1. **Разведывательный анализ макро-данных**

Используемые аббревиатуры добавок представлены в таблице 1.1

Таблица 1.1 – Аббревиатуры добавок в корм

|  |  |  |
| --- | --- | --- |
| **№** | **Аббревиатура** | **Описание** |
| 1 | ПР | Полусинтетический рацион (без добавления наночастиц) |
| 2 | ПДР | Полусинтетический рацион дефицитный по микроэлементам (дефицит Fe, Mn, Cu, Zn, Co, Mo, Se) |
| 3 | Целлюлоза | Добавка целлюлоза - 0,25 г/кг корма |
| 4 | Лактулоза | Добавка лактулоза 1 г/кг корма |
| 5 | Энтеросгель | Добавка энтеросгель 6 г/кг корма |
| 6 | АУ | Добавка активированный уголь 3 г/кг корма |
| 7 | Медь | УДЧ меди 1,7 мг/кг корма |
| 8 | Fe | УДЧ железа 17 мг/кг корма |
| 9 | Биф | Соя-бифидум - 1 г/кг корма |
| 10 | Спор | Споробактерин 0,25 мл/кг корма |
| 11 | Хит | Хитозан |

По данным корма имеется выборка из 22 куриц (по 2 для каждого вида рациона), с анализом на таксоны.

Проведем анализ различных видов корма. На рисунке 1.1 представлено распределение среднего количества общего числа таксонов.

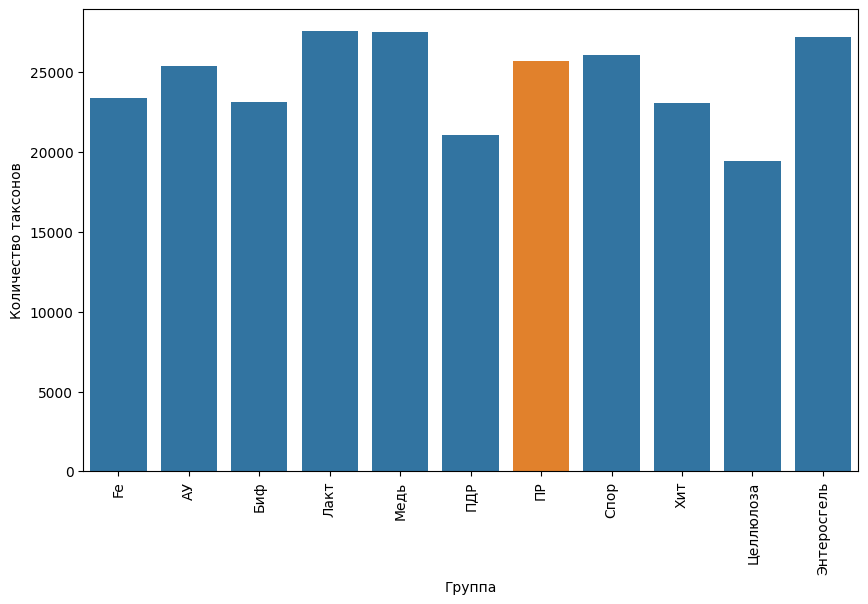


Рисунок 1.1 – Распределение среднего количества общего числа таксонов

На основе графика можно сделать вывод что среднее количество таксонов для полусинтетического рациона не выделяется среди остальных видов кормов.

Таксноны группируются по семействам, порядкам, классам и филумам. Проведем сравнительный анализ по каждому их объединений, для этого сгруппируем данные по соответствующей категории. Анализ проведем по доляем занимающей той или иной укрупненной группой от общего числа таксонов. Результаты анализа по филумам представлены на рисунке 1.2.

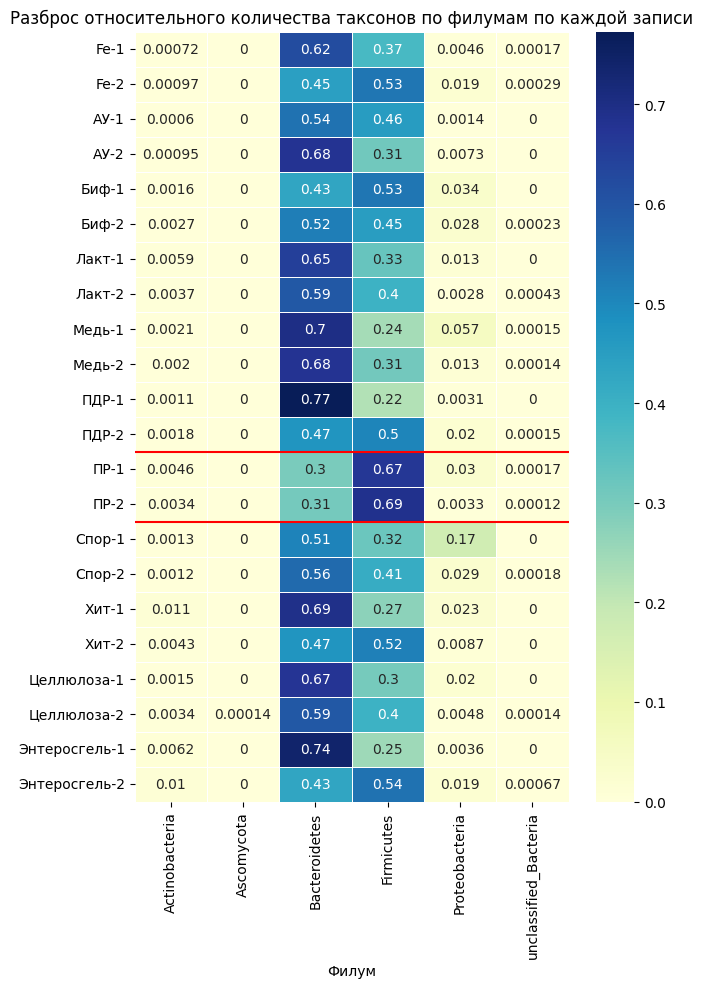


Рисунок 1.2 – Разброс относительного количества таксонов по филумам

Из рисунка можно заметить что основая доля таксонов во всех рационах приходится на филумы Bacteroidetes и Firmicutes. Причем доля Frimicutes в рационе *ПР* преобладает над долей Bacteroidetes более чем в 2 раза. Для рационов *АУ, Медь*, *Лакт*, *Спор*, *Целлюлоза* доля Bacteroidetes наоборот преобладает, для *Медь* более чем в 2 раза.

Проведем аналогичное исследование для групп по классам, соответствующее распределение долей представлено на тепловой карте рисунка 1.3.

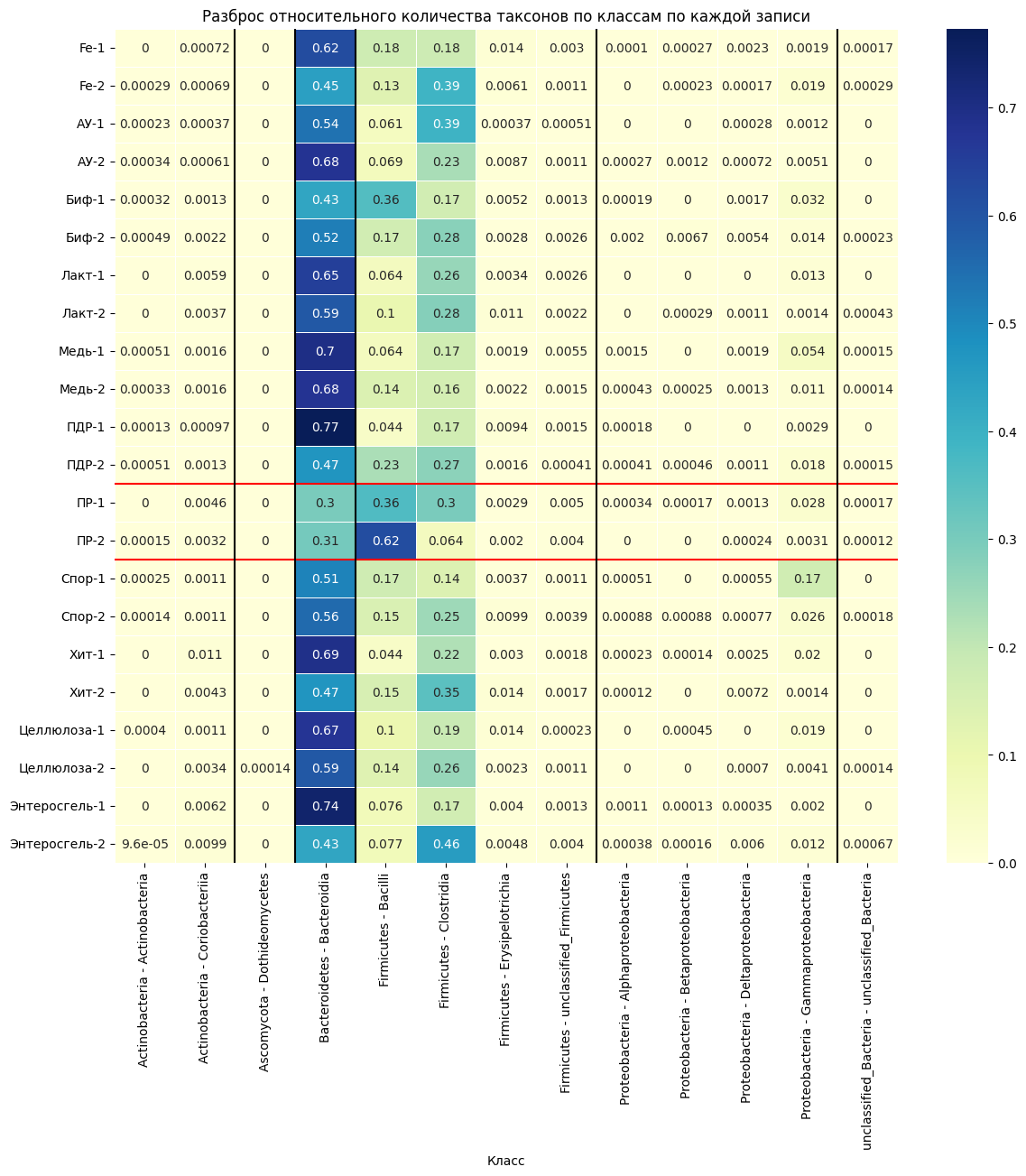


Рисунок 1.3 – Разброс относительного количества таксонов по классам

Из рисунка можно заметить что основая доля таксонов во всех рационах приходится на классы Bacteroidia, Bacilli и Clostridia.

Исследование для групп по порядкам, представлено на тепловой карте рисунка 1.4.

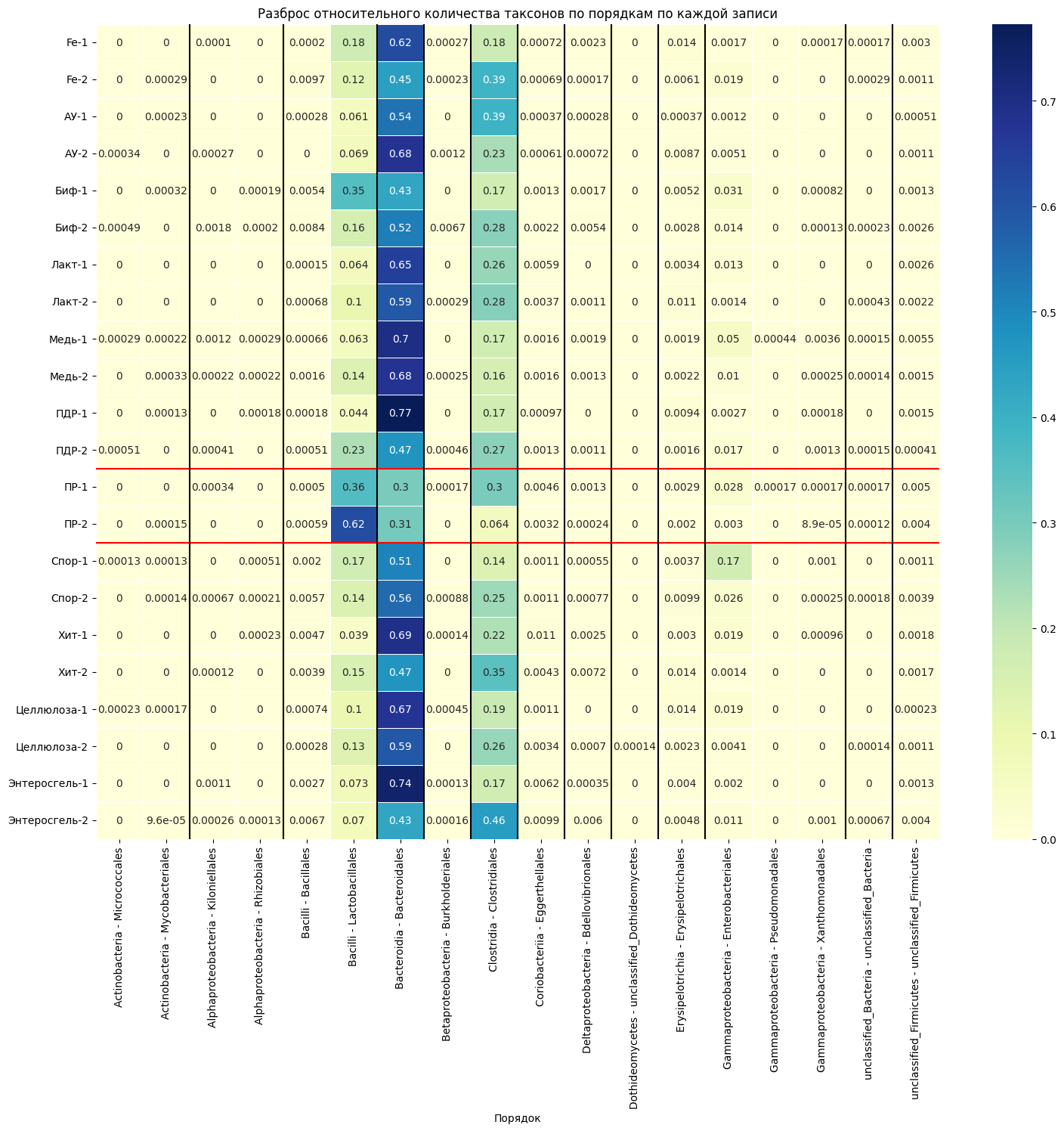


Рисунок 1.4 – Разброс относительного количества таксонов по порядкам

Из рисунка можно заметить что основая доля таксонов во всех рационах приходится на порядки Lactobacillales, Bacteroidales, Clostridiales.

Исследование для групп по семействам, представлено на тепловой карте рисунка 1.5.

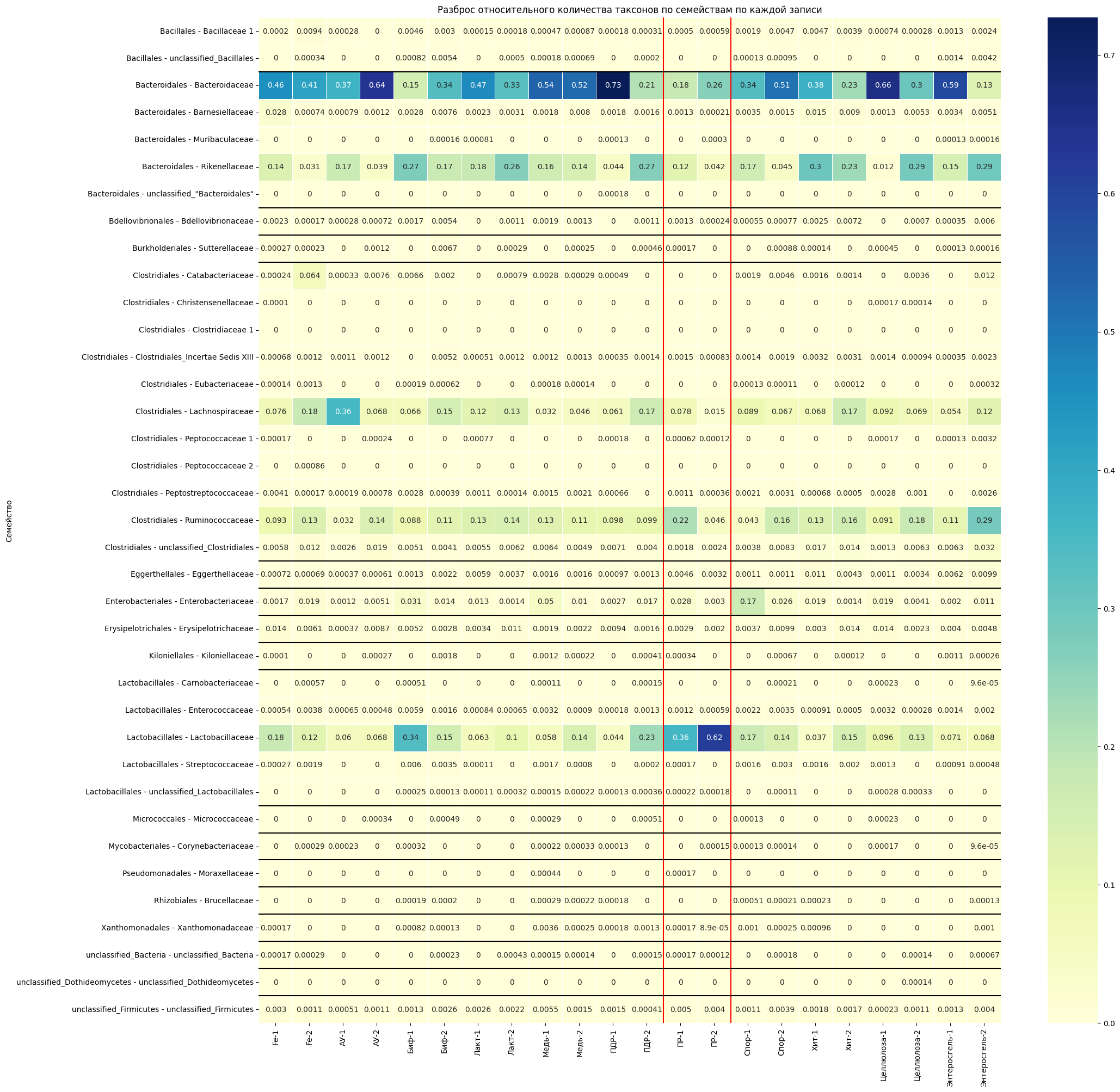


Рисунок 1.5 – Разброс относительного количества таксонов по семействам

Из рисунка можно заметить что основая доля таксонов во всех рационах приходится на семейства Bacteroidaceae, Rikenellaceae, Lachnospiraceae, Ruminococcaceae и Lactobacillaceae.

1. **Анализ разнообразия данных**
   1. **Альфа-разнообразие**

Альфа-разнообразие описывает разнообразие внутри сообщества. Существует несколько различных индексов, используемых для расчета -разнообразия, поскольку разные индексы отражают разные аспекты разнообразия и имеют разную чувствительность к разным факторам. Эти индексы были разработаны для решения конкретных исследовательских вопросов, учета различных экологических или популяционных характеристик или выделения определенных аспектов разнообразия.

* + 1. **Индекс Симпсона**

Индекс разнообразия Симпсона имеет вид:

где – встречаемость вида относительного общего содержания всех особей всех видов, – количество изучаемых видов.

Поскольку при возведении в квадрат малых отношений получаются очень малые величины, индекс Симпсона тем больше, чем сильнее доминирование одного или нескольких видов. Индекс Симпсона служит мерой связи числа степеней свободы внутривидовых и межвидовых взаимодействий. Его значение варьируется от 0 до 1, что значит бесконечное разнообразие или его отсутствие соответственно.

* 1. **Бета-разнообразие**

Бета-разнообразие измеряет расстояние между двумя или более отдельными объектами. Таким образом, оно описывает разницу между двумя сообществами или экосистемами.

Для расчёта -разнообразия используются несколько индексов, поскольку разные индексы подчёркивают различные аспекты композиционного сходства или различия между сообществами или участками.

Эти индексы были разработаны для решения конкретных исследовательских задач, работы с различными типами данных или для получения информации о различных аспектах -разнообразия.

* + 1. **Индекс Жаккара**

Индекс Жаккара измеряет долю общих видов в двух выборках, формула представлена ниже:

где:

* + – пересечение множеств и (элементы, общие для обоих множеств)
  + – объединение множеств и (все уникальные элементы из обоих множеств вместе)

1. **Кластеризация корма по химическим элементам**

Полученный датасет, выгруженный из скинутого excel-файла.

Изображение выглядит как снимок экрана

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.1 – оригинальный датасет

* + 1. **K-means**

Изображение выглядит как текст, линия, График, диаграмма

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.2 – метод Локтя для определения оптимального k

Из рисунка видно, что лучшие k для 3 или 4. При k=3 результаты плохие (нет разбиений на кластеры). Если выбрать k=4, то получаются следующие результаты:

Изображение выглядит как текст, снимок экрана, Шрифт, дизайн

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.3 – результаты кластеризации

Изображение выглядит как текст, диаграмма, линия, График

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.4 – PCA для k-means

* + 1. **Иерархическая кластеризация (AgglomerativeClustering)**

Для определения кластеров для иерархической кластеризации была использована дендорграмма:

Изображение выглядит как диаграмма, линия, График, Прямоугольник

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.5 – дендрограмма

Из дендограммы видно, что все корма объединились в 4 группы: 1 группа (цел, ПДР, энтер, лакт), 2 группа (акт.уг, хит, УДЧ железо, УДЧ меди), 3 группа (ПР), 4 группа (спор и биф). Также по результатам видно, что дендограмма полностью повторяет разбиение, которое получилось у метода K-means.

В результате работы метода получились выводы:

Изображение выглядит как текст, снимок экрана, Шрифт, меню

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.6 – Иерархическая кластеризация

Изображение выглядит как текст, диаграмма, График, линия

Контент, сгенерированный ИИ, может содержать ошибки.

Рисунок 3.7 – PCA для иерархической кластеризации

1. **Кластеризация корма по микробиоте (неудачн)**

Полученный датасет, выгруженный из скинутого excel-файла.

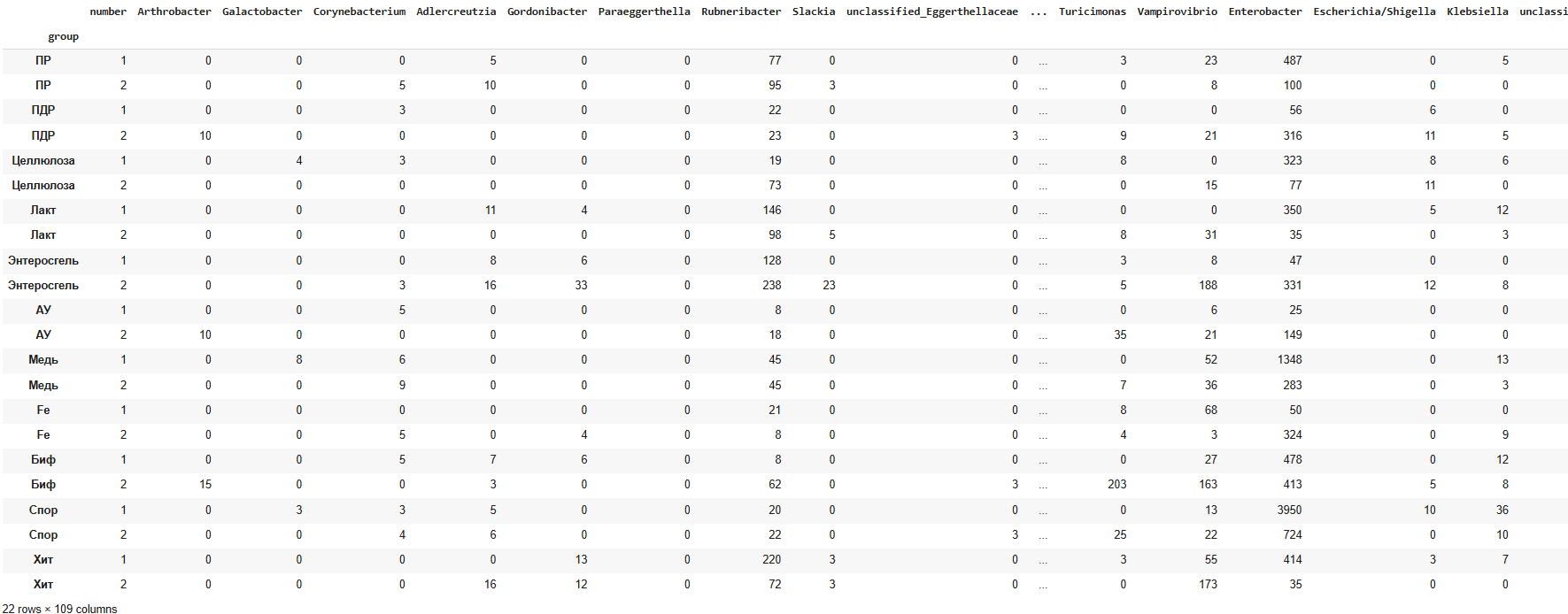


Рисунок 4.1 – оригинальный датасет

* + 1. **K-means**

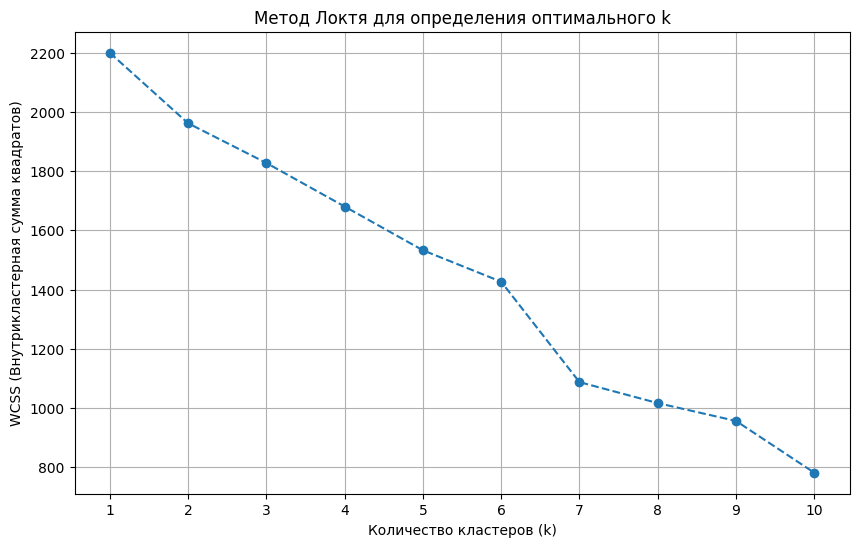


Рисунок 4.2 – метод Локтя для определения оптимального k

Из рисунка видно, что лучшие k для 6 или 7. Если выбрать k=6, то получаются следующие результаты:

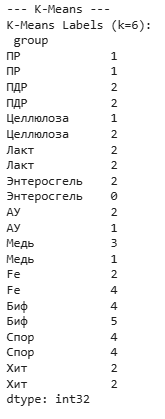


Рисунок 4.3 – результаты кластеризации

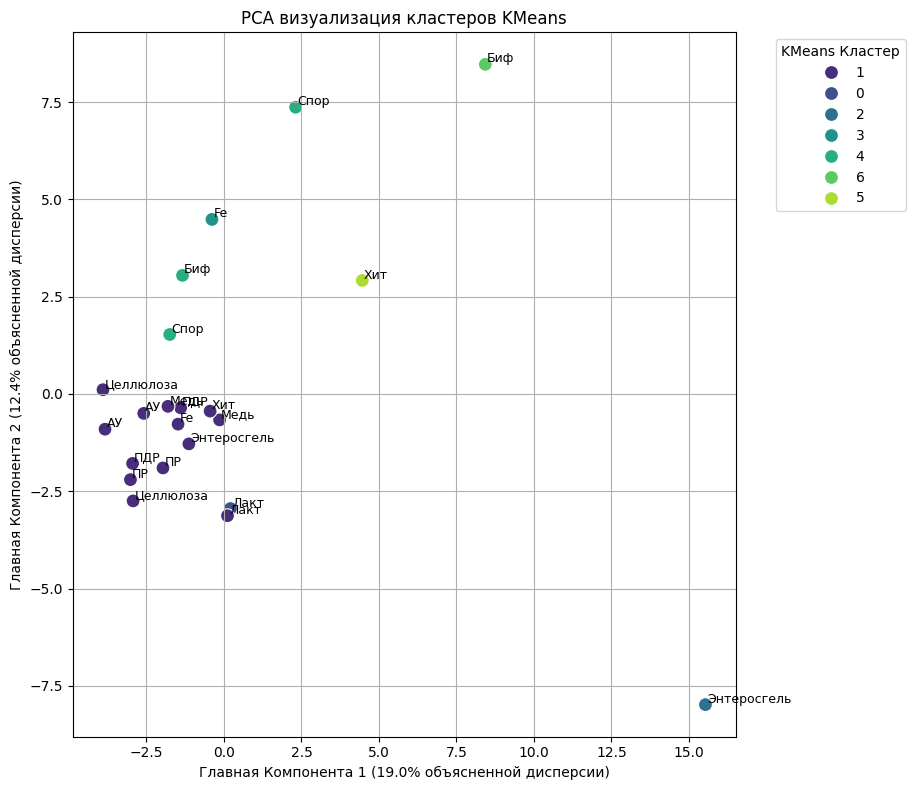


Рисунок 4.4 – PCA для k-means

* + 1. **Иерархическая кластеризация (AgglomerativeClustering)**

Для определения кластеров для иерархической кластеризации была использована дендорграмма:

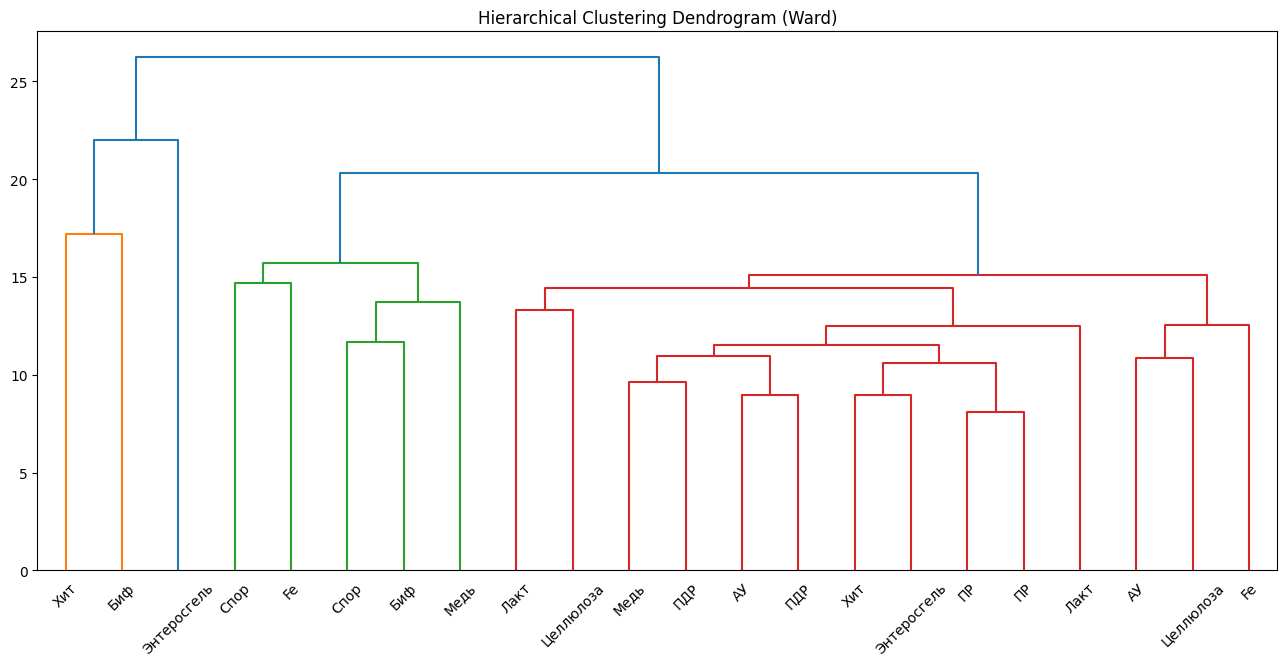


Рисунок 4.5 – дендрограмма

Из дендограммы видно, что все корма объединились в 4 группы: 1 группа (хит, биф), 2 группа (энтеросгель), 3 группа (спор1, fe, спор2, биф, медь), 4 группа (остальные).

В результате работы метода получились выводы:



Рисунок 4.6 – Иерархическая кластеризация

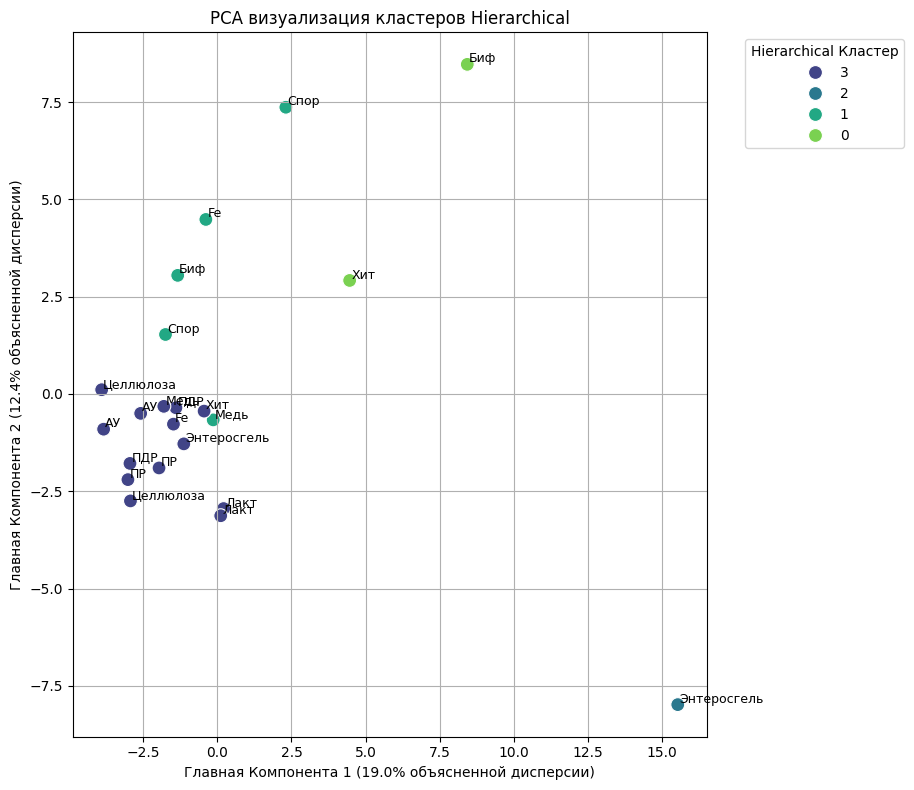


Рисунок 4.7 – PCA для иерархической кластеризации