

Отчет о Лабораторной №2

Работу выполнила: Лемешко Юлия Романовна

Введение

В данном исследовании проводится комплексный статистический анализ трех различных датасетов с целью проверки гипотез о распределении данных и построения доверительных интервалов для параметров распределений.

Датасеты и задачи:

1. **Babyboom** - анализ веса младенцев (нормальность), времени между рождениями (экспоненциальное распределение) и частоты рождений (распределение Пуассона)
2. **Euroweight** - проверка нормальности распределения веса монет в целом и по партиям
3. **Iris** - анализ нормальности распределения длины лепестков по видам ирисов

Цель - освоение проверки статистических гипотез, интервального оценивания

1.

Задание

По датасету "babyboom"

· Проверьте вес младенцев на нормальность. Сначала используйте все данные, не разделяя детей по полу, затем для группы мальчиков и девочек по отдельности. При проверке гипотез используйте точечные оценки параметров. Постройте доверительные интервалы для параметров нормального распределения.

Результат

Данные по весу младенцев в целом не совсем похожи на нормальное распределение - всего 2/6 критериев говорит о нормальности. Широковатые доверительные интервалы - особенно у дисперсии говорят о некоторой неопределенности в данных

Когда мы делим данные на мальчиков и девочек 4/6 тестов поддерживают нормальное распределение девочек и 5/6 мальчиков - мальчики ближе к нормальному распределению. Их доверительные интервалы тоже достаточно большие и накладываются друг на друга в среднем.

Задание

· Проверьте гипотезу о том, что время между рождением детей подчиняется экспоненциальному распределению. При проверке гипотезы используйте

Введение
точечные оценки параметров.
Результат
Проверяю данные на двух стат тестах - оба говорят об экспоненциальном распределении, т.е подтверждают гипотезу.
Задание
· Проверьте гипотезу, подчиняется ли количество рождений в час для каждого часа распределению Пуассона. При проверке гипотезы используйте точечные оценки параметров.
Результат
Проверка на 3х тестах - все отвергают нулевую гипотезу о том, что это распределение Пуассона.
2.
Задание
По датасету "Euroweight" Проверить гипотезу о нормальном распределении веса монет (сначала объединить все монеты в одну выборку, потом по каждому пакету в отдельности). При проверке гипотез используйте точечные оценки параметров.
Результат
При объединении всех монет в одну выборку данные показывают 2/6 принявших нулевую гипотезу критерия - слабая нормальность распределения. После разбиения данных по батчам - их 8 1,2,4,5,7 - показывают 6/6 пройденных тестов на нормальность, остальные 2/6
Задание
· Постройте доверительные интервалы для параметров нормального распределения.
Результат
Для всех данных очень низкая ширина доверительных интервалов, что говорит о такой же низкой неопределенности данных, данные доверительных интервалов по батчам также накладываются друг на друга, что указывает их схожесть
3.
Задание
По датасету 'iris' · Проверить гипотезу о нормальном распределении длины цветков, сгруппировав их по типу ириса. При проверке гипотез используйте точечные оценки параметров.
Результат

Введение

Распределение первого класса - 4/6 нормальное, второго и третьего 5/6 нормальные - предполагаем что распределение всех данных нормальное.

Задание

· Постройте доверительные интервалы для параметров нормального распределения.

Результат

Доверительные интервалы классов не накладываются друг на друга и узкие - определенность в данных и разница между группами данных по классам.

Вывод

. Babyboom dataset

- **Вес младенцев:** Нормальность подтверждается только при разделении по полу (особенно для мальчиков)
- **Время между рожденьями:** Соответствует экспоненциальному распределению ✓
- **Рождаемость по часам:** Не соответствует распределению Пуассона ×

2. Euroweight dataset

- **Вес монет:** Нормальность слабая в объединенной выборке, но сильная в 5 из 8 отдельных батчей
- **Точность оценок:** Очень узкие доверительные интервалы → высокая точность измерений

3. Iris dataset

- **Длина лепестков:** Нормальное распределение подтверждено для всех видов ирисов
- **Различия между видами:** Четкие статистические различия (непересекающиеся ДИ)