**Руководство по использованию программы для расчёта доверительных интервалов и определения необходимого количества проб, рhhhhрррпасчет через перцентили**

Скрипт разработан для планирования исследований бентоса. По уже имеющейся выборке можно спрогнозировать, какое количество проб требуется собрать в новом месте для достижения результата с определенной точность.

!!! определение требуемого количества проб этим методом позволяет получить точные данные только для относительно небольшого числа проб, сопоставимого с половиной оригинальной выборки. В случае прогноза на число проб, более половины от оригинальной выборки, этот метод занижает результат.

**Описание программы**

Программа анализирует, как изменяется точность (ширина доверительных интервалов) при увеличении объёма данных для трёх уровней значимости (70%, 90%, 95%). Основной результат — расчёт количества проб, необходимых для достижения заданной точности.

**Основные шаги алгоритма:**

1. **Загрузка данных** из файла sample.csv.
2. **Генерация псевдовыборок** (1000 перестановок исходных данных).
3. **Расчёт накопленных средних** для каждой псевдовыборки.
4. **Определение доверительных интервалов** (70%, 90%, 95%) с помощью перцентилей.
5. **Нормализация ширины интервалов** относительно среднего исходной выборки.
6. **Построение регрессионных моделей** (уравнение Хилла) для предсказания необходимого числа проб.
7. **Визуализация результатов** (графики и таблицы).

**Требования**

* **R** (версия 4.0+)
* **Пакеты R**:
  + ggplot2 (визуализация)
  + dplyr (обработка данных)
  + tidyr (преобразование данных)

Установка пакетов:

R

install.packages(c("ggplot2", "dplyr", "tidyr"))

**Инструкция по запуску**

1. **Подготовьте входные данные**:
   * Создайте файл sample.csv в рабочей директории.
   * Данные должны быть в **одном столбце** с заголовком x (десятичный разделитель — точка).
   * Пример структуры файла:

x

1.5

2.3

0.8

...

1. **Запустите скрипт** в R:
   * Скопируйте код в файл script.R.
   * Выполните:

R

source("script.R")

1. **Полученные результаты**:
   * Number of sumples.txt — таблица с расчётом числа проб для заданных уровней точности (20%, 30%, 40%, 50%, 60%).
   * hill\_models\_summary.txt и hill\_models\_summary\_formatted.txt — параметры регрессионных моделей.
   * combined\_hill\_plot.png — график зависимости ширины доверительного интервала от размера выборки.

**Интерпретация результатов**

**1. Основной вывод (**Number of sumples.txt**)**

Файл содержит таблицу вида:

text

Результаты расчетов (округленные до целых)

=========================================

Дата: 2024-03-15 12:30

70% 90% 95%

x=20 50 85 110

x=30 30 55 70

x=40 20 40 50

x=50 15 30 40

x=60 10 25 30

* **Строки (**x=20, x=30, ...**)** — требуемая точность (ширина доверительного интервала в % от среднего).
* **Столбцы (**70%, 90%, 95%**)** — уровни значимости.
* **Значения в таблице** — необходимое количество проб для достижения заданной точности.

**2. Параметры моделей (**hill\_models\_summary.txt**)**

Содержит коэффициенты уравнений Хилла для каждого уровня значимости:

text

Model top EC50 n R\_squared AIC Equation

norm\_range\_70 150.2 45.3 1.2 0.98 1200 y = 150.2 / (1 + (45.3/x)^1.2)

...

* top — максимальный размер выборки (асимптота).
* EC50 — значение x, при котором достигается половина от top.
* n — коэффициент наклона.
* R\_squared — качество модели (ближе к 1 — лучше).

**3. График (**combined\_hill\_plot.png**)**

* **Ось X** — ширина доверительного интервала (%).
* **Ось Y** — размер выборки.
* **Цветные линии** — предсказанные значения для 70%, 90%, 95%.
* **Точки** — реальные данные.

**Настройка параметров**

* **Изменение уровней значимости**:
  + В коде замените значения probs в разделах percentil70, percentil90, percentil95.
* **Изменение диапазона точности**:
  + В конце скрипта измените x\_values <- c(20, 30, 40, 50, 60) на нужные значения.

**Заключение**

Программа позволяет:

* Оценить, сколько проб нужно для достижения заданной точности.
* Сравнить поведение доверительных интервалов при разных уровнях значимости.
* Визуализировать зависимость точности от объёма данных.