|  |  |  |
| --- | --- | --- |
|  | | |
| Федеральное государственное бюджетное  образовательное учреждение высшего образования «Новосибирский государственный технический университет» | | |
|  | | |
| Кафедра теоретической и прикладной информатики | | |
| Лабораторная работа № 1 | | |
| по дисциплине «Статистический анализ нечисловых данных» | | |
|  | | |
|  | | |
|  | Бригада 10 | Субботин дмитрий |
| Группа ПМИ-12 | Швадченко артем |
| Вариант 2 |  |
|  |  |
|  |  |
|  |  |
| Преподаватели | Тимофеева анастасия юрьевна |
|  |  |
| Новосибирск, 2024 | | |

1. **Анализ исходных данных**

A1: {b, a} (качественные данные)

A2: [15,17;76,75] (количественные данные)

A3: [0;26,335] (количественные данные)

A4: {b, a} (качественные данные)

A8: [0;20] (количественные данные)

A9: {b, a} (качественные данные)

A11: [0;67] (количественные данные)

A12: {b, a} (качественные данные)

A13: {b, a} (качественные данные)

A14: [0;2000] (количественные данные)

A16: -, + (переменная класса)

1. **Ход работы**

**2.1 Из набора данных Вашего варианта сформируйте два массива данных:**

* массив количественных данных (все количественные данные оставить без изменений, для всех качественных данных кроме переменной класса произвести калибровку с учетом априорного шанса с поправкой Лапласа, значения переменной класса задать как 1, если положительный класс, 0 иначе);
* массив качественных данных (для всех качественных данных кроме переменной класса произвести калибровку с учетом априорного шанса с поправкой Лапласа, значения переменной класса задать как 1, если положительный класс, 0 иначе; для всех количественных данных произвести дискретизацию с равной частотой, в качестве границ интервалов взять выборочные квантили порядка 0, 0.2, 0.4, 0.6, 0.8, 1).

**Массив количественных данных:**

До калибровки:

Изображение выглядит как текст, число, Шрифт, снимок экрана

Автоматически созданное описание

После:

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

**Массив качественных данных:**

Границы интервалов:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Данные после дискретизации:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

**2.2 Рассчитайте все парные показатели взаимосвязи между переменными из набора количественных данных. Показатель взаимосвязи – коэффициент корреляции Пирсона**

Изображение выглядит как текст, Шрифт, линия, белый

Автоматически созданное описание

Парные показатели взаимосвязи между данными количественных переменных:

**Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание**

**2.3 Проверьте гипотезы о значимости взаимосвязей между переменными: о значимости коэффициента корреляции Пирсона с помощью стандартной процедуры и с помощью перестановочного критерия**

Уровень значимости

Проверка гипотезы о значимости взаимосвязей между переменными в массиве с количественными данными:

**Стандартная процедура:**

P-value:

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Если, значение p-value меньше, чем 0.05, то гипотеза о независимости отвергается (существует статистически значимая связь)

0-не отвергается

1-отвергается

Изображение выглядит как текст, снимок экрана, число, Параллельный

Автоматически созданное описание

**Перестановочный критерий:**

**Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание**

Если значение в столбце перестановочного критерия меньше, чем значение коэффициента корреляции Пирсона, то гипотеза о независимости отвергается

0-не отвергается

1-отвергается

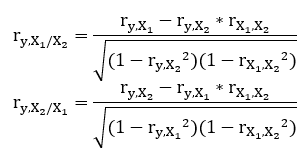
**Изображение выглядит как текст, снимок экрана, число, Параллельный

Автоматически созданное описание**

**Вывод:**

1. Переменные A2, A3, A4, A8, A9 и А11 демонстрируют статистически значимые взаимосвязи с другими переменными как по стандартной процедуре, так и по перестановочному критерию. Это указывает на то, что изменения в этих переменных могут оказывать влияние на другие переменные.
2. Переменные A1, A12, А13 и A14 часто показывают отсутствие значимых взаимосвязей с другими переменными, что может указывать на их потенциальную независимость от большинства других переменных. Это может быть важным моментом для определения менее значимых факторов в вашем анализе.
3. Сходство между результатами стандартной процедуры и перестановочного критерия говорит о надежности выводов.

**2.4 Из подмножества признаков, значимо влияющих на переменную класса, выберите пару признаков, наиболее сильно взаимосвязанных друг с другом. Обозначим их X1 и X2. Проверьте наличие частной корреляции между признаком X1 (X2) и откликом при исключении влияния X2 (X1) с помощью: частного коэффициента корреляции**

****

Наиболее сильно взаимосвязанные признаки: A2 и A8 с коэффициентом корреляции: 0.427373545 (Х1 – это признак А2, Х2 – это признак А8, Y – это А16)

Частная корреляция между A2 и A16 с учетом A8: 0.160948

Частная корреляция между A8 и A16 с учетом A2: 0.310746

Вывод:

Частная корреляция между признаком X1 и откликом при исключении влияния X2 равна .

Частная корреляция между признаком X2 и откликом при исключении влияния X1 равна.

Это означает, что после исключения влияния другого признака, взаимосвязь между каждым из признаков и откликом ослабляется, но все еще остается значимой

**2.5 Дайте интерпретацию полученным результатам исходя из практических соображений. Уделите особое внимание интерпретации взаимосвязей между объясняющими переменными и переменной класса**

Ранжирование данных по коэффициенту корреляции Пирсона от наиболее значимого к наименее:

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Сильные корреляции: Переменные с высокими коэффициентами (A9, A11) могут оказывать значительное влияние на переменную класса (A16).

Умеренные и слабые корреляции: Переменные с более низкими коэффициентами могут быть менее значимыми, но все равно могут дать полезную информацию.