Липецкий государственный технический университет

Кафедра прикладной математики

Индивидуальное домашнее задание №4 по курсу «Организация обработки данных» Вариант 2

| Студент | | Богомолов Е.А |
|------------------------------|---------------|-------------------|
| | подпись, дата | фамилия, инициалы |
| Группа | | |
| Руководитель | | |
| | | Левина Л.В. |
| ученая степень ученое звание | пошись пата | даминия випимер |

Липецк 2022 г.

Содержание

| Зғ | адание каф | редры | 3 |
|----|------------|-----------------------|----|
| 1. | Краткая | георитическая справка | 4 |
| 2. | Ход рабоз | гы | 6 |
| | 2.01 | 10 значений | 6 |
| Bı | ыводы | | 12 |

Задание кафедры

Дана выборка – 44 значения:

- 1. Для первых 10 элементов найти:
 - $\hat{M}(x), \hat{D}(x)$ обычным способом.
 - $\widetilde{M}(x),\widetilde{D}(x)$ по методу бутстреп.
 - $\widehat{D}\widetilde{M}(x), \widehat{D}\widetilde{D}(x).$
- 2. Для следующих 20 элементов:
 - $\hat{M}(x), \hat{D}(x)$ обычным способом.
 - $\widetilde{M}(x), \widetilde{D}(x)$ по методу бутстреп.
 - $\widehat{D}\widetilde{M}(x), \widehat{D}\widetilde{D}(x).$
- 3. Для всей выборки найти M(x), D(x)

1. Краткая теоритическая справка

В статистике и анализе данных бутстрапом называют статистическую процедуру, основанную на выборке с замещением для определения точности (смещения) выборочных оценок дисперсии, среднего, стандартного отклонения, доверительных интервалов и других структурных характеристик совокупности.

Метод разработан и впервые опубликован в 1972 году Бредли Эфроном.

В основе идеи бутстрапа лежит оценка структурных характеристик генеральной совокупности на основе перевыборки (resampling) из выборки. Иными словами, перевыборка по отношению к выборке рассматривается как выборка по отношению к генеральной совокупности.

Алгоритм работы метода следующий:

- Из генеральной совокупности формируется случайная выборка из N(t) наблюдений.
- К выборке применяется случайная перевыборка с возвратом (псевдовыборка) того же объема, но в которую некоторые наблюдения могут попасть несколько раз, а другие не попасть совсем.
- Процедура перевыборки повторяется достаточно много раз (несколько десятков, сотен или даже тысяч), и для каждого случая вычисляется среднее.
- Из полученного набора средних значений вычисляется среднее и рассматривается как среднее всей генеральной совокупности.

Важнейшим преимуществом бутстрапа являются: простота реализации; отсутствие необходимости гипотез о параметрах распределения данных; возможность оценивания многих статистических характеристик (среднего, дисперсии, стандартного отклонения, доверительных интервалов, квантилей, коэффициентов корреляции и др.).

К недостатку метода можно отнести использование малореалистичного предположения о независимости перевыборок и значительные вычислительные затраты при их многократном построении.

Метод оказывается особенно полезным, когда теоретическое распределение данных неизвестно или объем выборки мал для прямой статистической оценки.

В анализе данных бутстрап используется для оценки точности аналитических моделей.

2. Ход работы

2.0..1 10 значений

$$\widehat{M}(x) = \frac{1}{10} \sum_{u=1}^{10} x_u$$
, $\widehat{D}(x) = \frac{1}{10-1} \sum_{u=1}^{10} (x_u - M)^2$

Рисунок 1 – Формула для пункта «а»

$$M_{i}^{*} = \frac{1}{10} \cdot \sum_{j=1}^{10} V_{ij} X_{ij} \quad M^{*} = \frac{1}{10} \cdot \sum_{i=1}^{10} M_{i}^{*} \Delta = M^{*} - \widehat{M} \quad D_{i}^{*} = \frac{1}{9} \cdot \sum_{j=1}^{10} V_{ij} (X_{j} - M_{i})^{2}$$

$$D^{*} = \frac{1}{10} \sum_{i=1}^{10} D_{i}^{*} \Delta = D^{*} - \widehat{D}$$

Рисунок 2 – Формула для пункта «б»

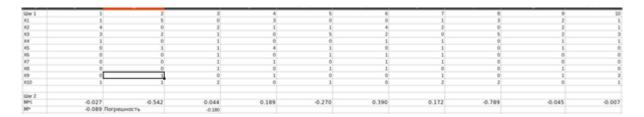


Рисунок 3 — Формирование подвыборок и вычисление математического ожидания для 10 повторений

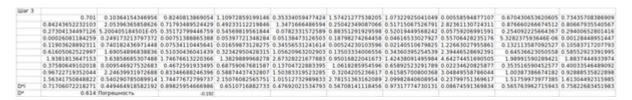


Рисунок 4 – Вычисление дисперсии для 10 повторений

| M* | -0.125 Погрешность | -0.216 |
|----|--------------------|--------|
| | | 0.220 |

Рисунок 5 – Математическое ожидание для 30 повторений

Рисунок 6 – Дисперсия для 30 повторений

| M* -0.085 Погрешность -0 | .177 | |
|--------------------------|------|--|
|--------------------------|------|--|

Рисунок 7 — Математическое ожидание для 50 повторений

| D* | 0.622 Γ | Тогрешность | -0.184 |
|----|---------|-------------|--------|

Рисунок 8 – Дисперсия для 50 повторений

| Оценки | |
|-----------|------|
| n=10 | |
| (M*i-M)^2 | 0.01 |
| DM | 0.16 |
| (D*i-D)^2 | 0.01 |
| DD | 0.10 |
| n=30 | |
| (M*i-M)^2 | 0.56 |
| DM | 0.17 |
| (D*i-D)^2 | 0.30 |
| DD | 0.10 |
| n=50 | |
| (M*i-M)^2 | 0.01 |
| DM | 0.14 |
| (D*i-D)^2 | 0.00 |
| | |

| | | _ | |
|----|------|-------------|-------|
| M* | 0.09 | Погрешность | -0.04 |

Рисунок 10 – Математическое ожидание для 10 повторений

| D* | 0.56 | Погрешность | -0.27 |
|----|------|-------------|-------|

Рисунок 11 – Дисперсия для 10 повторений

| M* | 0.10 Пог | решность | -0.03 |
|----|----------|----------|-------|

Рисунок 12 – Математическое ожидание для 30 повторений

| D* | 0.58 Погрешность | -0.26 |
|----|------------------|-------|
| | | |

Рисунок 13 – Дисперсия для 30 повторений

| 1 | M * | 0.10 | Погрешность | -0.73 |
|---|------------|------|-------------|-------|
| | | | | |

Рисунок 14 — Математическое ожидание для 50 повторений

| D* | 0.58 | Погрешность | -0.25 |
|----|------|-------------|-------|

Рисунок 15 – Дисперсия для 50 повторений

| Оценки | |
|-----------|------|
| n=10 | |
| (M*i-M)^2 | 0.07 |
| DM | 0.03 |
| (D*i-D)^2 | 0.00 |
| DD | 0.13 |
| n=30 | |
| (M*i-M)^2 | 0.04 |
| DM | 0.03 |
| (D*i-D)^2 | 0.37 |
| DD | 0.11 |
| n=50 | |
| (M*i-M)^2 | 0.03 |
| DM | 0.04 |
| (D*i-D)^2 | 0.09 |
| DD | 0.12 |

Рисунок 16 – Оценки для математического ожидания и дисперсии

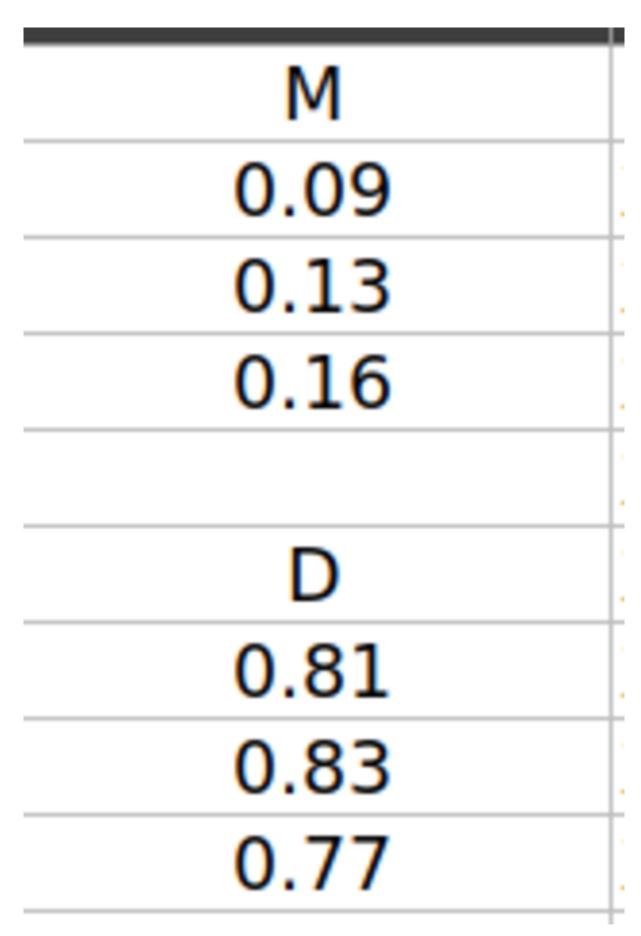


Рисунок 17 – Математическое ожидание и дисперсия для всей выборки

Выводы

Из материала, представленного в отчёте, можно сделать вывод, что метод бутстреп в сравнении с методом джекнайф имеет большую погрешность, но применение данного метода позволит легче интерпретировать полученные оценки.