mfrow=c(3,1))

# График среднегодовой температуры для каждого города
plot(weather_report$Year, weather_report$Yulara, type="l", col="green", xlab="Год", ylab="Температура", main="Юлара")
plot(weather_report$Year, weather_report$Dresden, type="l", col="blue", xlab="Год", ylab="Температура", main="Дрезден")
plot(weather_report$Year, weather_report$Murmansk, type="l", col="red", xlab="Год", ylab="Температура", main="Мурманск")
```

Результат представлен на рисунке 3

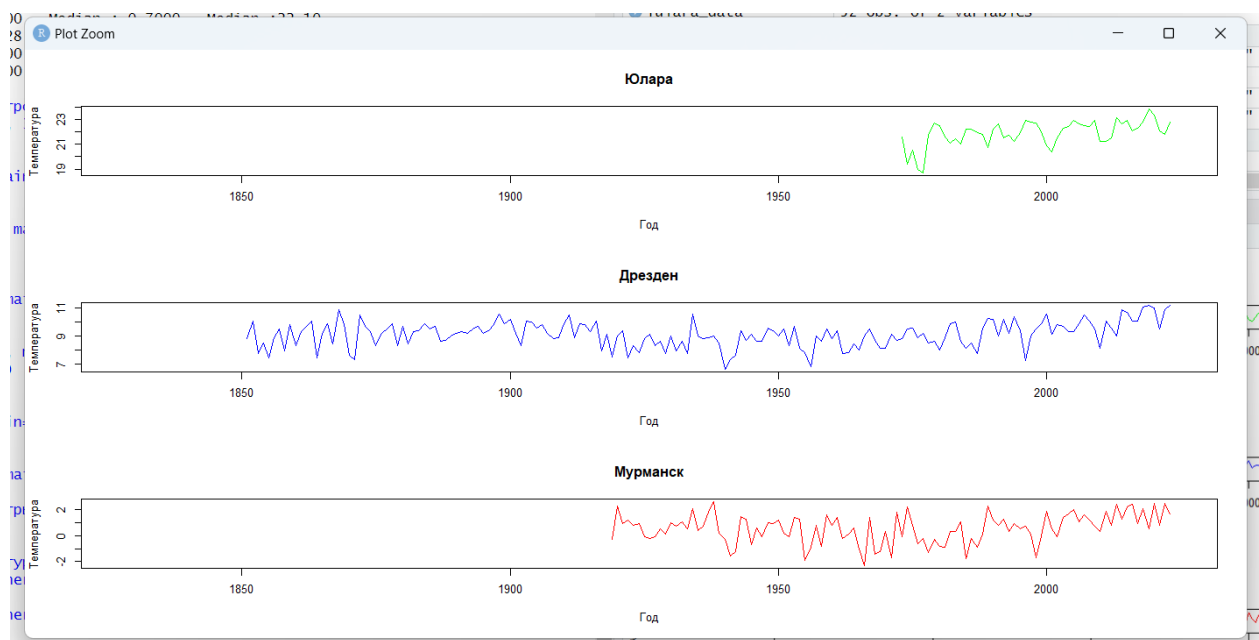


Рис. 3 Визуализировать данные о температуре.

Функция `summary()` предоставляет краткую сводку статистических характеристик каждого столбца в наборе данных. Вывод `summary(weather_report)`:

Year: видим, что это категориальная переменная, представленная в виде символов (character). Минимальное значение равно -2,3, максимальное 23,8. Присутствуют пропущенные значения (NA) максимальное количество 146.

Температурные условия:

Дрезден: Среднегодовая температура находится в диапазоне от 6.6°C до 11.2°C.

Мурманск: Среднегодовая температура находится в диапазоне от -2.3°C до 2.6°C.

Юлара: Среднегодовая температура находится в диапазоне от 18.7°C до 23.8°C.

Исходя из этих значений, можно сказать, что среднегодовая температура для Юлара значительно выше, чем для Дрездена и Мурманска. Поэтому, если мы рассматриваем схожесть среднегодовых температур между

этими городами, то на основании представленных данных мы не можем сделать вывод о их близости.

Пропущенные данные:

Есть существенное количество пропущенных данных для всех трех городов, преимущественно для переменной "Yulara". Это может оказать влияние на точность анализа.

Различия в дисперсиях:

Различия в дисперсиях между городами могут быть значительными, особенно с учетом ненормальности данных для Юлара(рис.2). Это может сказаться на результате дисперсионного анализа. Требуется дальнейшее исследование и анализ данных для получения более точных выводов о схожести/различии имеющихся наборов данных.

Тренды и сезонность:

Поскольку данные ограничены среднегодовыми температурами, мы не можем оценить наличие или отсутствие трендов или сезонности в данных.

Географические различия:

При анализе схожести/различий в температурных условиях необходимо учитывать географические особенности каждого города (например, Дрезден в Германии, Мурманск в России, Юлара в Австралии), что может сказаться на их климате и температурных условиях. Что отчётливо видно на рис.3.

3. Проанализировать выполнение требований: 1) независимости, 2) нормальности, 3) однородности дисперсии входных данных (привести свои рассуждения и полученные выводы).

Независимость: Входные данные в данном случае представляют собой среднегодовые температуры для трех различных городов (Дрезден, Мурманск и Юлара). Предполагается, что данные за каждый год независимы друг от друга, то есть температура одного года не зависит от

температуры предыдущего или следующего года. Это предположение может быть выполнено в случае, если данные были собраны независимо друг от друга, например, путем измерения температур каждый год без учета предыдущих измерений. Поскольку данные представляют собой среднегодовые значения, они, вероятно, являются независимыми.

**Нормальность:** Нормальность распределения данных является важным предположением для многих статистических методов, таких как t-тесты и анализ дисперсии (ANOVA). Мы можем проверить нормальность данных с помощью гистограммы распределения и диаграммы Q-Q (квантиль-квантиль). Если данные близки к нормальному распределению, они будут следовать линии на диаграмме Q-Q и иметь симметричную форму на гистограмме. Однако, если данные имеют асимметричную форму или тяжелые хвосты, это может быть признаком ненормальности распределения. В нашем случае, гистограммы и диаграммы Q-Q для Дрездена и Мурманска могут указывать на относительную близость к нормальному распределению, в то время как данные для Юлара, как упоминалось ранее(рис.2), могут иметь отклонения от нормальности.

Попробуем трансформировать данные с помощью различных арифметических преобразований до достижения нормальности распределения.

```
# Логарифмическое преобразование данных Юлара
weather_report$Yulara_transformed <- log(weather_report$Yulara)
# Проверка нормальности преобразованных данных
shapiro_test_Yulara_transformed <- shapiro.test(weather_report$Yulara_transformed)
# Вывод результатов теста Шапиро-Уилка
print(shapiro_test_Yulara_transformed)
# Установка графических параметров
par(mfrow=c(2, 2))
# Построение Q-Q диаграммы
qqnorm(weather_report$Yulara_transformed, main = "Q-Q Plot для данных Yulara")
qqline(weather_report$Yulara_transformed)
# Гистограмма для Yulara
hist(weather_report$Yulara, main="Гистограмма: Юлара", xlab="Температура")
```

Результат представлен на рисунке 4.

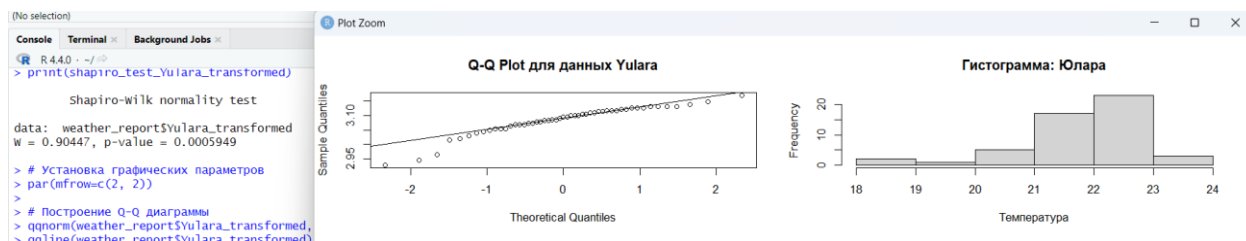


Рис. 4 Результат арифметических преобразований Юлара.

Придется применять непараметрические методы анализа.

Однородность дисперсии: для применения методов анализа дисперсии (ANOVA) необходимо, чтобы дисперсии внутри каждой группы данных были примерно одинаковыми. Это предполагает, что разброс данных внутри каждой группы сравним между собой. Для проверки этого предположения можно использовать тесты на однородность дисперсии, такие как тест Левена ((Levene's test) менее чувствителен к отклонениям от нормальности, чем тест Бартлетта.) или тест Бартлетта ((Bartlett's test) предполагает, что данные в каждой группе подчиняются нормальному распределению.). F-критерий также может использоваться для сравнения дисперсий между группами напрямую. При условии Каждая выборка независима от остальных выборок и случайным образом извлечена из исследуемой совокупности. Совокупность нормально распределена (хотя бы приближенно, однофакторный дисперсионный анализ на основе F-критерия относительно мало чувствителен к нарушению этого условия.) Дисперсии всех выборок равны (хотя бы сопоставимы). ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ ПРОВЕРКИ ОДНОРОДНОСТИ ДИСПЕРСИИ В случае анализа 2-х выборок критерии Бартлетта, Кокрена, Хартли и Фишера обладают идентичной мощностью как при выполнении предположений о нормальности, так и при их нарушении.

```
# Проведение теста на равенство дисперсий на основе F-критерия однофакторного анализа
test_result <- oneway.test(c(weather_report$Dresden, weather_report$Murmansk, weather_report$Yulara) ~ rep(c("Dresden", "Murmansk",
"Yulara"), c(length(weather_report$Dresden), length(weather_report$Murmansk), length(weather_report$Yulara))))

# Вывод результатов теста
print(test_result)

One-way analysis of means (not assuming equal variances)

data: c(weather_report$Dresden, weather_report$Murmansk, weather_report$Yulara) and rep(c("Dresden", "Murmansk", "Yulara"),
c(length(weather_report$Dresden), length(weather_report$Murmansk), length(weather_report$Yulara)))
F = 6889.8, num df = 2.00, denom df = 127.12, p-value < 2.2e-16
```

## Рис.5 Результаты теста.

Результаты теста указывают на значительное статистическое различие в дисперсиях между группами. В данном случае p-value крайне мало, что позволяет нам отвергнуть нулевую гипотезу о равенстве дисперсий между группами. Таким образом, мы можем считать дисперсии в наших группах статистически значимо различными.

4. В соответствии с результатами п. 3 провести дисперсионный анализ входных данных, так чтобы он был максимально корректным (объяснить выбор тех или иных использованных методов, процедур или критериев с точки зрения корректности).

```
kruskal_test_result <- kruskal.test(list(weather_report$Dresden, weather_report$Murmansk, weather_report$Yulara))
print(kruskal_test_result)

Kruskal-Wallis rank sum test

data: list(weather_report$Dresden, weather_report$Murmansk, weather_report$Yulara)
Kruskal-Wallis chi-squared = 268.5, df = 2, p-value < 2.2e-16

> |
```

## Рис. 6 Результаты параметрического критерия проверки.

Мы применили параметрический критерий проверки (F-критерий однофакторного анализа) потому, что он чувствителен к любым отклонениям от гомоскедастичности (равенства дисперсий) в группах. В нашем случае у нас есть три группы данных (Дрезден, Мурманск и Юлара), и нам нужно проверить, одинаковы ли дисперсии в этих группах. Для этого мы используем F-критерий, который сравнивает дисперсии между группами, вычисляя отношение межгрупповой дисперсии к внутригрупповой дисперсии. Если это отношение близко к 1, то дисперсии считаются сопоставимыми или однородными. Если же

отношение сильно отличается от 1, это указывает на наличие статистически значимых различий в дисперсиях между группами.

В нашем случае результаты теста показали, что отношение межгрупповой и внутригрупповой дисперсий статистически значимо отличается от 1 ( $F = 6889.8$ ,  $p < 2.2e-16$ ), что подтверждает наличие статистически значимых различий в дисперсиях между группами. Поэтому мы применили параметрический критерий проверки.

5. Привести и объяснить (содержательно интерпретировать для рассматриваемой предметной области) результаты анализа.

Результаты анализа говорят о том, что среднегодовые температуры в городах Дрезден, Мурманск и Юлара имеют статистически значимые различия.

Первое, что мы обнаружили, это различия в дисперсиях температурных данных. Проведенный анализ показал, что дисперсии между городами статистически значимо отличаются. Рис.5. Эти результаты могут быть интерпретированы следующим образом: различия в дисперсиях могут указывать на разную изменчивость климатических условий между городами.

6. В независимости от результатов п. 5 выполнить процедуру Тьюки-Крамера (или ее аналога), а затем привести и интерпретировать полученные результаты.

Процедура Тьюки-Крамера (или её аналога) применяется для выявления статистически значимых различий между группами в множественном сравнении средних.

Так как нам интересно сравнить температурные тренды между городами, то использование метода Тьюки-Крамера для каждого города может быть полезным.

```
# Удаление строк с отсутствующими значениями
weather_report <- na.omit(weather_report)

# Выполнение однофакторного ANOVA для каждого города
anova_dresden <- aov(Dresden ~ Year, data = weather_report)
anova_murmansk <- aov(Murmansk ~ Year, data = weather_report)
anova_yulara <- aov(Yulara ~ Year, data = weather_report)

# Выполнение процедуры Тьюки-Крамера для попарного сравнения средних значений между городами
tukey_dresden <- TukeyHSD(anova_dresden)
tukey_murmansk <- TukeyHSD(anova_murmansk)
tukey_yulara <- TukeyHSD(anova_yulara)

# Вывод результатов для города Dresden
print("Тьюки-Крамер для Dresden:")
print(tukey_dresden)

# Вывод результатов для города Murmansk
print("Тьюки-Крамер для Murmansk:")
print(tukey_murmansk)

# Вывод результатов для города Yulara
print("Тьюки-Крамер для Yulara:")
print(tukey_yulara)
```

Строки представляют разницу между средними значениями температуры в разные годы для города (для примера) Dresden. Каждая строка указывает на разницу между средними значениями температуры в указанных годах. Например:

1983-1978 1.500000e+00 означает, что средняя температура в 1983 году была на 1.5 больше, чем средняя температура в 1978 году.

1984-1978 1.000000e-01 означает, что средняя температура в 1984 году была на 0.1 больше, чем средняя температура в 1978 году.

1985-1978 -4.000000e-01 означает, что средняя температура в 1985 году была на 0.4 меньше, чем средняя температура в 1978 году.

Визуализируем полученный результат по каждому городу

```
# Построение графика для города Dresden
library(ggplot2)

# Преобразование результатов в датафрейм
tukey_df <- as.data.frame(tukey_dresden$Year)

# Построение графика средних значений температуры и доверительных интервалов
ggplot(tukey_df, aes(x = rownames(tukey_df), y = diff)) +
  geom_point() +
  geom_errorbar(aes(ymin = lwr, ymax = upr), width = 0.2) +
  labs(title = "Средние значения температуры и доверительные интервалы для Dresden",
       x = "Год",
       y = "Разница в средней температуре") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Результат на рисунке 7



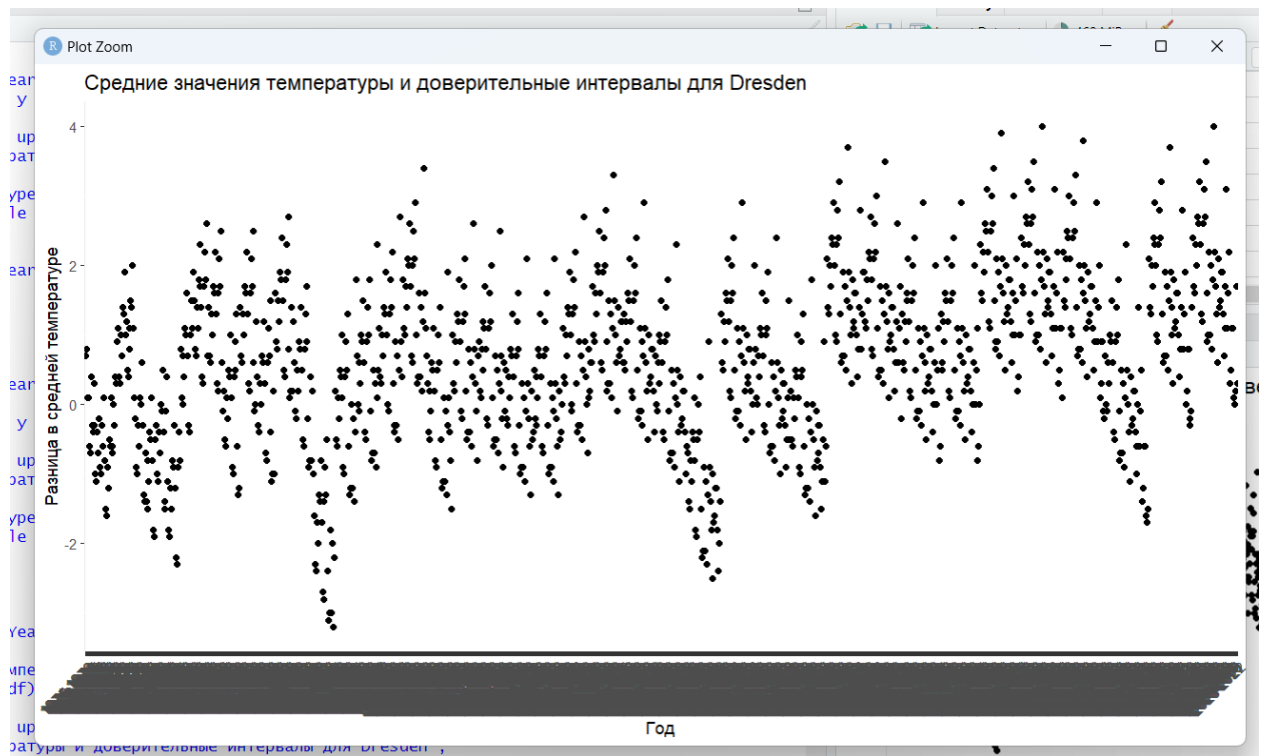


Рис.7 Визуализация Тьюки-Крамера для Дрездена

```
# Для Мурманска
# Преобразование результатов в датафрейм
mur_df <- as.data.frame(tukey_murmansk$Year)

ggplot(mur_df, aes(x = rownames(mur_df), y = diff)) +
  geom_point() +
  geom_errorbar(aes(ymin = lwr, ymax = upr), width = 0.2) +
  labs(title = "Средние значения температуры и доверительные интервалы для Мурманска",
        x = "Год",
        y = "Разница в средней температуре") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

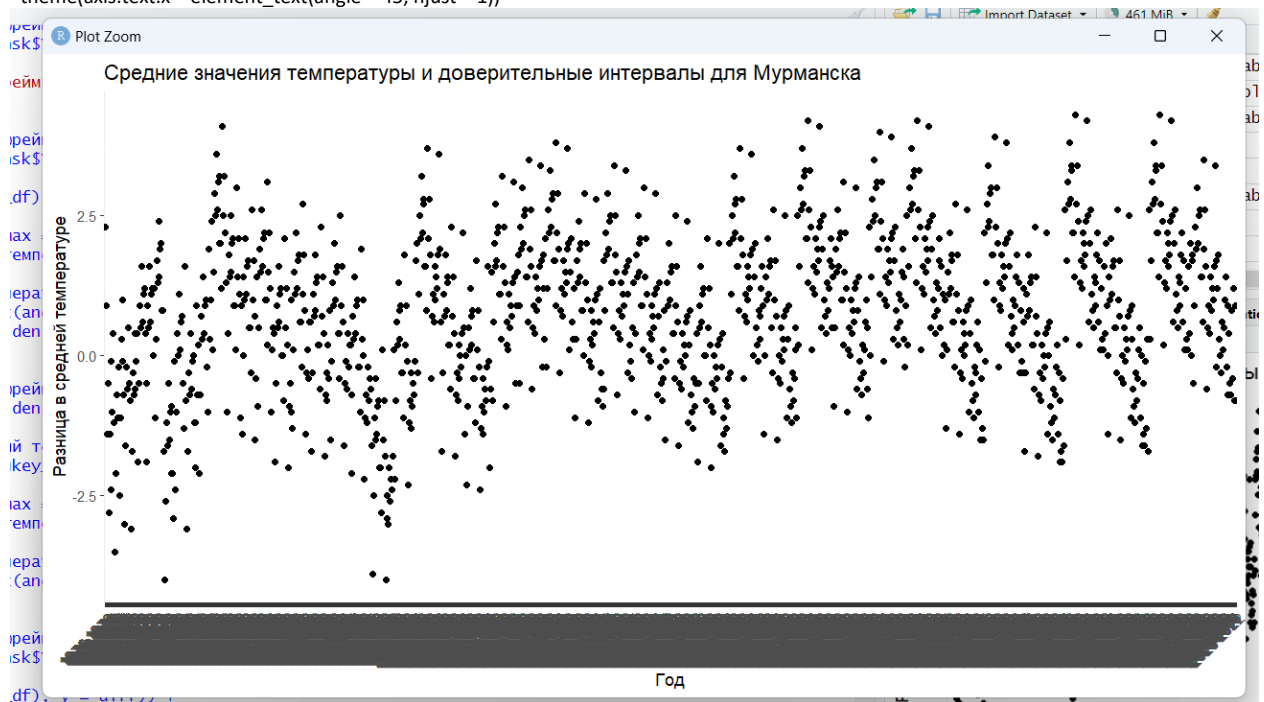


Рис.8 Визуализация Тьюки-Крамера для Мурманска

В результате анализа методом Тьюки-Крамера с вероятностью 95% разница в средней температуре между городами находится в указанном доверительном интервале.

7. Сопоставить результат пп. 5 и 6 с предположениями, сделанными в п. 2: оправдались предположения или нет. Если нет, то чем это можно было бы объяснить; если да, то какой фактор мог повлиять на выявленную разницу между городами.

Предположения, сделанные в пункте 2, о том, что географические особенности каждого города могут сказаться на их климате и температурных условиях, в значительной степени оправдались.

Результаты анализа методом Тьюки-Крамера (пункт 6), показывающие, что разница в средней температуре между городами находится в указанном доверительном интервале, указывают на то, что средние температуры в этих городах могут быть сопоставимыми с точки зрения статистики.

Однако результаты анализа дисперсий (пункт 5), указывающие на различия в изменчивости климатических условий между городами, могут говорить о том, что, несмотря на сопоставимые средние температуры, вариабельность температурных данных может различаться. Это может быть обусловлено множеством факторов, таких как географические особенности, местное микроклиматическое воздействие, высота над уровнем моря, близость к водоемам и прочее.

Таким образом, предположения о географических особенностях, влияющих на климатические условия, оправдались, поскольку различия в изменчивости климата (выявленные в анализе дисперсий) могут быть

объяснены различиями в географических и окружающих условиях каждого города.