МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

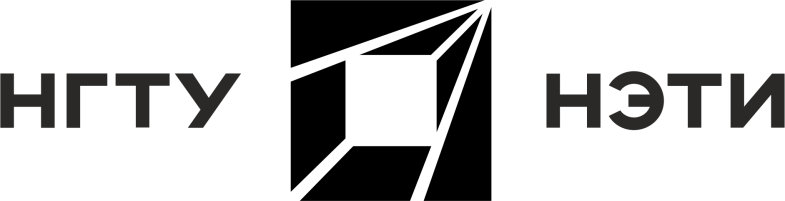
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра теоретической и прикладной информатики



**ЛАБОРАТОРНАЯ РАБОТА №2**

по дисциплине: Компьютерные технологии моделирования и анализа данных

на тему: Экспериментальное исследование робастности оценок

  Факультет: ФПМИ

Группа: ПММ-21

Выполнили: Сухих А.С., Черненко Д.

Вариант №2

Проверил: д.т.н., профессор Лемешко Б.Ю.

Дата выполнения: 10.11.22

Отметка о защите:

Новосибирск 2022

**Цель работы.** Исследование устойчивости оценок на наличие в выборке аномальных наблюдений. Исследование эффективности параметрической процедуры исключения аномальных наблюдений при использовании робастных оценок. Построение функций влияния Хампеля для ОМП. Исследование распределений статистик типа Граббса, предназначенных для анализа на аномальность сразу нескольких наблюдений.

|  |  |  |
| --- | --- | --- |
| № |  |  |
| 2 | Нормальное | Нормальное с масштабом 5 |

**Ход выполнения**

**1.** **Исследовать робастность оценок максимального правдоподобия (ОМП); ОМП по группированным данным; MD-оценок, минимизирующих расстояния, задаваемые статистиками Колмогорова, и Мизеса; -оптимальных оценок по выборочным квантилям**.



Были смоделированы выборки с засорением



с различной долей засорения v=0, 0.1, 0.2. Объем выборки — 1000.

Оценены параметры нормального распределения требуемыми методами:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Доля v  Оценка | v = 0 | | v = 0.1 | | v = 0.2 | |
| μ | σ | μ | σ | μ | σ |
| ОМП | 0.0025 | 0.9706 | 0.0368 | 1.6871 | 0.0556 | 2.2367 |
| ОМП по групп. | -0.0026 | 0.9862 | -0.0009 | 1.1264 | 0.0041 | 1.3357 |
| Колмогорова | -0.0065 | 0.9789 | -0.019 | 1.1572 | -0.0243 | 1.4416 |
| К-М-С, ɷ2 | -0.0116 | 0.976 | -0.0136 | 1.0969 | -0.0169 | 1.247 |
| L | -0.0013 | 0.9733 | -0.0000 | 1.0972 | 0.0028 | 1.2674 |

Группированная выборка составлена с помощью равночастотного группирования с числом интервалов 10.

**2. Отбраковка аномальных наблюдений. Используя параметрическую процедуру отбраковки аномальных наблюдений очистить полученные в п.1 выборки от аномальных наблюдений и проверить, как изменились результаты применения критериев для проверки согласия эмпирического распределения с распределением после удаления части наблюдений.**



Для выборки с v=0 аномальные наблюдения не были обнаружены. У выборки v=0.1 было удалено 17 наблюдений. У выборки v=0.2 было удалено 58 наблюдений.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Доля v  Оценка | | v = 0.1 | | v = 0.2 | |
| μ | σ | μ | σ |
| ОМП | исходная | 0.0368 | 1.6871 | 0.0556 | 2.2367 |
| чистая | 0.0044 | 1.2494 | 0.0012 | 1.4283 |
| ОМП по групп. | исходная | -0.0009 | 1.1264 | 0.0041 | 1.3357 |
| чистая | -0.0084 | 1.0794 | -0.0146 | 1.1754 |
| Колмогорова | исходная | -0.019 | 1.1572 | -0.0243 | 1.4416 |
| чистая | -0.0236 | 1.1015 | -0.0388 | 1.243 |
| К-М-С, ɷ2 | исходная | -0.0136 | 1.0969 | -0.0169 | 1.247 |
| чистая | -0.017 | 1.0672 | -0.0260 | 1.1548 |
| L | исходная | -0.0000 | 1.0972 | 0.0028 | 1.2674 |
| чистая | -0.0072 | 1.0683 | -0.0213 | 1.1601 |

**3. Построить функции влияния Хампеля для ОМП и ОМП по группированным данным.**

Функции влияния для параметров сдвига и масштаба нормального распределения представлены на рисунках 1 и 2:

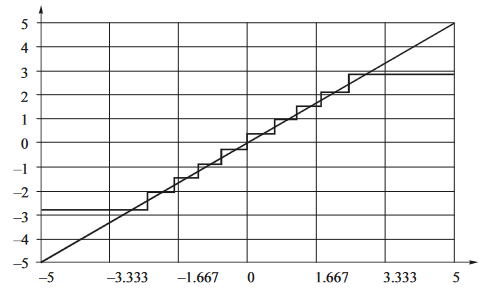


Рисунок 1. Функции влияния для параметра сдвига по не группированным (прямая) и сгруппированным (ступенчатая) данным

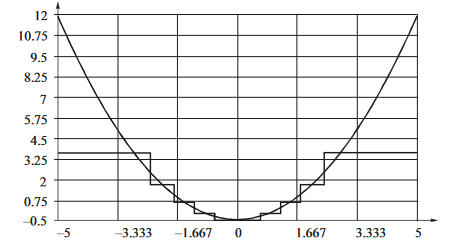


Рисунок 2. Функции влияния для параметра масштаба по не группированным (прямая) и сгруппированным (ступенчатая) данным

**4. Исследовать распределения статистик критериев типа Граббса, предназначенных для анализа на аномальность сразу двух минимальных и двух максимальных наблюдений в предположении о принадлежности выборки нормальному закону. Построить эмпирические распределения для статистик критериев типа Граббса, найти приближенные значения процентных точек. Применить критерий Граббса для отбраковки аномальных наблюдений по выборкам с засорением, полученным в п.1.**

Эмпирическое распределение статистик критериев типа Граббса для анализа на аномальность двух минимальных и двух максимальных наблюдений:

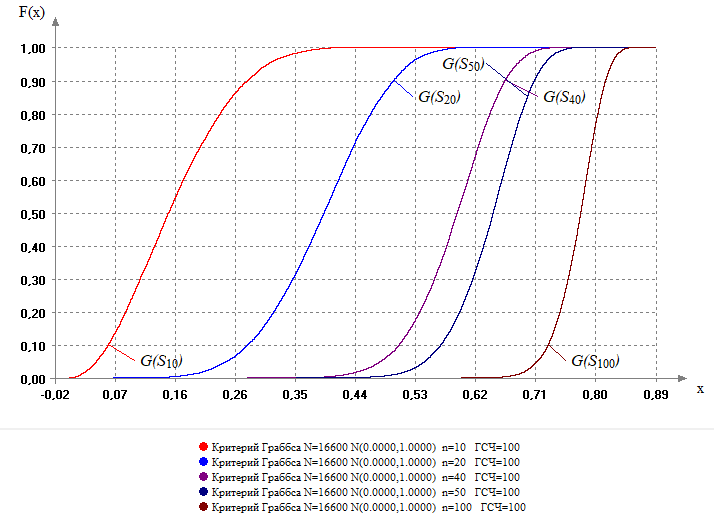


Рисунок 3. Распределения статистик для различных объемов выборки

Были определены значения статистик в нижних и верхних процентных точках:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Объем | Нижние процентные точки | | | | | Верхние процентные точки | | | | |
| 0,1% | 1% | 2,5%. | 5%. | 10%. | 0,1% | 1% | 2,5%. | 5%. | 10%. |
| n = 10 | 0.009 | 0.022 | 0.033 | 0.045 | 0.063 | 0.422 | 0.368 | 0.336 | 0.306 | 0.274 |
| n = 20 | 0.133 | 0.184 | 0.215 | 0.242 | 0.274 | 0.61 | 0.565 | 0.541 | 0.52 | 0.495 |
| n = 40 | 0.359 | 0.421 | 0.45 | 0.475 | 0.503 | 0.741 | 0.713 | 0.697 | 0.682 | 0.664 |
| n = 50 | 0.437 | 0.493 | 0.521 | 0.542 | 0.566 | 0.774 | 0.75 | 0.736 | 0.724 | 0.708 |
| n = 100 | 0.643 | 0.683 | 0.698 | 0.712 | 0.729 | 0.857 | 0.84 | 0.832 | 0.825 | 0.816 |

По выборкам, полученным в п.1. Были отбракованы аномальные наблюдения с помощью критерия Граббса для выборок с засорением 0,1 и 0,2:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Аномальных  точек  Уровень засорения | (0,1) | (1,0) | (1,1) | (1,2) | (2,1) | (2,2) |
| V=0.1 | S=7,4575  p=1, не отклоняется | S=5,9005  p=1, не отклоняется | S=0,9095,  p=0, отклоняется | S=0,8617,  p=0, отклоняется | S=0,886,  p=0, отклоняется | S=0.8383,  p=0, отклоняется |
| V=0.2 | S=6,1411  p=1, не отклоняется | S=5,0825  p=1, не отклоняется | S=0,9364,  p=0, отклоняется | S=0,9032,  p=0, отклоняется | S=0,9172,  p=0, отклоняется | S=0.8842,  p=0, отклоняется |

**Вывод**

По результатам, полученным в ходе выполнения лабораторной работы было установлено, что наиболее робастным методом для оценки параметров распределения является метод оценки максимального правдоподобия по группированным данным. Существенный вклад в робастность оказывает группирование данных, позволяющее снизить влияние выбросов на оценку параметров. Выбросы в выборках создавались с помощью засоренния распределения.

Также засоренное распределение было подвергнуто отбраковке аномальных наблюдений, что также позволило существенно повысить робастность оценки параметров.

Было произведено исследование эмпирического распределения статистик критерия Граббса для анализа на аномальность двух максимальных и двух минимальных наблюдений. Определены приближенные значения верхних и нижних процентных точек.

Также с помощью данного критерия были отбракованы одно минимальное и одно максимальное наблюдение.

Вопросы:

1. Что такое асимптотически эффективные оценки?
2. Как построить функцию влияния? Ответ: вычисляется количество информации Фишера как в первой лабе, потом рассчитывается производная для функции плотности распределения. Параметр может определяться как (x – mu) / sigma с фиксированием одного параметра и изменением другого
3. В критериях Граббса простая или сложная гипотеза?
4. Что такое процентные точки в критериях Граббса? Ответ: вероятности, при которых попадают статистики. Могут быть 0.1,0.05 и прочее. Собрать таблицу с выводом этих точек и соответствующими значениями статистик.
5. Почему критерий Граббса это критерий проверки нормальности, если он используется только для отбраковки аномальных значений?
6. Что моделируется при проверке на нормальность по критерию Граббса?
7. Если критерий гриббса проверяет на соответствие нормальности крайних наблюдений, то почему по отдельности он наблюдения не отклоняет,а когда делаешь и мин. и макс, то отклоняет?