МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ

ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра ИТиСБ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

(магистерская работа)

На тему Методика непараметрического анализа статистической однородности и связи экологических показателей водных объектов

Исполнитель Жильчук Анна Ивановна

(фамилия, имя, отчество)

Руководитель доктор технических наук, профессор

(ученая степень, ученое звание)

Завгородний Владимир Николаевич

(фамилия, имя, отчество)

«К защите допускаю»

Заведующий кафедрой

(подпись)

(ученая степень, ученое звание)

(фамилия, имя, отчество)

«__»_____2023 г.

Санкт-Петербург

2023

Реферат

Тема выпускной квалификационной работы: «Методика непараметрического анализа статистической однородности и связи экологических показателей воды в районах морской деятельности».
Работа содержит 58 страницу, 37 рисунков, 13 таблиц. При написании работы использовалось 16 литературных источников.
Объект исследования: данные экологических измерений реки Охта
Предмет исследования: статистические непараметрические критерии для оценки однородности и связи

В выпускной квалификационной работе проводится разработка методики непараметрического анализа однородностей и связей на примере результатов экологических измерений реки Охта. Проведены расчеты непараметрических критериев на примере временных рядов измерений экологических показателей.

Оглавление

Введение

С 2011 года река Охта стала относиться к категории загрязненных рек. На сегодняшний день это одна из самых загрязненных рек в Санкт-Петербурге и Ленинградской области и ее состояние не улучшается, поэтому вопрос организации системы мониторинга и анализа измеренных данных для нее стоит особенно остро. Более того, Охта является крупным притоком Невы.

Первые методики по разработке систем мониторинга систем водных ресурсов стали появляться в начале второй половины двадцатого века. В настоящее время разработаны методики по наблюдению за водными объектами и также существует ряд мер по их очищению. Связующим звеном между этими этапами является анализ данных, который позволяет обнаружить причины загрязнения и спрогнозировать дальнейшие изменения состояния.

Целью данной работы является анализ применение непараметрических методик анализа для оценки однородностей и связей для данных измерений реки Охта, проведенных студентами экологического факультета РГГМУ.

В работе решаются следующие задачи:

1. Проводится анализ возможностей применения непараметрических критериев для исследования однородностей и связей экологических явлений
2. Рассматриваются возможности ППП Anaconda относительно анализа экологических данных
3. Проводится расчет статистических критериев на данных измерений реки Охта

Глава 1. Возможности применения непараметрической и параметрической статистики для исследования однородности и связей экологических явлений на примере данных реки Охта

Система мониторинга качества воды предназначена для сбора данных для последующей её обработки для принятия управленческих решений и восстановления полной информации об объекте наблюдения.

В 1960-х и 1970-х годах началась разработка систем мониторинга качеством воды для оценки общего состояния воды. В них, как правило, не использовались какие-либо стратегии и методики проектирования. Неполноценность методической системы часто приводила к сбору данных качества воды без должного анализа или конечной цели. Места и частота измерения данных зачастую определялись удобством или другими субъективными критериями, а также не проводилась последующая оценка эффективности системы наблюдения.

На сегодняшний день Международная Метеорологическая организация (WMO) периодически выпускает руководства по разработке систем наблюдения за водными объектами. Также существует Water Framework Directive - директива Евросоюза, включающая в себя так же руководство по наблюдению за реками [1].

В общем случае система мониторинга качества воды может содержать огромное число этапов, которые в упрощенном виде представлены ниже:

1. Разработка системы наблюдения
    1.1. Местоположение станций
    1.2. Измеряемые данные
    1.3. Частота измерений
2. Сбор данных
    2.1. Методы сбора данных
    2.2. Единицы измерения
    2.3. Хранение данных
    2.4. Методы отбора проб

2.5. Перенос данных

3. Лабораторный анализ
   3.1. Техники анализа
   3.2. Эксплуатационные процедуры
   3.3. Контроль качества
   3.4. Запись данных
4. Обработка данных
   4.1. Прием данных
   4.2. Отбор и проверка
   4.3. Размещение и извлечение
   4.4. Отчетность
   4.5. Распространение
5. Анализ данных
   5.1. Основная статистика
   5.2. Регрессионный анализ
   5.3. Показатели качества
   5.4. Интерпретация “Контроля качества”
   5.5. Анализ временных рядов
   5.6. Модель качества воды
6. Использование информации для принятия решений
   6.1. Извлечение необходимой информации
   6.2. Формат отчетов
   6.3. Эксплуатационные процедуры
   6.4. Итоговая оценка

В системах мониторинга рек обычно используют следующие параметры:

- Основные показатели качества воды
  - Уровень воды
  - Содержание взвешенных веществ
  - Температура
  - pH
  - Электрическая проводимость

  - Растворенный кислород
  - Прозрачность
- Растворенные вещества
  - Кальций
  - Магний
  - Натрий
  - Калий
  - Хлориды
  - Фториды
  - Сульфаты
  - Щелочность
- Нутриенты
  - Нитраты и нитриты
  - Аммиак
  - Фосфаты
  - Силикаты
- Органические вещества
  - Хлорофилл

В системе мониторинга реки Охта используются в основном измерения растворенных веществ на водной поверхности и дне. В качестве постоянных мест для отбора проб использовались места с удобным доступом к воде (например мосты). Частота измерений приблизительно раз в год в первой половине июля.

Рисунок 1.1 Изображение реки Охта на карте с отмеченными станциями

1.2 Статистический анализ данных

Статистический анализ данных является ключевым в анализе данных измерений, поскольку эти данные зачастую должны соответствовать определенным статистическим критериям, к примеру случайность, независимость, однородность, стационарность [2].

Главным отличием непараметрических критериев является то, что для их применения не обязательно знать функцию распределения выборки. А поскольку согласно Центральной предельной теореме выборка достаточно большого размера, состоящая из случайных слабо зависимых величин, имеет функцию распределения, близкую к нормальной, то непараметрические критерии в основном используются для малых выборок. По этой же причине в качестве параметрических тестов чаще используются тесты, основанные на нормальном распределении. Также стоит отметить, что большинство непараметрических критериев является ранговыми, то есть для проверки

гипотезы используются порядковые номера величин, а не сами величины, поэтому имеется возможность использовать данные, представленные в любых шкалах измерения. За универсальность и простоту использования приходится расплачиваться: непараметрические критерии имеют меньшую мощность в сравнении с параметрическую и большую зависимость от табличных значений, из-за чего довольно затруднительно их использовать без пакетов прикладных программ.

Не стоит ожидать, что все величины из нескольких выборок будут совпадать. Тем не менее во многих случаях можно ожидать, что статистические параметры измерений будут такими же, либо отличаться на какое-то конкретное значение. Схожесть статистических параметров выборок называется однородностью. Чаще всего под однородностью понимаются однородности математического ожидания и дисперсии, иными словами, оцениваются различия мат. ожидания и дисперсии выборок. В работе будет проверяться как равенство или неравенство медиан с помощью непараметрических критериев Вилкоксона, Вилкоксона-Манна-Уитни, так и более общий случай однородности с помощью критерия Колмогорова-Смирнова для двух выборок. Тест Вилкоксона для связанных выборок и тест Вилкоксона-Манна-Уитни являются непараметрическими аналогами теста Стьюдента. [6]

Сегодня предполагается, что все явления в природе взаимосвязаны. Это касается и гидрологических и экологических явлений. К примеру, в данной работе производится анализ измерений биохимического потребления кислорода в реке Охта, которая связана с количеством кислорода, железа, фосфатов и нитритов в воде, что и будет проверено. Измерять мы будем корреляционную связь, смысл которой в обнаружении соответствия значений одной средней случайной величины средним значениям другой. То есть при изменении средней одной величины следует изменение среднего значение другого. Следует отличать статистическую и корреляционную связи: статистическая связь может означать закономерное изменение дисперсии,

эксцесса и так далее. Строгая функциональная связь, которая означает строгое соответствие одной величины значением одному или нескольким в природе не встречается из-за влияния случайных факторов на величины. Связь будет проверяться с помощью непараметрических критериев Кендалла и Спирмена, а также параметрического критерия согласия Пирсона.

1.3 Анализ статистических критериев

Имеется две выборки измерений жесткости воды на 1-3 станциях за 2006-2015 годы на уровне поверхности и дна:

Таблица 1.1 Примеры измерений жесткости воды

Номер станции	Измерения									
1	2.9	1.5	4.2	1.7	2.7	1.4	2.0	1.79	1.82	1.92
2	3.0	1.8	2.2	1.74	2.7	1.4	1.6	1.76	1.15	2.48

В примерах за нулевую гипотезу берется утверждение об отсутствии смещения проверяемого параметра, при альтернативной гипотезе о его наличии (двусторонний тест).

1.3.1 Критерий Вилкоксона-Манна-Уитни

Критерий Вилкоксона-Манна-Уитни используется для проверки однородности двух независимых выборок непрерывных величин x1, x2, x3, ..., xn и y1, y2, y3, ..., yn Нулевая гипотеза заключается в том, что разность медиан двух функций распределения равна нулю: Med X − Med Y = 0. Конкурирующие гипотезы:

$Med\ X - Med\ Y \neq 0$, $Med\ X - Med\ Y < 0$, $Med\ X - Med\ Y > 0$

Формула для расчета критерия:

$$U = n_1 * n_2 + \frac{n_1 * (n_1 + 1)}{2} - T\ \#(1.1)$$

, где T - ранговая сумма (T-статистика Вилкоксона), n1, n2 - объемы соответствующих выборок [3].

Для проверки нулевой гипотезы используют меньшее значение критерия из двух: U = min (U1, U2)

Если обе выборки имеют большую длину (больше 10), нулевое распределение U-статистики приближается к нормальному с параметрами

$$\mu_U = \frac{n_1 * n_2}{2} \#(1.2)$$

$$\sigma_U = \left[\frac{n_1 * n_2 * (n_1 + n_2 + 1)}{12}\right]^{\frac{1}{2}} \#(1.3)$$

Первую формулу следует использовать, если все значения в выборках различаются, либо совпадает небольшое их число [3]. В случае большого числа совпадений значение стандартного отклонения завышается и следует использовать

$$\sigma_U = \left[\frac{n_1 * n_2 * (n_1 + n_2 + 1)}{12} - \frac{n_1 * n_2}{12 * (n_1 + n_2) * (n_1 + n_2 - 1)} * \sum_{j=1}^{J} (t_j^3 - t_j)\right]^{\frac{1}{2}} \#(1.4)$$

Нулевая гипотеза может приниматься, если PX> Y = 1/2 даже при равенстве медиан. Тем не менее его допустимо применять для определения различий двух групп, поскольку большинство критериев не позволяют строго проверять те гипотезы, с которыми их связывают.

Также стоит отметить, что критерий не стоит использовать, если есть выраженная последовательная корреляция, поскольку:

- При отсутствии или небольшом сдвиге медиан наличие положительной последовательной корреляции увеличивает вероятность ошибки первого рода. Существование отрицательной последовательной корреляции в свою очередь увеличивает вероятность допустить ошибку второго рода
- Положительная последовательная корреляция в больших выборках также снижает вероятность критерия обнаружить сдвиг медиан, при этом отрицательная корреляция вероятность обнаружения сдвига незначительно увеличивает [3]

Проверим нулевую гипотезу:

1. представим выборки в виде одного вариационного ряда и найдем наблюдаемое значение критерия U
2. найдём по таблице нижнюю критическую точку U_{crit}

Располагаем обе выборки в виде одного пронумерованного вариационного ряда, присваивая одинаковым значениям признака среднее арифметическое значений их рангов:

Таблица 1.2 Ранжирование выборок для расчета статистики Манна-Уитни

Ранг	Значение
1	1.15
2.5	1.4
2.5	1.4
4	1.5
5	1.6
6	1.7
7	1.74
8	1.76
9	1.79
10	1.8
11	1.82
12	1.92
13	2
14	2.2
15	2.48
16.5	2.7
16.5	2.7
18	2.9
19	3
20	3.2

Найдем сумму порядковых номеров вариант выборок:

T1 = 2.5 + 4 + 6 + 9 + 11 + 12 + 13 + 16.5 + 18 + 20 = 112

T2 = 1 + 2.5 + 5 + 7 + 8 + 10 + 14 + 15 + 16.5 + 19 = 98

Затем найдем значение статистики U:

$$U_1 = 10 * 10 + \frac{10 * (10 + 1)}{2} - 112 = 43$$

$$U_2 = 10 * 10 + \frac{10 * (10 + 1)}{2} - 98 = 57$$

$$U = \min(U_1, U_2) = 43$$

При уровне значимости α = 0.05: Ucrit = 17, U> Ucrit, следовательно, нет оснований отклонить нулевую гипотезу Med X – Med Y ≠ 0

1.3.2 Критерий Вилкоксона для связанных выборок

Критерий Вилкоксона служит для проверки однородности двух связанных выборок непрерывных величин: x1, x2, x3, …, xn и y1, y2, y3, …, yn.

Нулевая гипотеза заключается в том, что функции распределения равны при любом значении аргумента x: F (x) = G(x). Конкурирующие гипотезы: F(x)≠ G(x), F (x) <G(x), F (x)> G(x).

Это один самых часто используемых непараметрических тестов для связанных выборок [9].

Данный критерий используется при выполнении следующих условий]:

1. Выборки связаны
2. Парные данные выбраны случайным образом независимо друг от друга
3. Данные измерены в порядковой или числовой шкале

При размере выборок n> 10 статистика приблизительно равна функции нормального распределения с параметрами

$$\mu_W = \frac{n * (n+1)}{4} \#(1.5)$$

$$\sigma_W = \left[\frac{n * (n+1) * (2 * n+1)}{24}\right]^{\frac{1}{2}} \#(1.6)$$

1. Рассчитаем разности значений с одинаковым порядковым номером в выборках и их модули
2. Проранжируем согласно полученным значениям
3. Найдем отдельно суммы рангов положительных T+ и отрицательных T-значений
4. Найдем по таблице критическое значение при заданном уровне значимости α и размере выборок n: Tα, n
5. Мы можем отклонить нулевую гипотезу, если хотя бы одна из полученных сумм меньше критического значения T [4]

Проверим на примере данных:

Таблица 1.3 Расчет рангов разности

Порядковый номер	Станции		Разность	Ранг разности
	Первая	Вторая		
1	2.9	3	-0.1	5
2	1.5	1.8	-0.3	6
3	4.2	2.2	2	10
4	1.7	1.74	-0.04	4
5	2.7	2.7	0	1.5
6	1.4	1.4	0	1.5
7	2	1.6	0.4	7
8	1.79	1.76	0.03	3
9	1.82	1.15	0.67	9
10	1.92	2.48	-0.56	8

$$T_+ = 10 + 7 + 3 + 9 = 29$$
$$T_- = 5 + 6 + 4 + 8 = 23$$

$T_{0.05,10} = 8 < 23$, следовательно, оснований отклонить нулевую гипотезу нет при уровне значимости $\alpha = 0.05$

1.3.3 Критерий Колмогорова-Смирнова для двух выборок

Главным предназначением теста Колмогорова-Смирнова является определение равенства или неравенства распределений двух выборок. Главная его особенность заключается в том, что тест чувствителен не только к смещению медиан, но и к смещениям других параметров, например дисперсий, иными словами, в отличие от представленных выше критериев критерий проверяет однородность функций распределения в более строгом смысле.

Также этот критерий можно использовать для проверки одной выборки, если мы заранее задаем множество распределений, к которому, как мы предполагаем, принадлежит выборка.

При проверке сложных гипотез критерий перестает быть свободным от распределения, поэтому в данной работе тест будет рассматриваться только для простых гипотез.

Нулевая гипотеза:

$$H_0: F(t) = G(t)$$

для любого t

Альтернативная гипотеза:

$$H_1: F(t) \neq G(t)$$

для некоторого t

Для проверки гипотезы используют статистику J Колмогорова-Смирнова

$$J = \frac{m * n}{d} * \max|F_m(t) - G_n(t)| \#(1.7)$$

Где $F_m(t)$ и $G_n(t)$ - эмпирические функции распределения для проверяемых выборок.

Для расчета критерия вычислим накопленные частоты выборок, рассчитаем их долю от полной суммы для каждой выборки и найдем наибольше значение модуля соответствующих разностей.

Таблица 1.4 Расчет накопленных частот для расчета критерия Колмогорова-Смирнова

Год	Станция 1	Станция 2	Сумма 1	Сумма 2	$F_1^*(x)$	$F_2^*(x)$	$\lvert F_1^*(x) - F_2^*(x)\rvert$
2006	2.9	3.0	2.9	3	0.13	0.15	0.019
2007	1.5	1.8	4.4	4.8	0.2	0.24	0.041
2008	4.2	2.2	8.6	7	0.39	0.35	0.039
2009	1.7	1.74	10.3	8.74	0.47	0.44	0.029
2010	2.7	2.7	13	11.44	0.59	0.58	0.016
2011	1.4	1.4	14.4	12.84	0.66	0.65	0.009
2012	2.0	1.6	16.4	14.44	0.75	0.73	0.02
2013	1.79	1.76	18.19	16.2	0.83	0.82	0.013
2014	1.82	1.15	20.01	17.35	0.91	0.87	0.038
2015	1.92	2.48	21.93	19.83	1	1	0

Найдем значение статистики:

$$\lambda = \max|F_1^*(x) - F_2^*| * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}} \ \#(1.8)$$

$$\lambda = 0.041 * \sqrt{\frac{10 * 10}{10 + 10}} = 0.092$$

При уровне значимости 0.05 $\lambda_{0.05}$ = 1.36, значит можно считать, что измерения на первой и второй станциях описываются одной функцией распределения.

1.3.4 Коэффициент ранговой корреляции Кендалла

Критерий τ – Кендалла предназначен для оценки связи между двумя признаками и является непараметрической альтернативой коэффициенту корреляции Пирсона. Коэффициент возвращает значения от -1 до 1, где 1 сильная корреляция, -1 - сильная отрицательная корреляция, 0 – отсутствие корреляции [10]

Коэффициент τ

$$\tau = \frac{P - Q}{\frac{1}{2} * N(N-1)} \#(1.9)$$

, где P– совпадения, Q – инверсии, N – объем выборки

Рассмотрим пример

Составим таблицу со значениями и соответствующими рангами

Таблица 1.5 Расчет числа совпадений и инверсий для коэффициента корреляции Кендалла

Год	Станция 1	Станция 2	Ранг 1	Ранг 2	P	Q
2006	2.9	3.0	9	10	0	1
2007	1.5	1.8	2	6	4	4
2008	4.2	2.2	10	7	0	0
2009	1.7	1.74	3	4	5	2
2010	2.7	2.7	8	9	1	1
2011	1.4	1.4	1	2	8	1
2012	2.0	1.6	7	3	3	0
2013	1.79	1.76	4	5	4	2
2014	1.82	1.15	5	1	5	0
2015	1.92	2.48	6	8	2	2

Итого P = 32, Q = 13

Рассчитаем значение коэффициента

$$\tau = \frac{32 - 13}{\frac{1}{2} * 10 * (10-1)} = 0.422$$

Значение коэффициента корреляции τ>0.3 говорит о значимой связи между выборками

1.3.5 Коэффициент ранговой корреляции Спирмена

Коэффициент ранговой корреляции Спирмена – это непараметрический коэффициент линейной зависимости между признаками. Поскольку коэффициент оперирует рангами, он инвариантен к преобразованиям шкалы измерения [11].

Расчет коэффициента проводится по формуле

$$p = 1 - \frac{6 \sum d^2}{n^3 - n} \ \#(1.10)$$

Где d^2 – квадрат разностей рангов, n – количество признаков, используемых при расчете.

Пример:

Составим таблицу рангов

Таблица 1.5 расчет рангов и квадратов разности рангов для нахождения коэффициента корреляции Спирмена

Год	Станция 1	Станция 2	Ранг 1	Ранг 2	$(d_x-d_y)^2$
2006	2.9	3.0	9	10	1
2007	1.5	1.8	2	6	16
2008	4.2	2.2	10	7	9
2009	1.7	1.74	3	4	1
2010	2.7	2.7	8	9	1
2011	1.4	1.4	1	2	1
2012	2.0	1.6	7	3	16
2013	1.79	1.76	4	5	1
2014	1.82	1.15	5	1	16
2015	1.92	2.48	6	8	4
Сумма			55	55	66

Рассчитаем коэффициент корреляции

$$p = 1 - \frac{6 * 66}{10^3 - 10} = 0.6$$

Значение коэффициента равное 0.6 означает сильную связь между признаками.

1.3.6 Параметрические критерии

Непараметрические критерии для однородностей и связей будут сравниваться с параметрическими критериями Стьюдента и Пирсона.

Критерий Стьюдента является самым популярным критерием в статистике и используется для проверки смещения мат. ожиданий двух выборок. Использоваться будет критерий Стьюдента для связанных выборок, поскольку результаты измерений экологических показателей могут коррелировать друг с другом. Данный критерий подразумевает, что функции распределения обеих выборок нормальные.

Коэффициент корреляции Пирсона применяют для обнаружения связей между двумя выборками. Поскольку это параметрический критерий, то выборки должны быть нормально распределены. Дополнительно требуется, чтобы признаки были измерены в шкале отношений, либо в интервальной, а также выборки должны быть одинакового размера. Также критерий позволяет обнаружить только линейную корреляцию и неустойчив к выбросам.

1.4 Выбор данных для статистического анализа

Основным признаком для статистического анализа однородностей и связей было выбрано биохимическое потребление кислорода, связь которой будет проверятся с температурой воды, pH, содержанием фосфатов, нитратов, кислорода, кальция и других показателей.

Главной причиной такого выбора послужило то, что БПК – это скорее биологический показатель, чем химический, который зависит от многих других показателей, например температуры воды, pH, содержания металлов и кислорода. Этот показатель отображает уменьшение содержания кислорода, в герметично закрытом сосуде за определенное количество

времени, измеряемое в днях. Диапазон времени обычно от 1 до 40 дней. В нашем случае это 5 дней. Обычно в России и некоторых европейских странах этот показатель измеряют в миллиграммах на литр. С помощью него можно определить степень окисления органических веществ. Оптимальные условия – это нейтральный pH, баланс количества фосфора и азота. Особенно остро микроорганизмы реагируют на изменение температуры. В водах, где присутствуют токсичные и дезинфицирующие вещества, БПК отображает степень загрязненности вод этими веществами, поскольку микробиологическая культура в них практически отсутствует. В случае реки Охта, которая с 2011 года классифицируется как грязная из-за обилия донных отложений, содержащих токсичные вещества.

Содержание кислорода в воде также является одним из ключевых показателей при биологической и химической оценке водных ресурсов. В реке концентрация кислорода обычно колеблется в довольно широких пределах от 0 до 14 миллиграмм на литр. Значение признака очень сильно зависит от сезона и времени суток, поскольку растворимость веществ в воде зависит от ее температуры. При высокой минерализации (большом количестве растворимых твердых веществ) растворимость кислорода в воде также падает. Распределение концентрации кислорода по вертикали может быть довольно неравномерно при отсутствии течений.

Температура воды – это показатель, сильно влияющий на остальные химико-биологические показатели. Имеет очень высокую зависимость от времени суток и сезона. Также на температуру воды может оказывать влияние скорость течения.

Главными источниками соединений железа в поверхностных водах являются процессы химического выветривания горных пород, сопровождающиеся их механическим разрушением и растворением. В процессе взаимодействия с содержащимися в природных водах минеральными и органическими веществами образуется сложный комплекс соединений железа, находящихся в воде в растворенном, коллоидном и

взвешенном состоянии. Значительные количества железа поступают с подземным стоком и со сточными водами предприятий металлургической, металлообрабатывающей, текстильной, лакокрасочной промышленности и с сельскохозяйственными стоками [7] ПДК железа в речных водах – 0,1 мг/дм3

Нитриты представляют собой промежуточную ступень в цепи бактериальных процессов окисления аммония до нитратов (нитрификация - только в аэробных условиях) и, напротив, восстановления нитратов до азота и аммиака (денитрификации - при недостатке кислорода). Подобные окислительно-восстановительные реакции характерны для станций аэрации, систем водоснабжения и собственно природных вод. Кроме того, нитриты используются в качестве ингибиторов коррозии в процессах водоподготовки технологической воды и поэтому могут попасть и в системы хозяйственно-питьевого водоснабжения. Широко известно также применение нитритов для консервирования пищевых продуктов. В поверхностных водах нитриты находятся в растворенном виде. В кислых водах могут присутствовать небольшие концентрации азотистой кислоты (HNO2) (не диссоциированной на ионы). Повышенное содержание нитритов указывает на усиление процессов разложения органических веществ в условиях более медленного окисления NO2- в NO3-, что указывает на загрязнение водного объекта, т.е. является важным санитарным показателем. Концентрация нитритов в поверхностных водах составляет сотые (иногда даже тысячные) доли миллиграмма в 1 дм3 [8].

Соединения минерального фосфора поступают в природные воды в результате выветривания и растворения пород, содержащих ортофосфаты (апатиты и фосфориты) и поступления с поверхности водосбора в виде орто-, мета-, пиро- и полифосфат-ионов (удобрения, синтетические моющие средства, добавки, предупреждающие образование накипи в котлах, и т. п.), а также образуются при биологической переработке остатков животных и растительных организмов. Избыточное содержание фосфатов в воде, особенно в грунтовой, может быть отражением присутствия в водном

объекте примесей удобрений, компонентов хозяйственно-бытовых сточных вод, разлагающейся биомассы.

Концентрация фосфатов в природных водах обычно очень мала - сотые, редко десятые доли миллиграммов фосфора в 1 дм3, в загрязненных водах она может достигать нескольких миллиграммов в 1 дм [7].

Выводы по главе

Данные экологических измерений реки Охта, измеренные студентами РГГМУ, являются репрезентативными, поскольку измерения физических данных проводились через одинаковые отрезки времени в один сезон в одних и тех же местах. Непараметрические методы анализа однородностей и связей имеют преимущество перед параметрическими для представленных данных, поскольку размеры выборок в наборе измерений относительно небольшие, а также функции распределения физических данных не обязательно нормально распределены.

Глава 2. Характеристика инструментальных средств статистического анализа программного продукта Anaconda

2.1 Особенности пакета прикладных программ Anaconda

Пакет прикладных программ Anaconda представляет из себя среду с несколькими интегрированными средами разработки и языком программирования общего назначения python с предустановленным окружением с модулями для статистического анализа данных и их представления в графическом виде, самими часто используемыми из которых являются модули NumPy, matplotlib, pandas. Главным преимуществом данного пакета прикладных программ является то, что он распространяется на бесплатной основе, позволяет изолировать наборы модулей для различных проектов и открытый исходный код.

Используемый в данной работе модуль NumPy имеет синтаксис, максимальной схожий с языком, используемым в ППП MATLAB, что существенно упрощает переход с него. В отличие от MATLAB, модули языка python распространяются с открытым исходным кодом, при этом давая практически те же возможности для использования. А поскольку исходный код используемых пакетов доступен каждому, любой может ознакомиться с реализациями алгоритмов анализа.

В качестве основной среды разработки будет использоваться Jupyter Notebook. Ключевая особенностью этой среды является то, что он позволяет редактировать код в браузере, при этом само выполнение программы может происходить на удаленном сервере. Также среда исполняет код на лету и позволяет изменять уже ранее блоки кода. У среды разработки имеется поддержка latex выражений для оформления написанного кода для документирования программ. На сегодняшний день это одно из самых популярных решений для анализа данных и машинного обучения, поскольку его поддерживают крупнейшие облачные сервисы Microsoft azure и Amazon web services [16].

Основные модули для статистического анализа, используемые в этой работе – NumPy, matplotlib, SciPy и pandas.

NumPy – это открытый проект, разрабатывающийся с целью внедрения средств научных вычислений на языке python. Основой пакета являются многомерные массивы и его производные и методы для обработки этих массивов, например: математическая обработка данных массива, изменения структуры, дискретные преобразования Фурье, методы линейной алгебры, генерирование случайных чисел с заданными параметрами и так далее. Синтаксис пакета намеренно основан на синтаксисе языка MATLAB [13].

NumPy в данной работе необходим для работы обработки данных в таблицах.

Matplotlib используется для визуализации данных, pandas в свою очередь позволяет работать с массивами как с таблицами и поддерживает чтение и сохранение в Excel.

Matplotlib – это библиотека для визуализации данных. Она позволяет строить как 2D, так и 3D графики, основана на принципах объектно-ориентированного программирования (ООП). Как и NumPy, разрабатывалась с оглядкой на реализацию в языке программы MATLAB [12].

Имеет поддержку быстрого построения графиков и диаграмм, например:

- Диаграммы рассеяния
- Поля градиентов
- Спектральные диаграммы
- Контурные графики
- Круговые диаграммы

Помимо этого, matplotlib интегрирован в пакет pandas для быстрого построения графиков по табличным данным.

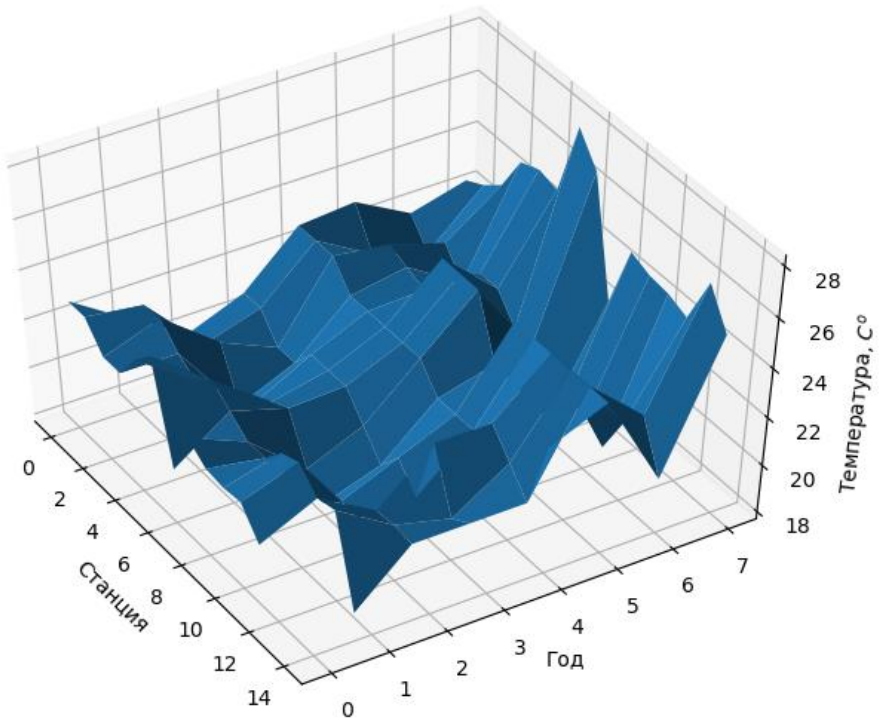

Рисунок 2.1 Пример визуализации данных с помощью matplotlib: 3D-график температур на поверхности, начиная с 2006 года

Pandas – это библиотека для работы с блоками данных в языке python. Широко используется в проектах как академической, так и коммерческой направленности. Главной особенностью является возможность работы с форматами CSV, Excel, SQL и HDF5, что делает ее одной из немногих библиотек, способных одинаково эффективно работать как с текстовыми файлами, так и с таблицами баз данных. Единственным недостатком перед Ms Excel можно назвать необходимость в базовом знании языка python для работы и отсутствие интерактивности таблиц. Из недостатков можно выделить неспособность библиотеки распознавать объединенные поля Excel [14].

```
In [3]: Year2006 = SurfaceVals[0]
        Year2006
```

Out[3]:

	index	Station	Horizon	pH	Temp	Oxygen	BOD5	PO	Chlorides	Alkalinity	Ca	Mg	HardnessOfWater	Fe	Nitrites	Phosphates
0	1	1	surface	6.85	23.0	1.13	3.46	6.40	34.79	1.40	22.50	21.65	2.9	1.36	59.1	118.30
1	3	2	surface	6.92	23.0	2.79	3.12	5.88	28.40	1.00	19.20	24.80	3.0	1.16	40.1	96.46
2	5	3	surface	6.91	22.0	2.87	3.63	6.92	41.89	1.35	22.44	14.35	2.3	1.32	63.1	129.22
3	7	4	surface	6.83	22.0	4.24	4.91	7.27	36.21	1.25	20.44	22.86	2.9	1.12	54.1	100.10
4	9	5	surface	6.97	23.2	5.01	3.92	7.27	41.89	1.25	20.04	21.89	2.8	1.12	49.1	81.90
5	15	8	surface	7.02	24.0	2.97	6.25	7.84	39.05	1.45	26.45	32.71	4.0	1.00	68.1	87.84
6	17	9	surface	6.99	20.0	3.62	4.07	6.93	33.37	1.45	31.26	27.36	3.8	1.08	74.1	80.52
7	19	10	surface	6.77	22.0	2.11	2.77	6.75	34.08	1.35	27.25	27.36	3.6	1.03	66.1	102.48
8	21	11	surface	7.21	21.4	4.66	5.42	7.78	34.08	1.20	18.99	13.98	2.1	1.36	42.1	78.36
9	24	12	surface	7.06	21.3	4.32	4.98	5.36	29.11	1.20	20.04	24.32	2.8	1.28	44.1	67.34
10	25	13	surface	7.21	21.4	4.89	4.79	7.96	34.08	1.15	17.64	20.67	3.3	0.80	47.1	74.62
11	27	14	surface	7.01	20.3	2.65	3.37	7.79	34.79	1.45	20.84	29.43	2.4	1.32	50.1	107.38

Рисунок 2.2 Пример вывода таблицы формата Ms Excel в среде Jupyter Notebook с помощью библиотеки pandas: данные на поверхности реки Охта за 2006 год

SciPy – это библиотека для научных и инженерных вычислений для языка python. Состоит из множества подмодулей:

- Special – модуль, содержащий методы из математической физики. Главное достоинство – возможность параллельного выполнения операций за счет того, что модуль написан на языке C.
- Integrate – модуль для работы с интегралами
- Optimize – модуль, содержащий методы оптимизации, например метод наименьших квадратов и поиск корней уравнения.
- Interpolate – модуль с методами интерполяции: нахождения параметров различных функций по заданным точкам.
- FFT – модуль с методами преобразования Фурье
- Signal – модуль, содержащий в себе функции цифровых фильтров и методы синтеза фильтров
- Linalg – модуль для работы с функциями линейной алгебры
- Spatial – модуль для расчетов триангуляции, построения выпуклых фигур и диаграмм Вороного
- Stats – модуль для статистического анализа данных. Позволяет генерировать наборы случайных данных с заданными

параметрами, содержит подавляющее большинство статистических тестов, использующихся для научных вычислений.

- Ndimage – модуль для цифровой обработки изображений. От других библиотек для работы с изображениями отличается широким набором встроенных методов для анализа изображений и высокой скоростью работы.
- IO – модуль, дающий возможность писать и читать программы написанные с помощью скриптового языка ППП MATLAB в среде выполнения языка python [15].

Выбор ППП anaconda в качестве инструмента для статистического анализа обусловлен также тем, что использование скриптового языка для анализа не только расширяет функциональные возможности, поскольку необходимые операции можно описать самостоятельно, но и позволяет впоследствии восстановить все предыдущие выполненные операции по обработке для обнаружения каких-либо ошибок, что может быть критическим фактором при выборе среды для анализа данных, особенно для больших проектов.

2.2 Выполнение расчетов и формирование результатов статистического анализа

2.2.1 Примеры выполнения расчетов и визуализации анализа однородности выборок

Проверка гипотезы о нулевом смещении медиан на примере измерений БПК-5 за 2 года

Примеры статистического анализа однородности проводятся на примере данных БПК-5 за 2010 и 2011 годы на дне на различных станциях.

Таблица 2.1 Значения БПК5

№ станции	2010г.	2011г.
1	5,53	
2	3,45	6,24
3	3,73	5,53
4	5,31	5,72
5	4,10	5,72
6	3,99	5,64
7	5,14	5,42
8	4,20	5,11
Ок1	2,47	5,59
9	4,89	5,48
10	3,87	5,36
11	4,79	5,91
12	4,67	3,73
13	4,90	5,77
14	3,61	5,82

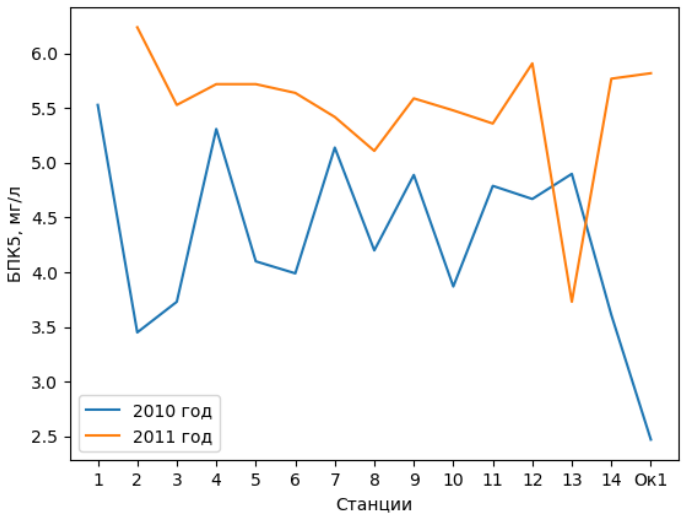

Рисунок 2.3 График значений БПК5 на станциях

Из графика видно, что состояние реки Охта в 2011 году ухудшилось в сравнении с 2010-м годом, как и указано ранее. Следовательно, мы ожидаем, что нулевая гипотеза будет отклоняться при уровне значимости 5%, то есть p-значение должно быть меньше 0,05.

Значения на первой станции не учитывались, поскольку в 2011 году значение БПК-5 не измерено.

```
In [108]: print(
              "Непараметрические тесты",
              "Тест Вилкоксона",
              stats.wilcoxon(Y2010[1:], Y2011[1:]),
              "Тест Вилкоксона-Манна-Уитни",
              stats.mannwhitneyu(Y2010[1:], Y2011[1:]),
              "Тест Колмогорова-Смирнова",
              stats.ks_2samp(Y2010[1:], Y2011[1:]),
              sep="\n",
              end="\n\n",
          )
          print(
              "Параметрический тест Стьюдента",
              stats.ttest_ind(Y2010[1:], Y2011[1:]),
              sep="\n",
          )

Непараметрические тесты
Тест Вилкоксона
WilcoxonResult(statistic=6.0, pvalue=0.003510363671504142)
Тест Вилкоксона-Манна-Уитни
MannwhitneyuResult(statistic=12.5, pvalue=4.679768154233677e-05)
Тест Колмогорова-Смирнова
Ks_2sampResult(statistic=0.8571428571428571, pvalue=1.8845066630771294e-05)

Параметрический тест Стьюдента
Ttest_indResult(statistic=-4.932030261789233, pvalue=4.0206807670726626e-05)
```

Рисунок 2.4 Пример расчета тестов в среде jupyter notebook

Результаты всех тестов позволяют нам отклонить нулевую гипотезу, то есть, как минимум, медианы функций выборок смещены относительно друг друга. Как можно заметить, наименьшее p-значение у тестов Колмогорова-Смирнова, поскольку он проверяет однородность в более строгом смысле (проверяется не только отсутствие смещения медиан) и можно предположить, что есть различия других параметров функций распределения выборок помимо медианы. Меньшее значение p-критерия теста Стьюдента в сравнении с непараметрическими тестами может быть связано с тем, что при расчете критерия предполагается, что дисперсии обеих выборок одинаковы.

Таблица 2.2 Значения содержания кислорода

	№ Станции	
год	3	4
2006	2.74	4.2
2007	0.81	1.06
2008	0.315	0.343
2009	0.0	1.17
2010	3.4	0.74
2011	0.92	0.99
2012	2.02	2.32
2013	1.58	0.42
2014	2.268	2.436

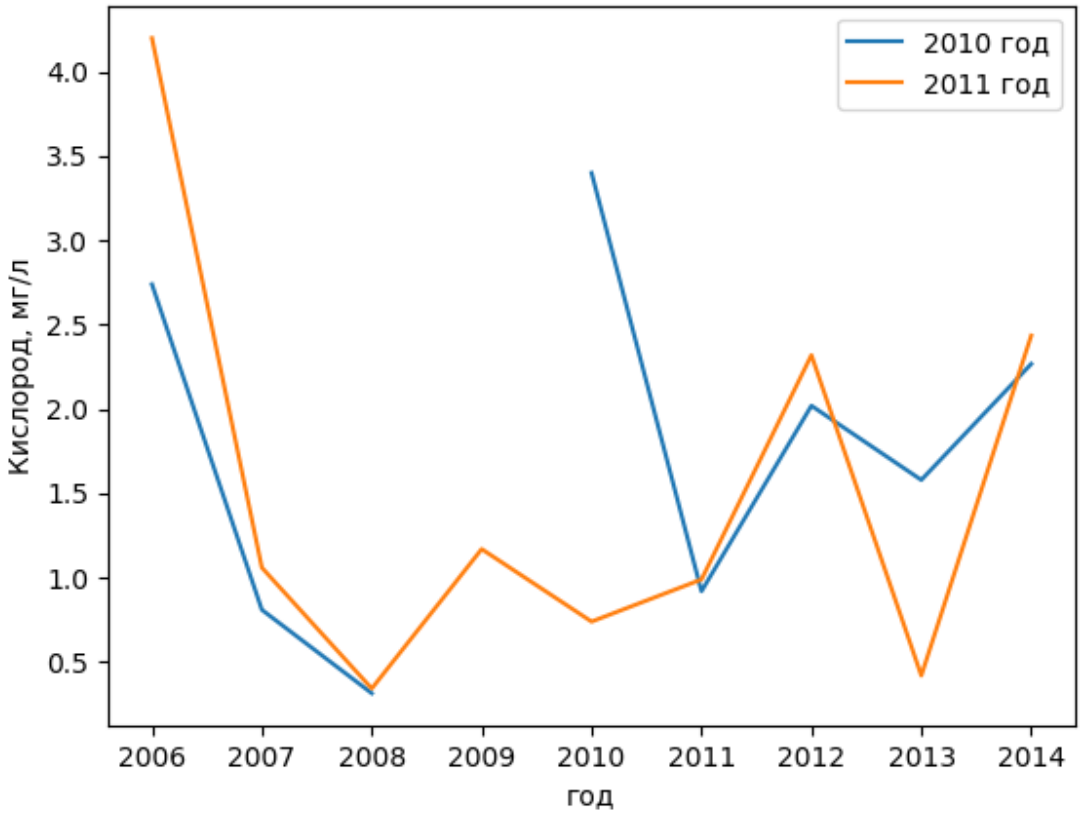

Рисунок 2.5 График измерений содержания кислорода

```
Непараметрические тесты
Тест Вилкоксона
WilcoxonResult(statistic=15.0, pvalue=0.3742593192802244)
Тест Вилкоксона-Манна-Уитни
MannwhitneyuResult(statistic=40.0, pvalue=0.5)
Тест Колмогорова-Смирнова
Ks_2sampResult(statistic=0.2222222222222222, pvalue=0.9894693541752365)

Параметрический тест Стьюдента
Ttest_indResult(statistic=0.07364109819013323, pvalue=0.9422087613706008)
```

Рисунок 2.6 Пример расчетов критериев для выявления однородностей

Как видно из графика, данные измерений отличаются довольно слабо.

Таблица 2.3 Значения pH

	№ Станции	
год	3	4
2006	6.88	6.85
2007	6.89	7.13
2008	7.14	7.11
2009	6.9	6.97
2010	7.14	7.18
2011	7.3	7.21
2012	7.65	7.23
2013	6.53	6.84
2014	7.01	7.05

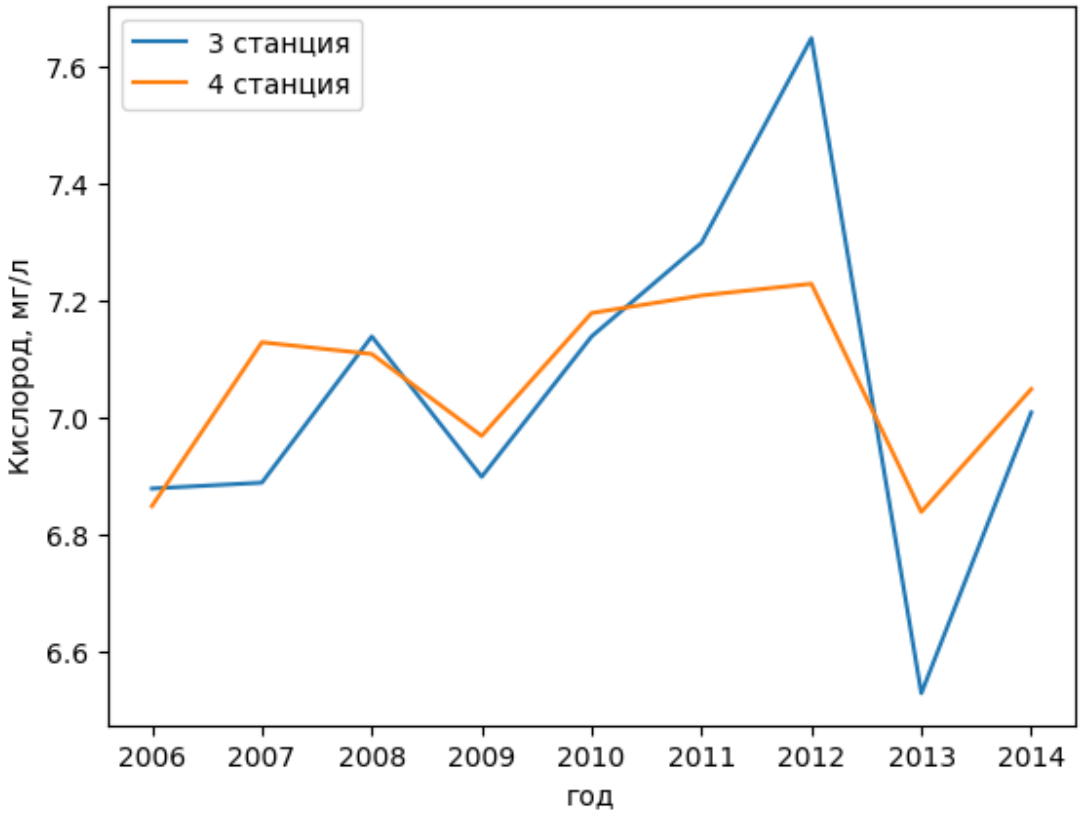

Рисунок 2.7 График изменения содержания pH

Непараметрические тесты
Тест Вилкоксона
```
WilcoxonResult(statistic=18.0, pvalue=0.5936305914425295)
```
Тест Вилкоксона-Манна-Уитни
```
MannwhitneyuResult(statistic=39.0, pvalue=0.4648001417660872)
```
Тест Колмогорова-Смирнова
```
Ks_2sampResult(statistic=0.2222222222222222, pvalue=0.9894693541752365)
```

Параметрический тест Стьюдента
```
Ttest_indResult(statistic=-0.12499566636813811, pvalue=0.9020839995301597)
```

Рисунок 2.8 Результаты тестов на однородность для pH

Таблица 2.4 значения содержания железа

	№ Станции	
год	3	4
2006	1.32	1.04
2007	1.68	3.76
2008	1.64	0.84
2009	4.33	3.24
2010	1.67	3.05
2011	1.14	0.89
2012	1.6	1.34
2013	-	-
2014	2.21	1.95

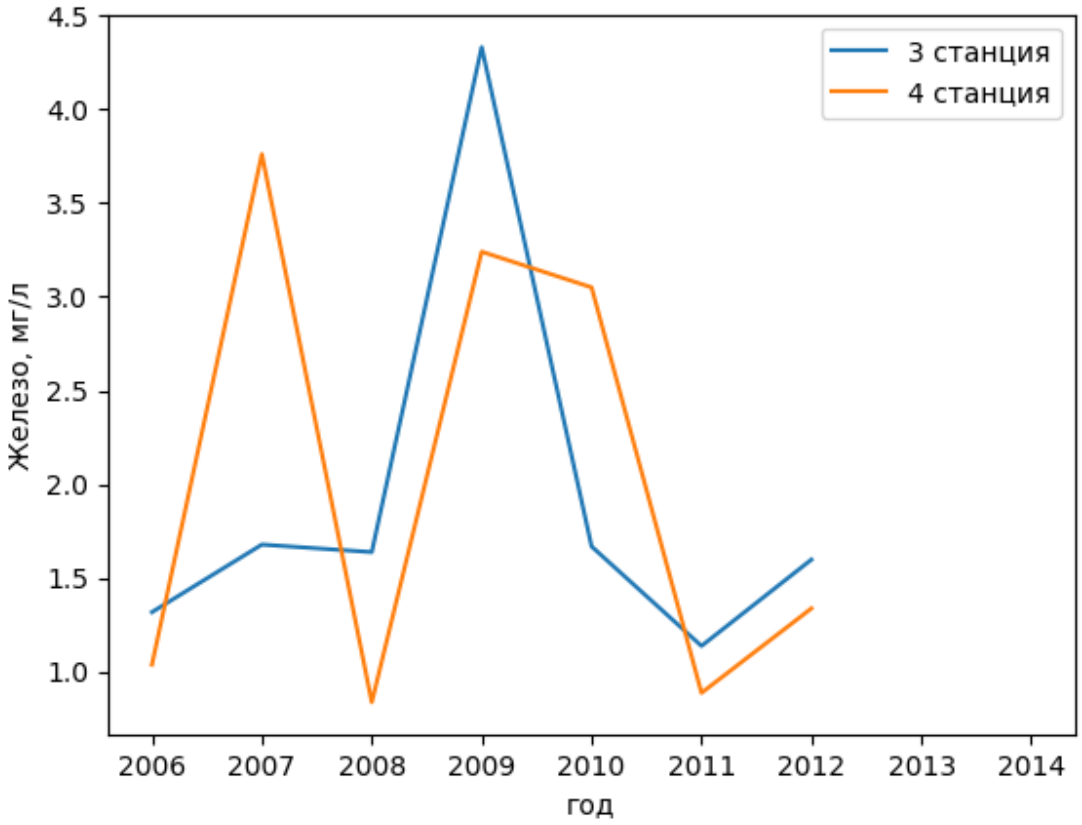

Рисунок 2.7 График изменения содержания железа

Непараметрические тесты
Тест Вилкоксона
WilcoxonResult(statistic=15.0, pvalue=0.6740473527050261)
Тест Вилкоксона-Манна-Уитни
MannwhitneyuResult(statistic=29.0, pvalue=0.3964477515894803)
Тест Колмогорова-Смирнова
Ks_2sampResult(statistic=0.375, pvalue=0.6601398601398599)

Параметрический тест Стьюдента
Ttest_indResult(statistic=-0.1185727764374884, pvalue=0.9072983598833703)

Рисунок 2.8 Результаты тестов на однородность для содержания железа в воде

Таблица 2.5 Значения содержания фосфатов

	№ Станции	
год	3	4
2006	129.22	100.1
2007	234.24	204.96
2008	269.8	177.5
2009	612.74	220.56
2010	81.4	148.7
2011	124.08	97.76
2012	146.4	126.5
2013	97.05	176.04
2014	131.72	118.37

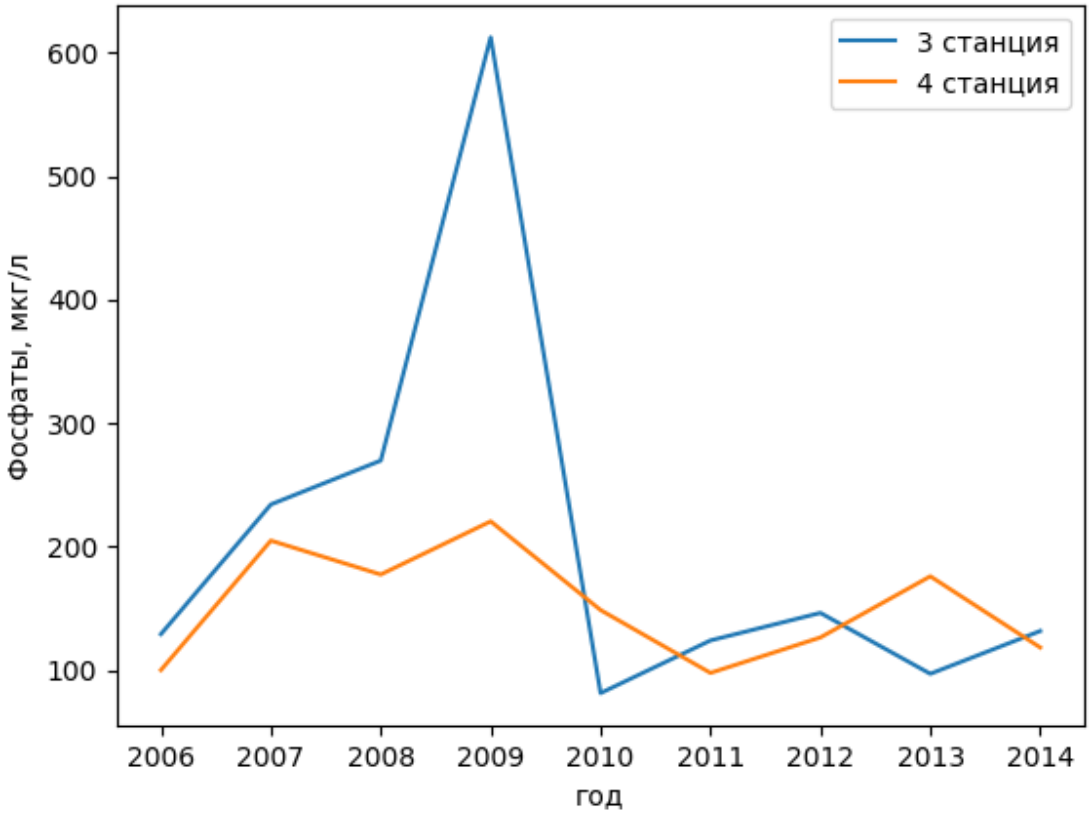

Рисунок 2.8 График изменения содержания фосфатов

```
Непараметрические тесты
Тест Вилкоксона
WilcoxonResult(statistic=13.0, pvalue=0.26039294361048326)
Тест Вилкоксона-Манна-Уитни
MannwhitneyuResult(statistic=39.0, pvalue=0.46481826225353534)
Тест Колмогорова-Смирнова
Ks_2sampResult(statistic=0.3333333333333333, pvalue=0.7301110654051831)

Параметрический тест Стьюдента
Ttest_indResult(statistic=0.8863152195614998, pvalue=0.38857536491299594)
```

Рисунок 2.9 Результаты тестов на однородность для фосфатов

Таблица 2.6 Значения содержания нитритов

	№ Станции	
год	3	4
2006	61.1	51.1
2007	133.1	162.1
2008	105.1	119.1
2009	111.1	87.1
2010	51.1	85.1
2011	174.0	144.0
2012	119.5	123.5
2013	62.7	72.3
2014	70.19	80.99

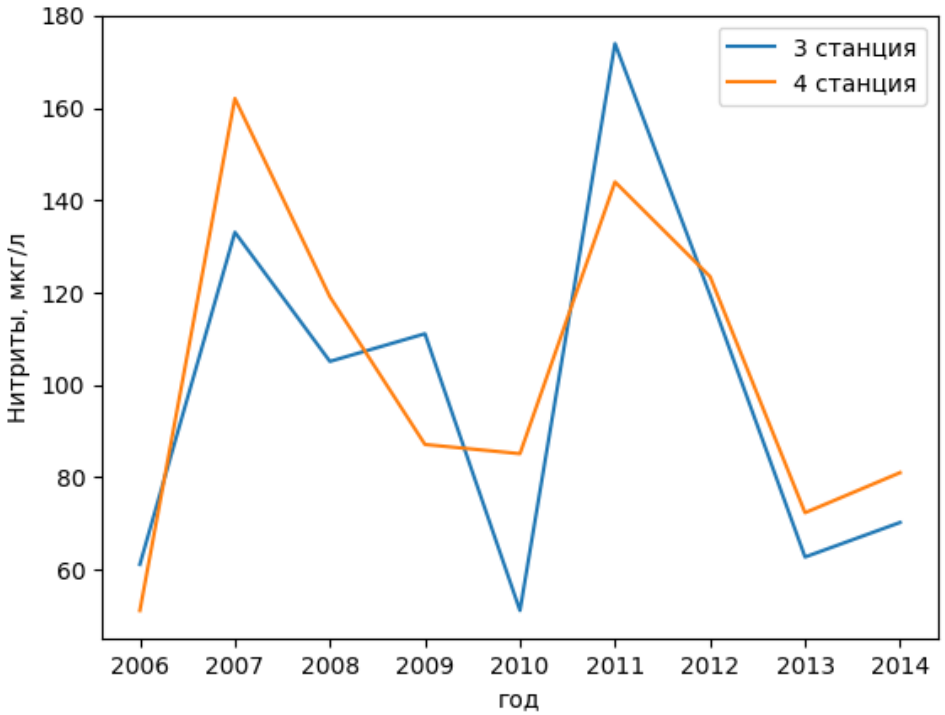

Рисунок 2.10 Изменение содержания нитритов в воде

```
Непараметрические тесты
Тест Вилкоксона
WilcoxonResult(statistic=17.0, pvalue=0.5146697234497355)
Тест Вилкоксона-Манна-Уитни
MannwhitneyuResult(statistic=35.5, pvalue=0.3454754683021908)
Тест Колмогорова-Смирнова
Ks_2sampResult(statistic=0.3333333333333333, pvalue=0.7301110654051831)

Параметрический тест Стьюдента
Ttest_indResult(statistic=-0.22848748628984647, pvalue=0.8221619369837582)
```

Рисунок 2.10 Пример расчета критериев для показаний содержания нитритов в воде

2.2.2 Примеры выполнения расчетов и визуализации анализа связей выборок

Примеры статистического анализа однородности проводятся на примере данных БПК-5 и содержания кислорода в воде 2011 год на дне на различных станциях.

Таблица 2.7 БПК-5 и содержание кислорода на различных станциях за 2011 год

№ Станции	БПК5, мг/л	Содержание кислорода, мл/л
1	5,53	
2	3,45	6,24
3	3,73	5,53
4	5,31	5,72
5	4,10	5,72
6	3,99	5,64
7	5,14	5,42
8	4,20	5,11
Ок1	2,47	5,59
9	4,89	5,48
10	3,87	5,36
11	4,79	5,91
12	4,67	3,73
13	4,90	5,77
14	3,61	5,82

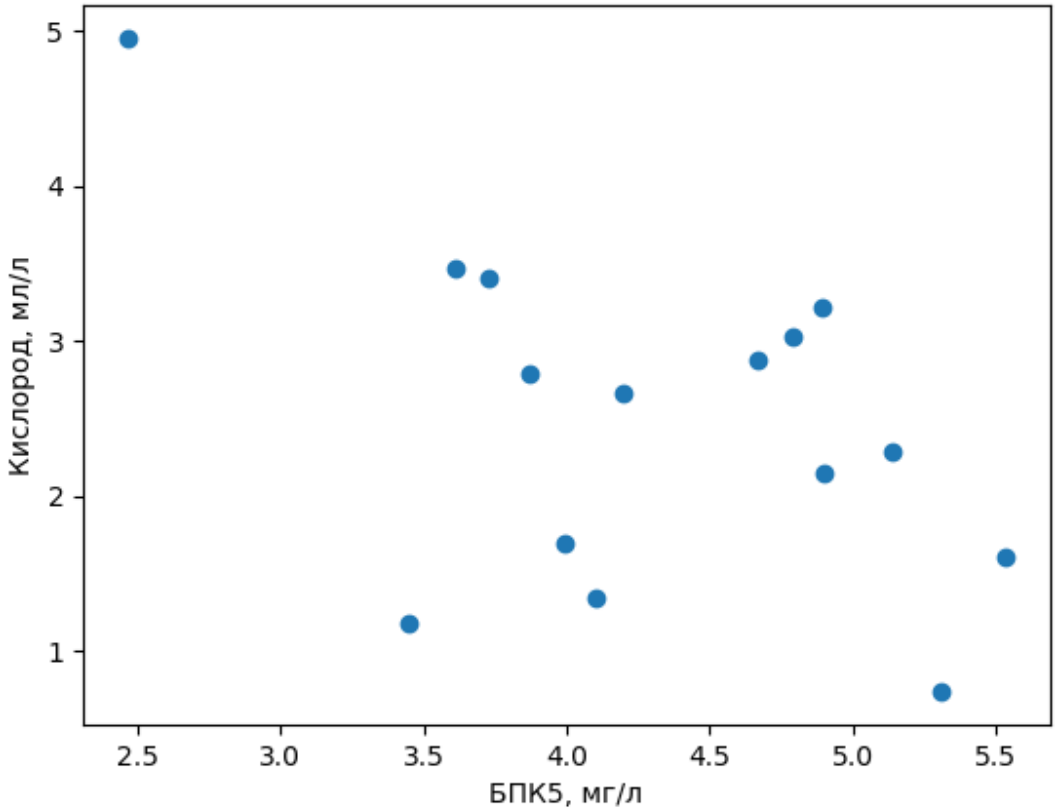

Рисунок 2.11 График значений содержания кислорода и БПК5 в 2011 году

По графику можно заметить небольшую отрицательную линейную зависимость значений.

```
In [144]: print(
              "Коэффициент корреляции Кендалла",
              stats.kendalltau(bod, oxy),
              "Коэффициент корреляции Спирмена",
              stats.spearmanr(bod, oxy),
              "Коэффициент корреляции Пирсона",
              stats.pearsonr(bod, oxy),
              sep="\n",
          )

Коэффициент корреляции Кендалла
KendalltauResult(correlation=-0.3142857142857143, pvalue=0.11424717239544455)
Коэффициент корреляции Спирмена
SpearmanrResult(correlation=-0.42142857142857143, pvalue=0.11769939585541649)
Коэффициент корреляции Пирсона
(-0.5284126823954302, 0.042871234254703815)
```

Рисунок 2.12 Пример расчетов коэффициентов корреляции для содержания кислорода

Все три теста показывают наличие значительной корреляции между значениями БПК и содержания кислорода, что может говорить о том, что эти признаки могут иметь линейную связь. Столь высокое значение для параметрического теста Пирсона можно объяснить его чувствительностью к выбросам.

Также стоит привести примеры для других показателей, от которых может зависеть биохимическое потребление кислорода.

Таблица 2.8 Данные измерений на различных станциях за 2010 год

№ Станции	БПК5, мг/л	Температура, С	pH	Содержание железа, Мг/л	Содержание нитритов, Мкг/л	Содержание фосфатов, Мкг/л
1	5,53	21.5	7.07	1.96	33.1	95.6
2	3,45	22.0	7.13	2.11	79.1	127.4
3	3,73	21.5	7.14	1.67	51.1	81.4
4	5,31	21.5	7.18	3.05	85.1	148.7
5	4,10	20.0	7.21	4.07	111.1	293.5
6	3,99	20.0	7.25	4.21	79.1	293.5
7	5,14	19.5	7.33	5.23	149.1	233.6
8	4,20	20.0	7.35	3.05	117.1	113.8
Ок1	2,47	18.0	7.68	2.9	83.1	107.8
9	4,89	19.5	7.23	3.63	89.1	106.2
10	3,87	19.5	7.32	3.92	85.1	106.2
11	4,79	21.0	7.31	2.54	83.1	102.7
12	4,67	21.0	7.23	2.47	103.1	113.3
13	4,90	19.0	7.34	3.78	87.1	95.6
14	3,61	22.0	7.06	0.54	10.9	19.1

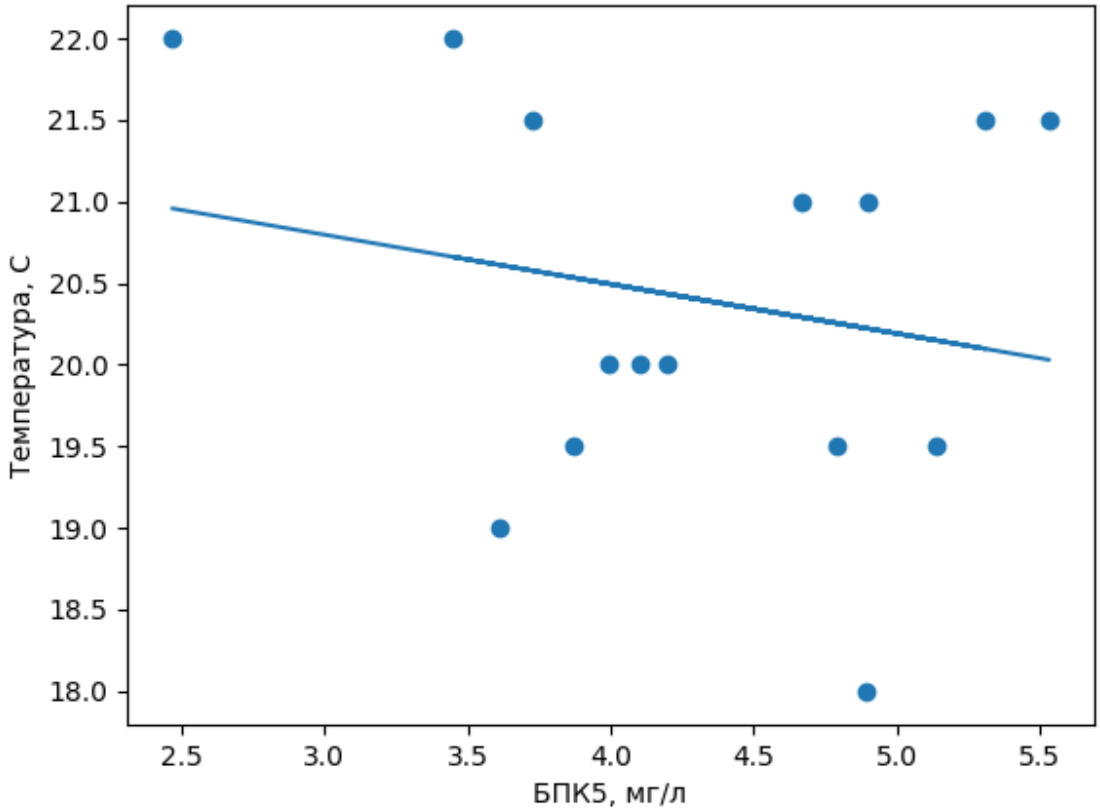

Рисунок 2.13 График зависимости БПК от температуры

Коэффициент корреляции Кендалла
-0.08052512555716988
Коэффициент корреляции Спирмена
-0.16276168399558924
Коэффициент корреляции Пирсона
-0.20849786801362705

Рисунок 2.14 Результаты расчета коэффициентов корреляции для температуры

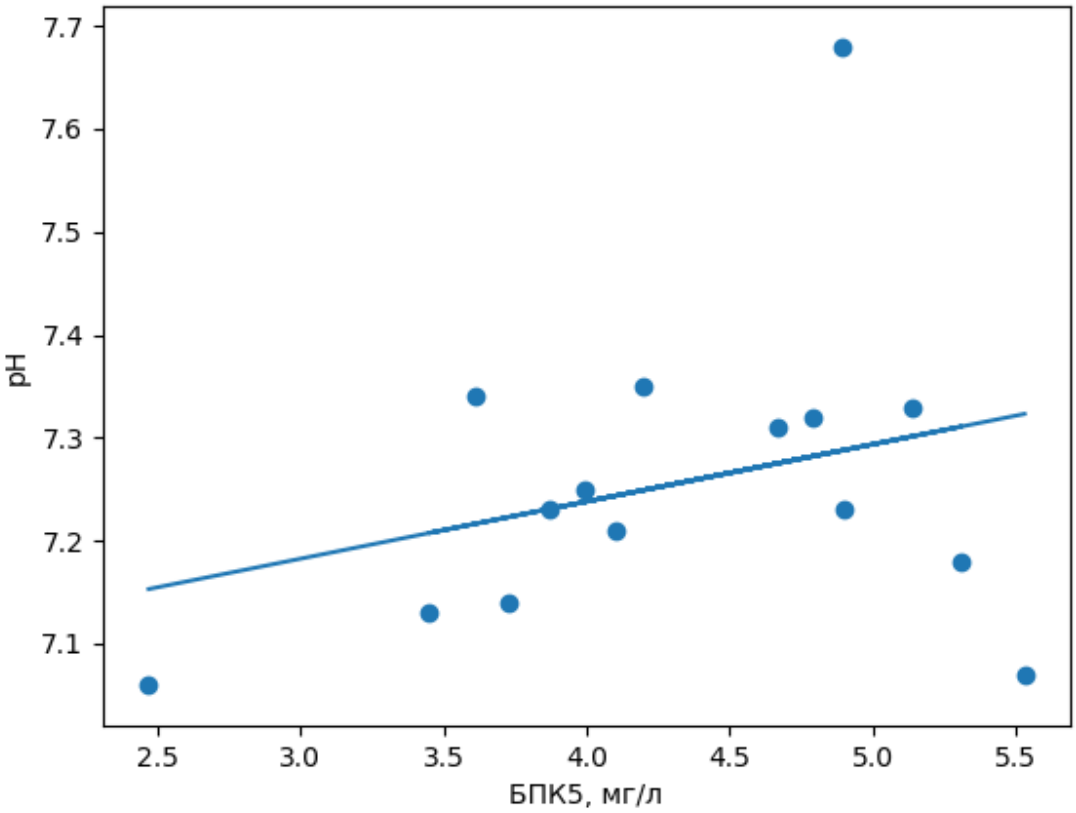

Рисунок 2.15 График зависимости БПК от pH

```
Коэффициент корреляции Кендалла
0.19138975058773822
Коэффициент корреляции Спирмена
0.20911536500314523
Коэффициент корреляции Пирсона
0.3033308936716506
```

Рисунок 2.16 Результаты расчета коэффициентов корреляции для pH

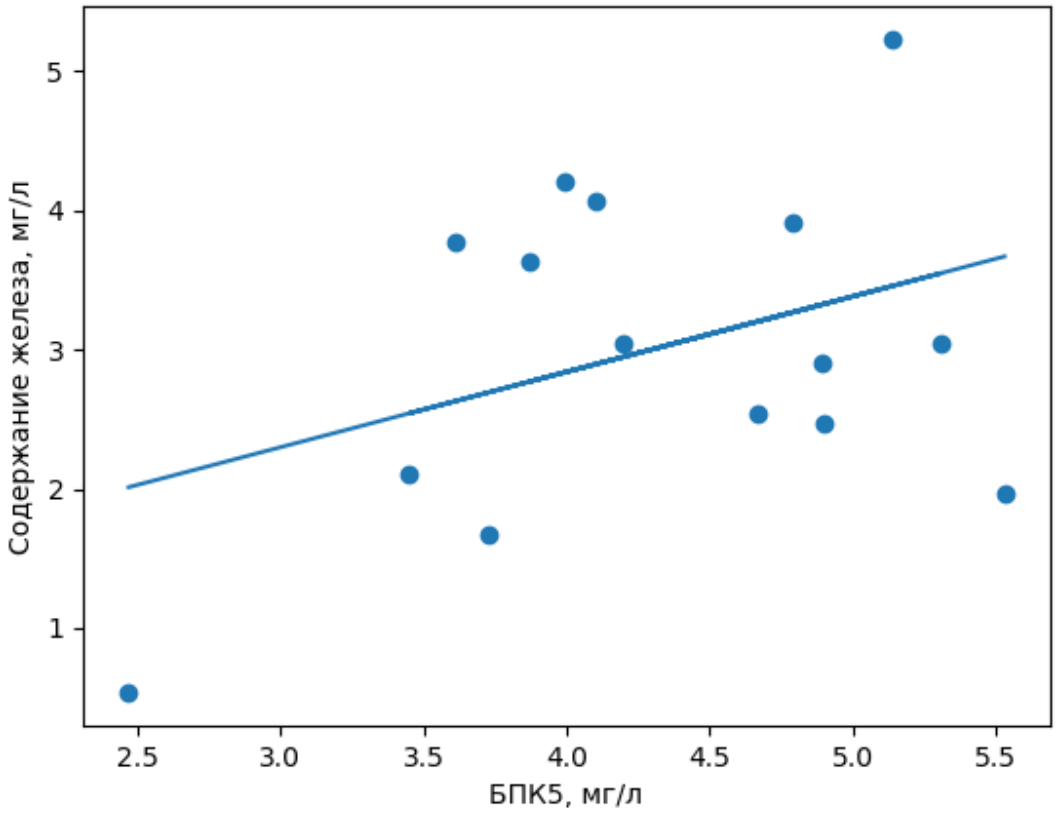

Рисунок 2.17 График зависимости БПК от содержания железа в воде

```
Коэффициент корреляции Кендалла
0.09569487529386911
Коэффициент корреляции Спирмена
0.1894549460712256
Коэффициент корреляции Пирсона
0.3773090489701253
```

Рисунок 2.18 Результаты расчета коэффициентов корреляции для содержания железа

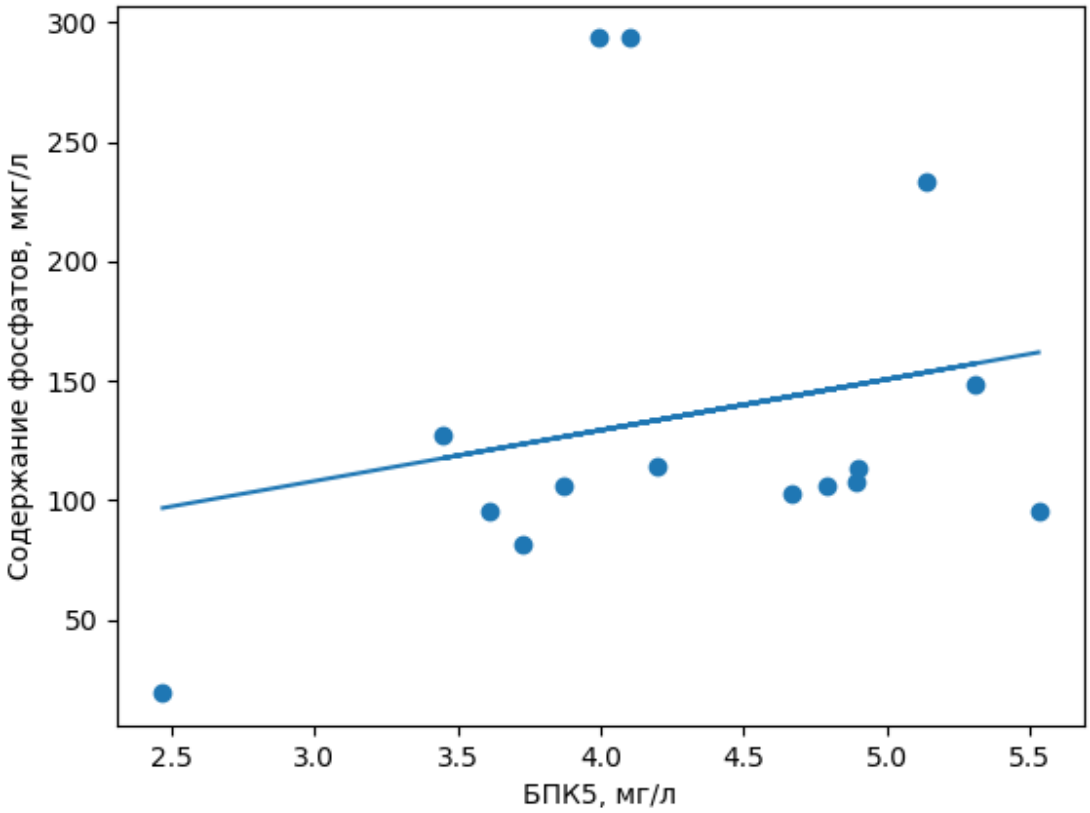

Рисунок 2.19 График зависимости БПК от содержания фосфатов

```
Коэффициент корреляции Кендалла
0.2125827130725644
Коэффициент корреляции Спирмена
0.2793206135062723
Коэффициент корреляции Пирсона
0.22686231413157917
```

Рисунок 2.20 Результаты расчета коэффициентов корреляции для содержания фосфатов

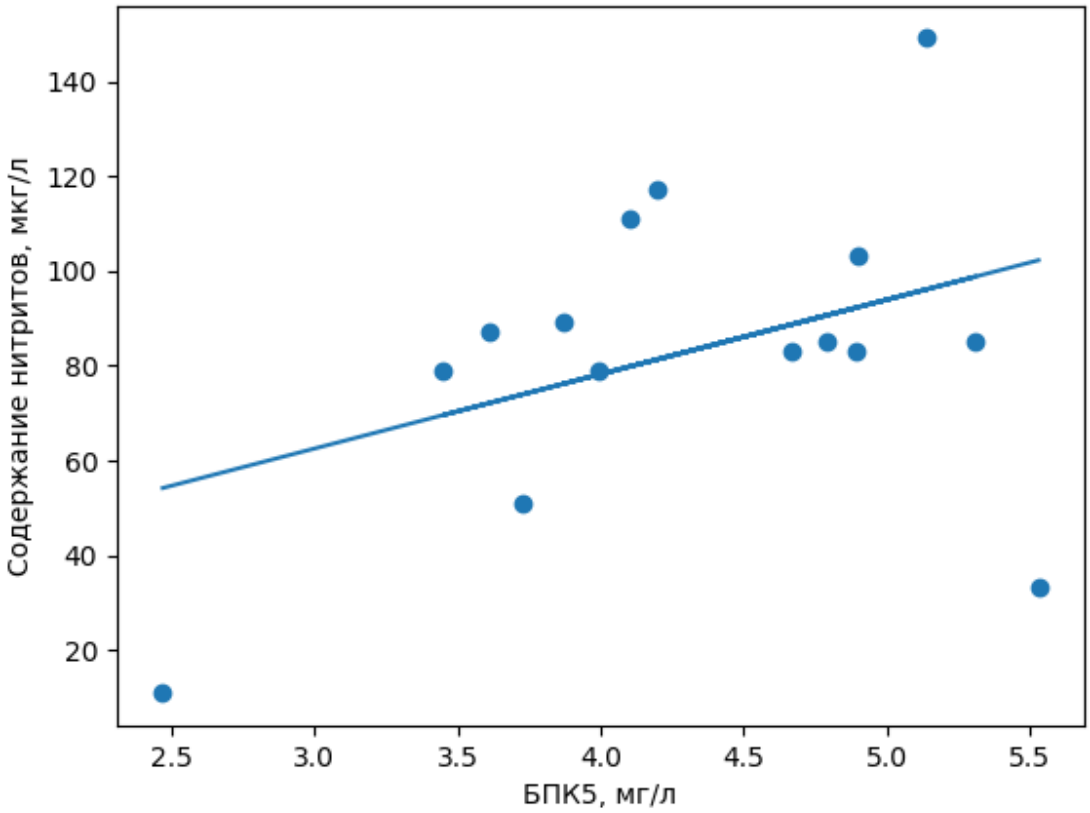

Рисунок 2.21 График зависимости БПК от содержания нитритов

```
Коэффициент корреляции Кендалла
0.25123411544939434
Коэффициент корреляции Спирмена
0.28648268051925363
Коэффициент корреляции Пирсона
0.38904077140649873
```

Рисунок 2.22 Результаты расчета коэффициентов корреляции для содержания нитритов

По визуальной оценке, стохастическая связь между содержанием нитритов и БПК довольно заметна, при этом непараметрические критерии Кендалла и Спирмена показывают незначительную корреляцию между признаками. Параметрический тест Пирсона в свою очередь находит значительную связь.

Выводы по главе

ППП Anaconda позволяет проводить полноценный статистический анализ и визуализацию данных для статистического анализа экологических показателей реки Охта. В ходе данной главы с помощью различных тестов была установлена неоднородность биохимического потребления кислорода в два разных года и корреляционные связи признака с другими экологическими показателями.

Глава 3. Методика непараметрического анализа однородностей и связей многомерных рядов

3.1 Анализ однородностей временных рядов

Анализ однородности нескольких временных рядов имеет большую важность, поскольку позволяет нам обнаружить смещение параметров функций распределения относительно друг друга. Относительно данных экологических измерений реки Охта это может означать, что в различных местах реки оказываются различные влияния на экологическое состояние, например, сброс отходов в реку.

Мы будем брать уровень значимости 0.05 как критическое значение для отклонения нулевой гипотезы.

Сначала рассмотрим результаты применения t-критерия Стьюдента для проверки однородностей на различных станциях на поверхности и на дне для БПК-5.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.157675	0.026583	0.034834	0.017837	0.026310	0.137227	0.091535	0.069141	0.078543	0.029827	0.121467	0.663761	0.443130
2	0.157675	1.000000	0.161917	0.126420	0.239827	0.317480	0.655368	0.778297	0.368601	0.359638	0.402036	0.410365	0.621545	0.081955
3	0.026583	0.161917	1.000000	0.680931	0.931689	0.916405	0.285008	0.135247	0.431709	0.624569	0.536143	0.607135	0.050271	0.063938
4	0.034834	0.126420	0.680931	1.000000	0.640627	0.746554	0.207718	0.118424	0.297876	0.425421	0.366025	0.418097	0.049500	0.077434
5	0.017837	0.239827	0.931689	0.640627	1.000000	0.736516	0.286328	0.156399	0.280473	0.563462	0.705048	0.480829	0.034964	0.013022
6	0.026310	0.317480	0.916405	0.746554	0.736516	1.000000	0.389035	0.135927	0.261092	0.647059	0.957568	0.730550	0.034199	0.031932
7	0.137227	0.655368	0.285008	0.207718	0.286328	0.389035	1.000000	0.789457	0.648861	0.588113	0.657031	0.646804	0.313751	0.082834
8	0.091535	0.778297	0.135247	0.118424	0.156399	0.135927	0.789457	1.000000	0.364296	0.384812	0.437938	0.456173	0.344627	0.148800
9	0.069141	0.368601	0.431709	0.297876	0.280473	0.261092	0.648861	0.364296	1.000000	0.848317	0.947024	0.905229	0.118308	0.118424
10	0.078543	0.359638	0.624569	0.425421	0.563462	0.647059	0.588113	0.384812	0.848317	1.000000	0.911872	0.959831	0.146473	0.152265
11	0.029827	0.402036	0.536143	0.366025	0.705048	0.957568	0.657031	0.437938	0.947024	0.911872	1.000000	0.957823	0.163680	0.042018
12	0.121467	0.410365	0.607135	0.418097	0.480829	0.730550	0.646804	0.456173	0.905229	0.959831	0.957823	1.000000	0.184966	0.103277
13	0.663761	0.621545	0.050271	0.049500	0.034964	0.034199	0.313751	0.344627	0.118308	0.146473	0.163680	0.184966	1.000000	0.490680
14	0.443130	0.081955	0.063938	0.077434	0.013022	0.031932	0.082834	0.148800	0.118424	0.152265	0.042018	0.103277	0.490680	1.000000

Рисунок 3.1 Таблица p-значений критерия Стьюдента для данных на поверхности

Только в 22 случаях из 182 можно отклонить нулевую гипотезу об отсутствии смещения математических ожиданий. Критерий Стьюдента обнаружил различия мат. ожиданий между первой и пятой, шестой станциями и четырнадцатой. Это может говорить о том, что на в этих местах есть факторы, довольно сильно влияющие на БПК.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.936754	0.468638	0.962528	0.910612	0.759319	0.964693	0.997171	0.530817	0.383047	0.898365	0.721663	0.695454	0.665813
2	0.936754	1.000000	0.395788	0.892016	0.937270	0.772876	0.873368	0.931364	0.431725	0.312117	0.883235	0.671203	0.738806	0.777948
3	0.468638	0.395788	1.000000	0.460273	0.388092	0.233840	0.327357	0.453477	0.771684	0.890155	0.756855	0.723228	0.270570	0.936479
4	0.962528	0.892016	0.460273	1.000000	0.844536	0.665087	0.986902	0.964071	0.515737	0.366482	0.839002	0.747198	0.642789	0.915031
5	0.910612	0.937270	0.388092	0.844536	1.000000	0.844256	0.989835	0.971120	0.513190	0.333870	0.773474	0.617829	0.659713	0.973031
6	0.759319	0.772876	0.233840	0.665087	0.844256	1.000000	0.779183	0.886923	0.266455	0.184751	0.450193	0.445078	0.745669	0.734126
7	0.964693	0.873368	0.327357	0.986902	0.989835	0.779183	1.000000	0.966300	0.247822	0.219182	0.521832	0.505789	0.580633	0.673872
8	0.997171	0.931364	0.453477	0.964071	0.971120	0.886923	0.966300	1.000000	0.509361	0.365293	0.823260	0.610819	0.683667	0.664842
9	0.530817	0.431725	0.771684	0.515737	0.513190	0.266455	0.247822	0.509361	1.000000	0.624693	0.973255	0.988661	0.281195	0.432669
10	0.383047	0.312117	0.890155	0.366482	0.333870	0.184751	0.219182	0.365293	0.624693	1.000000	0.359420	0.656804	0.211378	0.429050
11	0.898365	0.883235	0.756855	0.839002	0.773474	0.450193	0.521832	0.823260	0.973255	0.359420	1.000000	0.555387	0.629246	0.171408
12	0.721663	0.671203	0.723228	0.747198	0.617829	0.445078	0.505789	0.610819	0.988661	0.656804	0.555387	1.000000	0.385955	0.575987
13	0.695454	0.738806	0.270570	0.642789	0.659713	0.745669	0.580633	0.683667	0.281195	0.211378	0.629246	0.385955	1.000000	0.427042
14	0.665813	0.777948	0.936479	0.915031	0.973031	0.734126	0.673872	0.664842	0.432669	0.429050	0.171408	0.575987	0.427042	1.000000

Рисунок 3.2 Таблица p-значений критерия Стьюдента для данных на дне

Для данных измерений на дне тест Стьюдента не нашел ни одного смещения мат. ожиданий для измерений БПК.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.020879	0.015156	0.057804	0.010862	0.049950	0.066316	0.085831	0.173071	0.109745	0.020879	0.109745	0.260393	0.463071
2	0.020879	1.000000	0.059336	0.114128	0.213524	0.483840	0.386271	0.878482	0.444587	0.241121	0.059336	0.284503	0.284503	0.172955
3	0.015156	0.059336	1.000000	0.721277	0.858955	0.865772	0.241121	0.059336	0.444587	0.386271	0.241121	0.721277	0.012515	0.115851
4	0.057804	0.114128	0.721277	1.000000	0.593955	0.400814	0.092601	0.168807	0.332880	0.313938	0.414520	0.444587	0.020879	0.115851
5	0.010862	0.213524	0.858955	0.593955	1.000000	0.483840	0.213524	0.173071	0.514670	0.952765	0.952765	0.858955	0.050612	0.046399
6	0.049950	0.483840	0.865772	0.400814	0.483840	1.000000	0.262618	0.068364	0.262618	0.674424	1.000000	0.888638	0.092892	0.115851
7	0.066316	0.386271	0.241121	0.092601	0.213524	0.262618	1.000000	0.721277	0.507624	0.444587	0.507624	0.374259	0.284503	0.115851
8	0.085831	0.878482	0.059336	0.168807	0.173071	0.068364	0.721277	1.000000	0.284503	0.386271	0.386271	0.332880	0.386271	0.248864
9	0.173071	0.444587	0.444587	0.332880	0.514670	0.262618	0.507624	0.284503	1.000000	0.878482	0.858955	0.959354	0.092601	0.172955
10	0.109745	0.241121	0.386271	0.313938	0.952765	0.674424	0.444587	0.386271	0.878482	1.000000	0.798859	0.878482	0.123485	0.172955
11	0.020879	0.059336	0.241121	0.414520	0.952765	1.000000	0.507624	0.386271	0.858955	0.798859	1.000000	0.878482	0.114128	0.115851
12	0.109745	0.284503	0.721277	0.444587	0.858955	0.888638	0.374259	0.332880	0.959354	0.878482	0.878482	1.000000	0.092601	0.115851
13	0.260393	0.284503	0.012515	0.020879	0.050612	0.092892	0.284503	0.386271	0.092601	0.123485	0.114128	0.092601	1.000000	0.753152
14	0.463071	0.172955	0.115851	0.115851	0.046399	0.115851	0.115851	0.248864	0.172955	0.172955	0.115851	0.115851	0.753152	1.000000

Рисунок 3.3 Таблица p-значений критерий Вилкоксона для данных на поверхности

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.374259	0.313938	0.593955	1.000000	0.888638	0.952765	0.858955	0.514670	0.109745	0.735317	0.674424	0.173071	1.000000
2	0.374259	1.000000	0.138641	0.138641	1.000000	0.888638	0.767097	0.952765	0.313938	0.020879	0.916512	0.400814	0.374259	1.000000
3	0.313938	0.138641	1.000000	0.207578	0.483840	0.161429	0.260393	0.313938	0.678402	0.514670	0.612090	0.779435	0.213524	1.000000
4	0.593955	0.138641	0.207578	1.000000	0.888638	0.779435	0.858955	0.858955	0.313938	0.050612	0.498962	0.575403	0.342829	1.000000
5	1.000000	1.000000	0.483840	0.888638	1.000000	0.779435	0.888638	0.674424	0.674424	0.400814	0.735317	0.483840	0.779435	1.000000
6	0.888638	0.888638	0.161429	0.779435	0.779435	1.000000	1.000000	0.779435	0.207578	0.262618	0.498962	0.400814	0.483840	1.000000
7	0.952765	0.767097	0.260393	0.858955	0.888638	1.000000	1.000000	0.593955	0.173071	0.109433	0.612090	0.483840	0.514670	0.285049
8	0.858955	0.952765	0.313938	0.858955	0.674424	0.779435	0.593955	1.000000	0.593955	0.313938	0.865772	0.483840	0.313938	0.592980
9	0.514670	0.313938	0.678402	0.313938	0.674424	0.207578	0.173071	0.593955	1.000000	0.514670	0.735317	0.575403	0.374259	0.108809
10	0.109745	0.020879	0.514670	0.050612	0.400814	0.262618	0.109433	0.313938	0.514670	1.000000	0.612090	0.398025	0.050612	0.285049
11	0.735317	0.916512	0.612090	0.498962	0.735317	0.498962	0.612090	0.865772	0.735317	0.612090	1.000000	0.612090	0.612090	0.285049
12	0.674424	0.400814	0.779435	0.575403	0.483840	0.400814	0.483840	0.483840	0.575403	0.398025	0.612090	1.000000	0.161429	1.000000
13	0.173071	0.374259	0.213524	0.342829	0.779435	0.483840	0.514670	0.313938	0.374259	0.050612	0.612090	0.161429	1.000000	0.592980
14	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.285049	0.592980	0.108809	0.285049	0.285049	1.000000	0.592980	1.000000

Рисунок 3.4 Таблица p-значений критерий Вилкоксона для данных на дне

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.108187	0.013638	0.021077	0.021130	0.032941	0.092663	0.066660	0.038694	0.038694	0.013638	0.055981	0.395541	0.235585
2	0.108187	1.000000	0.092938	0.120572	0.144657	0.215279	0.311588	0.454861	0.236338	0.213678	0.213678	0.213678	0.311588	0.086743
3	0.013638	0.092938	1.000000	0.395630	0.429910	0.458118	0.120661	0.120661	0.224761	0.425053	0.236338	0.454861	0.022577	0.064103
4	0.021077	0.120572	0.395630	1.000000	0.329339	0.247100	0.136428	0.106061	0.311523	0.381098	0.236255	0.454844	0.018710	0.188821
5	0.021130	0.144657	0.429910	0.329339	1.000000	0.437367	0.144657	0.092663	0.213388	0.298121	0.395541	0.361966	0.021130	0.022664
6	0.032941	0.215279	0.458118	0.247100	0.437367	1.000000	0.281618	0.077975	0.215279	0.356495	0.437367	0.479046	0.015600	0.046063
7	0.092663	0.311588	0.120661	0.136428	0.144657	0.281618	1.000000	0.484925	0.285375	0.285375	0.298282	0.272599	0.136518	0.086743
8	0.066660	0.454861	0.120661	0.106061	0.092663	0.077975	0.484925	1.000000	0.172352	0.213678	0.236338	0.106147	0.120661	0.148977
9	0.038694	0.236338	0.224761	0.311523	0.213388	0.215279	0.285375	0.172352	1.000000	0.484925	0.469860	0.285375	0.052055	0.114883
10	0.038694	0.213678	0.425053	0.381098	0.298121	0.356495	0.285375	0.213678	0.484925	1.000000	0.454861	0.410265	0.060472	0.114883
11	0.013638	0.213678	0.236338	0.236255	0.395541	0.437367	0.298282	0.236338	0.469860	0.454861	1.000000	0.366865	0.092938	0.086743
12	0.055981	0.213678	0.454861	0.454844	0.361966	0.479046	0.272599	0.106147	0.285375	0.410265	0.366865	1.000000	0.120661	0.046348
13	0.395541	0.311588	0.022577	0.018710	0.021130	0.015600	0.136518	0.120661	0.052055	0.060472	0.092938	0.120661	1.000000	0.148977
14	0.235585	0.086743	0.064103	0.188821	0.022664	0.046063	0.086743	0.148977	0.114883	0.114883	0.086743	0.046348	0.148977	1.000000

Рисунок 3.5 Таблица p-значений критерия Вилкоксона-Манна-Уитни для данных на поверхности

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.429910	0.298026	0.329339	0.500000	0.479061	0.213388	0.464818	0.165694	0.188612	0.399149	0.479061	0.345475	0.500000
2	0.429910	1.000000	0.213269	0.313426	0.396448	0.479061	0.239964	0.500000	0.154815	0.188612	0.423920	0.416757	0.429910	0.331260
3	0.298026	0.213269	1.000000	0.165068	0.094469	0.077975	0.060945	0.125377	0.329339	0.464800	0.304443	0.396373	0.239851	0.500000
4	0.329339	0.313426	0.165068	1.000000	0.200060	0.500000	0.329339	0.500000	0.108068	0.188489	0.304640	0.281618	0.298026	0.331260
5	0.500000	0.396448	0.094469	0.200060	1.000000	0.356596	0.247418	0.479061	0.135074	0.186015	0.304640	0.281762	0.318251	0.331260
6	0.479061	0.479061	0.077975	0.500000	0.356596	1.000000	0.479061	0.479061	0.159213	0.135074	0.221644	0.247418	0.437413	0.500000
7	0.213388	0.239964	0.060945	0.329339	0.247418	0.479061	1.000000	0.464818	0.154815	0.213388	0.304640	0.135074	0.144657	0.331260
8	0.464818	0.500000	0.125377	0.500000	0.479061	0.479061	0.464818	1.000000	0.165694	0.144657	0.399149	0.215449	0.329422	0.500000
9	0.165694	0.154815	0.329339	0.108068	0.135074	0.159213	0.154815	0.165694	1.000000	0.361966	0.500000	0.318251	0.144657	0.191367
10	0.188612	0.188612	0.464800	0.188489	0.186015	0.135074	0.213388	0.144657	0.361966	1.000000	0.221644	0.376269	0.165694	0.331260
11	0.399149	0.423920	0.304443	0.304640	0.304640	0.221644	0.304640	0.399149	0.500000	0.221644	1.000000	0.221644	0.304640	0.095215
12	0.479061	0.416757	0.396373	0.281618	0.281762	0.247418	0.135074	0.215449	0.318251	0.376269	0.221644	1.000000	0.318251	0.500000
13	0.345475	0.429910	0.239851	0.298026	0.318251	0.437413	0.144657	0.329422	0.144657	0.165694	0.304640	0.318251	1.000000	0.331260
14	0.500000	0.331260	0.500000	0.331260	0.331260	0.500000	0.331260	0.500000	0.191367	0.331260	0.095215	0.500000	0.331260	1.000000

Рисунок 3.6 Таблица p-значений критерия Вилкоксона-Манна-Уитни для данных на дне

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.351707	0.125874	0.125874	0.125874	0.087024	0.730111	0.351707	0.125874	0.351707	0.033566	0.125874	0.989469	0.474026
2	0.351707	1.000000	0.417524	0.417524	0.730111	0.660140	0.994458	0.786930	0.786930	0.786930	0.417524	0.417524	0.786930	0.142857
3	0.125874	0.417524	1.000000	0.994458	0.989469	0.980109	0.417524	0.417524	0.786930	0.994458	0.994458	0.994458	0.167821	0.142857
4	0.125874	0.417524	0.994458	1.000000	0.730111	0.660140	0.786930	0.417524	0.786930	0.994458	0.786930	0.786930	0.167821	0.142857
5	0.125874	0.730111	0.989469	0.730111	1.000000	0.980109	0.730111	0.351707	0.730111	0.730111	0.989469	0.989469	0.033566	0.142857
6	0.087024	0.660140	0.980109	0.660140	0.980109	1.000000	0.660140	0.282673	0.980109	0.980109	0.660140	0.980109	0.018648	0.142857
7	0.730111	0.994458	0.417524	0.786930	0.730111	0.660140	1.000000	0.786930	0.786930	0.786930	0.417524	0.417524	0.786930	0.142857
8	0.351707	0.786930	0.417524	0.417524	0.351707	0.282673	0.786930	1.000000	0.786930	0.786930	0.786930	0.167821	0.417524	0.142857
9	0.125874	0.786930	0.786930	0.786930	0.730111	0.980109	0.786930	0.786930	1.000000	0.994458	0.994458	0.786930	0.167821	0.142857
10	0.351707	0.786930	0.994458	0.994458	0.730111	0.980109	0.786930	0.786930	0.994458	1.000000	0.994458	0.994458	0.417524	0.474026
11	0.033566	0.417524	0.994458	0.786930	0.989469	0.660140	0.417524	0.786930	0.994458	0.994458	1.000000	0.786930	0.167821	0.142857
12	0.125874	0.417524	0.994458	0.786930	0.989469	0.980109	0.417524	0.167821	0.786930	0.994458	0.786930	1.000000	0.167821	0.142857
13	0.989469	0.786930	0.167821	0.167821	0.033566	0.018648	0.786930	0.417524	0.167821	0.417524	0.167821	0.167821	1.000000	0.474026
14	0.474026	0.142857	0.142857	0.142857	0.142857	0.142857	0.142857	0.142857	0.142857	0.474026	0.142857	0.142857	0.474026	1.000000

Рисунок 3.7 таблица p-значений критерия Колмогорова-Смирнова для данных на поверхности

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000000	0.989469	0.351707	0.730111	0.980109	0.660140	0.125874	0.989469	0.125874	0.351707	0.962704	0.660140	0.989469	1.0
2	0.989469	1.000000	0.351707	0.730111	0.980109	0.980109	0.351707	0.989469	0.351707	0.730111	0.962704	0.660140	0.989469	0.6
3	0.351707	0.351707	1.000000	0.125874	0.087024	0.282673	0.125874	0.351707	0.730111	0.989469	0.962704	0.980109	0.351707	1.0
4	0.730111	0.730111	0.125874	1.000000	0.660140	0.980109	0.730111	0.730111	0.351707	0.351707	0.575175	0.282673	0.730111	0.6
5	0.980109	0.980109	0.087024	0.660140	1.000000	0.660140	0.660140	0.660140	0.282673	0.282673	0.575175	0.282673	0.660140	0.6
6	0.660140	0.980109	0.282673	0.980109	0.660140	1.000000	0.660140	0.980109	0.660140	0.660140	0.962704	0.660140	0.660140	1.0
7	0.125874	0.351707	0.125874	0.730111	0.660140	0.660140	1.000000	0.730111	0.730111	0.351707	0.962704	0.282673	0.125874	0.6
8	0.989469	0.989469	0.351707	0.730111	0.660140	0.980109	0.730111	1.000000	0.730111	0.730111	1.000000	0.282673	0.730111	1.0
9	0.125874	0.351707	0.730111	0.351707	0.282673	0.660140	0.730111	0.730111	1.000000	0.730111	0.962704	0.660140	0.125874	0.6
10	0.351707	0.730111	0.989469	0.351707	0.282673	0.660140	0.351707	0.730111	0.730111	1.000000	0.575175	0.980109	0.730111	0.6
11	0.962704	0.962704	0.962704	0.575175	0.575175	0.962704	0.962704	1.000000	0.962704	0.575175	1.000000	0.575175	0.575175	0.6
12	0.660140	0.660140	0.980109	0.282673	0.282673	0.660140	0.282673	0.282673	0.660140	0.980109	0.575175	1.000000	0.660140	1.0
13	0.989469	0.989469	0.351707	0.730111	0.660140	0.660140	0.125874	0.730111	0.125874	0.730111	0.575175	0.660140	1.000000	1.0
14	1.000000	0.600000	1.000000	0.600000	0.600000	1.000000	0.600000	1.000000	0.600000	0.600000	0.600000	1.000000	1.000000	1.0

Рисунок 3.8 таблица p-значений критерия Колмогорова-Смирнова для данных на дне

Для данных БПК-5 на дне реки Охта ни один критерий не отклоняет нулевую гипотезу об однородности выборок. Главной причиной является то, что на показатели на дне реки обладают не такой сильной изменчивостью под воздействием внешних факторов. При этом тест Колмогорова-Смирнова для двух выборок для данных на поверхности обнаружил наименьшее число пар выборок, для которых отклоняется нулевая гипотеза.

Можно заключить, что признаки, измеренные на дне реки Охта более однородны, чем на поверхности, скорее всего, из-за большего внешнего влияния на признаки на поверхности.

3.2 Анализ связей временных рядов

Анализ связей измерений экологических данных позволит нам узнать, насколько сильно зависят результаты измерений БПК от других признаков. Если проверка однородности может помочь локализовать аномальные изменения показателей, то при анализе связей можно выявить, какие работы необходимо проводить для восстановления экологического состояния реки Охта.

Результаты расчетов представлены в виде таблиц корреляции

	pH	Temp	Oxygen	Fe	Nitrites	Phosphates
1	0.252696	0.15737	-0.402865	**0.749881**	-0.370444	0.372001
2	**0.579246**	0.125537	0.176148	-0.0848683	**0.427055**	0.103131
3	**0.633268**	0.195241	0.055211	0.258622	**0.506436**	0.351059
4	0.07794	**0.432108**	0.201788	0.161616	-0.0107633	-0.157241
5	0.2305	0.17064	-0.401463	-0.348248	-0.0292893	-0.0902496
6	**0.798784**	0.052812	-0.545818	-0.604254	0.0251706	0.0610679
7	**0.660267**	-0.0125582	0.161255	0.181636	-0.361432	0.051941
8	0.480772	0.298174	-0.298802	-0.487302	-0.088554	-0.0455972
9	0.120405	-0.176352	-0.331006	**0.683074**	0.00184658	**0.459846**
10	0.463089	**0.53021**	**0.624964**	-0.120986	0.0988659	0.099248
11	0.371102	0.000868336	-0.0531276	-0.304209	-0.2064	**0.553503**
12	0.353915	**0.405146**	**0.574346**	0.211589	-0.323636	**0.618136**
13	0.335724	**0.612585**	**0.54289**	-0.114832	-0.152986	0.294621
14	-0.20964	-0.0982819	0.252634	0.234964	-0.130958	0.284017

Рисунок 3.9 Таблица корреляции для коэффициента корреляции Пирсона для измерений на дне

	pH	Temp	Oxygen	Fe	Nitrites	Phosphates
1	0.29277	0.0500626	-0.39036	0.6	-0.428571	0.047619
2	0.166667	0.0281718	0.0555556	0	0.222222	0.222222
3	0.366234	0.228665	0.0714286	-0.285714	0.277778	0.111111
4	0.253546	0.228571	0.140859	0.109109	0.140859	-0.0845154
5	0.422577	0	-0.30989	0.0714286	0.0571429	0.0845154
6	0.428571	-0.0377964	-0.428571	-0.333333	-0.181848	0.285714
7	0.428571	-0.0363696	0.0714286	0.142857	-0.357143	-0.0714286
8	0.388889	0.285831	-0.222222	-0.327327	-0.0845154	-0.111111
9	0.333333	-0.117851	-0.333333	0.428571	0.111111	0.222222
10	0.197203	0.412479	0.611111	-0.0714286	0.140859	-0.111111
11	0.366234	0.0281718	0.0555556	-0.214286	-0.140859	0.388889
12	0.142857	0.29277	0.428571	0.2	-0.238095	0.52381
13	0.109109	0.428571	0.357143	-0.333333	0.0714286	0.214286
14	0.0555556	-0.0363696	0.333333	0.285714	-0.0555556	0.277778

Рисунок 3.10 Таблица корреляции для коэффициента корреляции Кендалла для измерений на дне

	pH	Temp	Oxygen	Fe	Nitrites	Phosphates
1	0.324337	0.0727393	-0.594619	0.657143	-0.571429	0.178571
2	0.25	0.0836827	0.133333	0	0.25	0.283333
3	0.451887	0.218495	0.142857	-0.309524	0.383333	0.2
4	0.309626	0.252101	0.133892	0.131739	0.175734	-0.133892
5	0.476992	0.025534	-0.351468	-0.0238095	-0.121849	0.0836827
6	0.619048	-0.0243975	-0.52381	-0.428571	-0.203596	0.404762
7	0.642857	-0.0598813	0.119048	0.178571	-0.404762	-0.0714286
8	0.616667	0.428587	-0.233333	-0.419169	-0.150629	-0.1
9	0.516667	-0.110663	-0.516667	0.547619	0.183333	0.483333
10	0.259416	0.612905	0.75	-0.047619	0.184102	-0.2
11	0.502096	-0.0251048	0.0833333	-0.285714	-0.184102	0.533333
12	0.214286	0.378394	0.642857	0.314286	-0.321429	0.678571
13	0.227549	0.547619	0.547619	-0.428571	0	0.357143
14	0.05	-0.0359288	0.516667	0.428571	-0.0833333	0.4

Рисунок 3.11 Таблица корреляции коэффициента корреляции Спирмена для измерений на дне

Все коэффициенты корреляции выделяют сильную корреляцию между БПК и pH. Стоит отметить, что непараметрические коэффициенты показывают, что при сильной корреляции БПК с pH корреляция с содержанием кислорода довольно низкая, либо отрицательная и наоборот. Происходит это из-за того, что большое количество кислорода повышает кислотность среды. Незначительную отрицательную корреляцию на четырнадцатой станции показал только коэффициент Пирсона.

Все коэффициенты корреляции отметили сильную связь содержания железа с БПК на первой и девятой станциях и полное отсутствие значительной обратной корреляции для фосфатов для всех станций.

	pH	Temp	Oxygen	Fe	Nitrites	Phosphates
1	0.410813	-0.000625127	0.472315	-0.18681	-0.0950648	-0.475427
2	0.271266	0.0983718	-0.448631	-0.136206	-0.000237962	-0.0334709
3	0.703424	0.00104183	-0.303849	0.0850717	0.605955	0.510871
4	0.126309	0.238744	-0.136367	0.0260913	0.0769098	0.224962
5	-0.459203	0.410363	-0.09035	-0.0876869	-0.522738	-0.464524
6	0.355182	0.224589	-0.359591	0.402623	0.257548	0.295274
7	0.209174	0.473703	0.166168	-0.595703	-0.494514	-0.549196
8	0.502485	0.29806	-0.125116	-0.47189	-0.135328	0.0707749
9	0.432507	0.683769	-0.0751426	0.251256	-0.0602451	0.388571
10	0.0460112	0.155812	0.683643	0.475776	-0.252476	-0.100514
11	0.0943495	0.173262	0.290483	-0.292013	-0.291655	0.282699
12	0.501166	0.51607	0.367021	-0.266329	0.0209473	0.647345
13	0.0635993	0.252296	0.717019	0.151932	-0.533374	0.46445
14	-0.14049	-0.175722	0.488303	0.36514	-0.571254	0.0025017

Рисунок 3.12 Таблица корреляции для коэффициента корреляции Пирсона для измерений на поверхности

	pH	Temp	Oxygen	Fe	Nitrites	Phosphates
1	0.359573	0.0281718	-0.0222222	-0.0555556	0.0666667	0.0666667
2	0.277778	0.113389	-0.333333	-0.254588	0.111111	0.0555556
3	0.466667	-0.0845154	-0.111111	-0.0845154	0.422222	0.466667
4	0.0898933	0.203091	-0.0222222	0.0555556	0.0222222	0.2
5	-0.0449467	0.514286	-0.13484	0.197203	-0.13484	-0.26968
6	0.333333	0.109109	-0.222222	0.285714	0.222222	0.277778
7	0.327327	0.25	-0.109109	-0.444444	-0.400066	-0.254588
8	0.422222	0.342997	-0.0666667	-0.388889	-0.13484	0.0222222
9	0.333333	0.514496	-0.0222222	0.111111	-0.179787	0.333333
10	-0.0449467	0.140859	0.422222	0.388889	-0.111111	-0.2
11	0.0681994	0.197203	0.333333	-0.222222	-0.333333	0.244444
12	0.359573	0.0845154	0.333333	-0.253546	-0.111111	0.511111
13	0.155556	0.222222	0.6	0.111111	-0.155556	0.288889
14	0	-0.30429	0.388889	0.333333	-0.422222	0.0222222

Рисунок 3.13 Таблица корреляции для коэффициента корреляции Кендалла для измерений на поверхности

	pH	Temp	Oxygen	Fe	Nitrites	Phosphates
1	0.455929	-0.0418414	-0.0424242	-0.2	0.0545455	0.030303
2	0.4	0.170783	-0.466667	-0.28743	0.15	-0.0166667
3	0.6	-0.108788	-0.151515	-0.092051	0.50303	0.563636
4	0.164134	0.288177	-0.0909091	0.133333	0.0181818	0.236364
5	-0.0486324	0.634454	-0.115502	0.209207	-0.151976	-0.334348
6	0.55	0.263478	-0.333333	0.357143	0.25	0.4
7	0.251502	0.372727	-0.155691	-0.638554	-0.419169	-0.299407
8	0.6	0.369761	-0.0909091	-0.566667	-0.170214	0.0909091
9	0.454545	0.680696	0.00606061	0.15	-0.206688	0.442424
10	-0.133739	0.117156	0.636364	0.5	-0.187879	-0.187879
11	0.121953	0.284521	0.381818	-0.283333	-0.357576	0.272727
12	0.425534	0.175734	0.50303	-0.343099	-0.0909091	0.636364
13	0.284848	0.25	0.769697	0.116667	-0.224242	0.418182
14	0.0364743	-0.304636	0.533333	0.533333	-0.563636	0.0666667

Рисунок 3.14 Таблица корреляции для коэффициента корреляции Спирмена для измерений на поверхности

Для корреляционных связей по данным на поверхности наблюдается схожая картина: по мере отдаления от Невы, корреляционная связь содержания кислорода и БПК-5 резко возрастает начиная с десятой станции. Из этого можно сделать вывод, что корреляционные связи для признаков на дне и на поверхности довольно схожи.

Выводы по главе

Исходя полученных значений можно сделать определенные выводы:

1. Данные экологических измерений на дне на различных станциях более однородны, чем данные измерений на поверхности за счет меньше подверженности внешним факторам. Больше всего выделяются результаты на поверхности для первой станции, которая находится ближе всего к Неве, что могло быть вполне ожидаемо.
2. Для корреляционной связи заметна изменчивость тесноты связи с pH в зависимости от тесноты связи с содержанием кислорода в воде. При этом теснота связи измерений БПК и pH почти всегда либо положительная, либо слабая. Заметна значимая корреляция БПК с pH на станциях с 6-й по 8-ю. Также непараметрические критерии показывают либо значительную положительную, либо практически отсутствие зависимости БПК от температуры. Коэффициенты корреляции для содержания железа, нитритов и фосфатов могут быть как положительными, так и отрицательными.

Заключение

В ходе работы были проведены:

1. Анализ непараметрических критериев для проверки однородностей и связей
2. Рассмотрены возможности прикладного пакета программ для анализа данных
3. Проведены расчеты непараметрических критериев на примере временных рядов измерений экологических показателей

С помощью измерений показателей и применения статистического анализа удалось определить, что на поверхности реки данные могут обладать большей изменчивостью, чем на глубине, так же предположительно установлено место, в котором присутствуют факторы, оказывающие влияние на экологическое состояние реки Охта.

Нам удалось установить, что результаты расчетов параметрических критериев заметно отличаются от непараметрических аналогов как для определения однородностей, так и связей. Учитывая также малые объемы выборок и то, что в гидрологии и экологии функции распределения переменных зачастую отличаются от нормального распределения, следует использовать для дальнейшего анализа данных реки Охта непараметрические критерии.

Список использованных источников

1. Robert O. Strobl - Network design for water quality monitoring of surface freshwaters: A review Journal of Environmental Management 87 (2008) 639–648
2. Wilks, Daniel S. - Statistical methods in the atmospheric sciences / Daniel S. Wilks. – 4-е издание (2019) - стр. 160-161
3. S Yue, CY Wang - The influence of serial correlation on the Mann–Whitney test for detecting a shift in median – Elsevier, 2002
4. Гмурман В. Е. – Теория вероятностей и математическая статистика – 9-е изд. – М.: Высш.шк., 2003
5. Елисеева И.И., Юзбашев М.М. – Общая теория статистики – 5-е изд., 2004
6. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников/ А.И Кобзарь; – М.: ФИЗМАТЛИТ, 2006 –
7. Коник О.А. Контроль качества воды, атмосферного воздуха и почвы. Учебное пособие / О.А.Коник, Т.В.Шахова; - СЫКТЫВКАР: СЛИ, 2013.- 145с.
8. Зенин А.А. Гидрохимический словарь / А.А. Зенин, Н.В. Белоусова; - Л.: Гидрометеоиздат,1988. - 240с.
9. Критерий Вилкоксона для проверки однородности выборо [Электронный ресурс] — URL: https://docplayer.ru/35735873-Kriteriy-vilkoksona-w-dlya-proverki-odnorodnosti-vyborok-v-1-2.html Дата обращения: 15.04.2020
10. Коэффициент корреляции Кенделла[Электронный ресурс] — URL: http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%BE%D1%8D%D1%84%D1%84%D0%B8%D1%86%D0%B8%D0%B5%D0%BD%D1%82_%D0%BA%D0%BE%D1%80%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D0%B8%D0%B8_%D0%9A%D0%B5%D0%BD%D0%B4%D0%B5%D0%BB%D0%BB%D0%B0 Дата обращения: 20.04.2020
11. Коэффициент ранговой корреляции Спирмена [Электронный ресурс] — URL: https://math.semestr.ru/corel/spirmen.php#:~:text=%D0%9A%D0%BE%D1%8D%D1%84%D1%84%D0%B8%D1%86%D0%B8%D0%B5%D0%BD%D1%82%20%D1%80%D0%B0%D0%BD%D0%B3%D0%BE%D0%B2%D0%BE%D0%B9%20%D0%BA%

D0%BE%D1%80%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D0%B8%D0%B8%20%D0%A1%D0%BF%D0%B8%D1%80%D0%BC%D0%B5%D0%BD%D0%B0%20%2D%20%D1%8D%D1%82%D0%BE,%D1%80%D0%B0%D0%BD%D0%B3%D0%B0%D0%BC%D0%B8%20%D0%BE%D1%82%20%D1%81%D0%BB%D1%83%D1%87%D0%B0%D1%8F%20%D0%BE%D1%82%D1%81%D1%83%D1%82%D1%81%D1%82%D0%B2%D0%B8%D1%8F%20%D1%81%D0%B2%D1%8F%D0%B7%D0%B8. Дата обращения: 23.04.2020
12. Matplotlib history [Электронный ресурс] — URL: https://matplotlib.org/users/history.html Дата обращения: 27.04.2020
13. What is NumPy? [Электронный ресурс] — URL: https://numpy.org/doc/stable/user/whatisnumpy.html Дата обращения: 27.04.2020
14. Pandas project description [Электронный ресурс] — URL: https://pypi.org/project/pandas/ Дата обращения: 27.04.2020
15. Scientific computing tools for Python [Электронный ресурс] — URL: https://www.scipy.org/about.html Дата обращения: 15.04.2020
16. The Jupyter Notebook [Электронный ресурс] — URL: https://jupyter-notebook.readthedocs.io/en/stable/notebook.html Дата обращения: 27.04.2020