|  |  |  |
| --- | --- | --- |
|  | | |
| 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» | | |
|  | | |
|  | | |
|  | Группа ПМ-13 |  |
| Бригада 5 | зиянуров артём |
| Вариант 5 | вострецова екатерина |
|  |  |
|  |  |
| Преподаватель | Тимофеева анастасия юрьевна |
|  |  |
|  |  |
| Новосибирск, 2024 | | |

# Данные

Attribute Information:

A1: b, a.

A2: continuous.

A3: continuous.

A4: u, y, l, t.

A5: g, p, gg.

A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.

A7: v, h, bb, j, n, z, dd, ff, o.

A8: continuous.

A9: t, f.

A10: t, f.

A11: continuous.

A12: t, f.

A13: g, p, s.

A14: continuous.

A15: continuous.

A16: +,- (class attribute)

# Ход работы

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 класса1 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 2 ).**

В данном варианте в качественными данными являются:

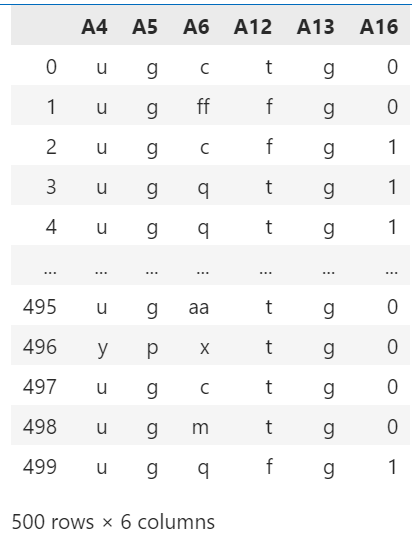
A4, A5, A6, A12, A13

Количественными:

A2, A3, A11, A14, A15

A16 является переменной класса

**Данные ДО калибровки**

****

**Значения после калибровки**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Столбец А4** | **Всего** | **1** | **0** | **с** | **Ответ** |
| u | **385** | **192** | **193** | **1,23214286** | **0,4467223273** |
| y | **114** | **31** | **83** | **0,2361623612** |
| l | **1** | **1** | **0** | **0,6187845298** |

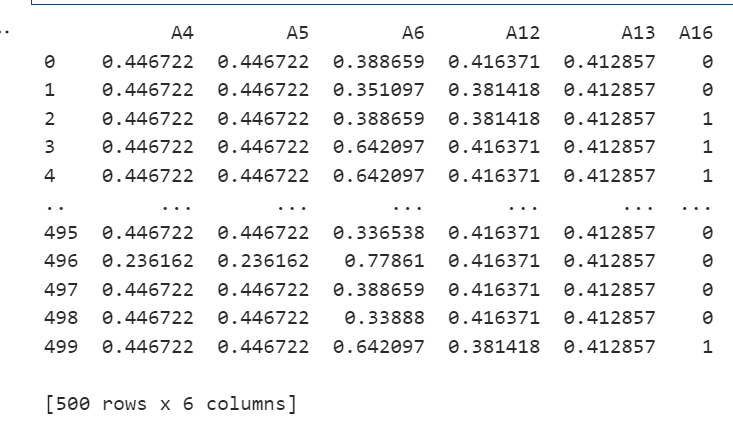
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Столбец А5** | **Всего** | **1** | **0** | **с** | **Ответ** |
| **g** | **385** | **192** | **193** | **1.23214286** | **0,4467223273** |
| **p** | **114** | **31** | **83** | **0,2361623612** |
| **gg** | **1** | **1** | **0** | **0,6187845298** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Столбец А6** | **Всего** | **1** | **0** | **с** | **Ответ** |
| **C** | **105** | **46** | **59** | **1.23214286** | **0,388659184** |
| **D** | **19** | **5** | **14** | **0,245076586** |
| **cc** | **26** | **18** | **8** | **0,631454005** |
| **I** | **43** | **8** | **35** | **0,168674698** |
| **J** | **10** | **3** | **7** | **0,288659793** |
| **K** | **36** | **11** | **25** | **0,272506082** |
| **M** | **29** | **11** | **18** | **0,338880484** |
| **R** | **3** | **2** | **1** | **0,549019607** |
| **Q** | **59** | **41** | **18** | **0,642096642** |
| **W** | **51** | **28** | **23** | **0,495121951** |
| **X** | **30** | **25** | **5** | **0,778609625** |
| **E** | **14** | **9** | **5** | **0,574948665** |
| **Aa** | **37** | **14** | **23** | **0,336538461** |
| **ff** | **38** | **3** | **35** | **0,082717873** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Столбец А12** | **Всего** | **1** | **0** | **с** | **Ответ** |
| **T** | **231** | **108** | **123** | **1.23214286** | **0,416371077** |
| **F** | **269** | **116** | **153** | **0,381418092** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Столбец А13** | **Всего** | **1** | **0** | **с** | **Ответ** |
| **G** | **459** | **213** | **246** | **1.23214286** | **0,412856995** |
| **P** | **1** | **0** | **1** | **0,288659793** |
| **s** | **40** | **11** | **29** | **0,245076586** |

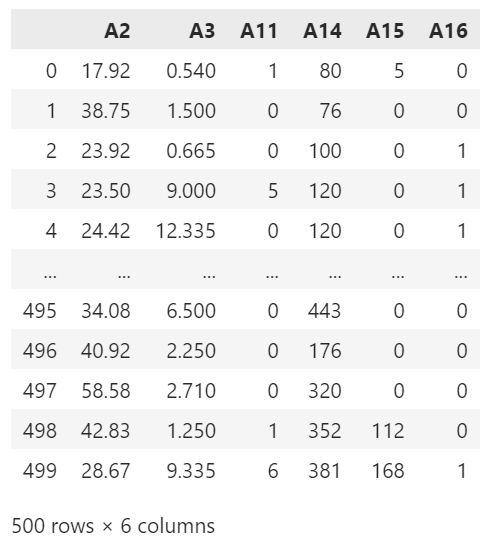
**Данные после калибровки с учетом априорного шанса с поправкой Лапласа:**

****

**К количественным данным относятся столбцы**

A2, A3, A11, A14, A15

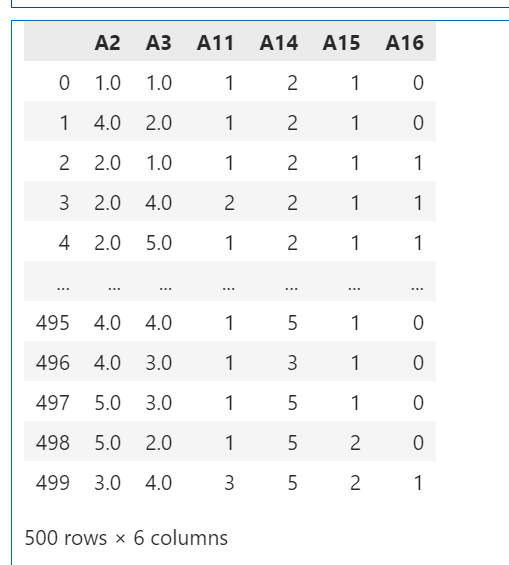
Данные ДО дискретизации:



Получились следующие границы интервалов:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **A2** | **A3** | **A11** | **A14** | **A15** |
| [15.17; 21.814]  (21.814; 25.468]  (25.468;32.282]  (32.282; 40.92]  (40.92; 74.83] | [0.0; 0.822]  (0.822; 2.0]  (2.0;4.415]  (4.415; 9.5]  (9.5; 28.0] | [0.0; 1]  (1; 5.0]  (5.0;67.0] | [0.0; 21.6]  (21.6; 120.0]  (120.0;181.0]  (181.0; 290.0]  (290.0; 980.0] | [0.0; 50]  (50; 587.8]  (587.8;51100.0] |

Данные после дискретизации выглядят так:

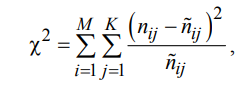
****

1. **Рассчитайте все парные показатели взаимосвязи между переменными из набора данных, соответствующего Вашему варианту.**

**Нашему варианту соответствует m=0**

**m=0 – статистика хи-квадрат, коэффициент сопряженности Пирсона, коэффициент Крамера**

Парные показатели взаимосвязи между переменными в массиве с качественными данными, **статистика хи-квадрат**

****

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **A4** | **A5** | **A6** | **A12** | **A13** |
| **A4** |  | 39,05016 | 103,8378 | 20,18749 | 25,65594 |
| **A5** | 39,05016 |  | 103,8378 | 20,18749 | 25,65594 |
| **A6** | 103,8378 | 103,8378 |  | 84,97514 | 90,44359 |
| **A12** | 20,18749 | 20,18749 | 84,97514 |  | 6,79327 |
| **A13** | 25,65594 | 25,65594 | 90,44359 | 6,79327 |  |

**Значения Хи-квадрат критического**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Альфа=0.01** | | | | |
|  | 0,554 | 5,81 | 0,297 | 0,554 |
| 0,554 |  | 5,81 | 0,297 | 0,554 |
| 5,81 | 5,81 |  | 5,23 | 5,81 |
| 0,297 | 0,297 | 5,23 |  | 0,297 |
| 0,554 | 0,554 | 5,81 | 0,297 |  |

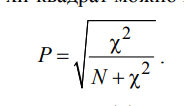
**1 – если хи квадрат больше чем хи квадрат критическое**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **A4** | **A5** | **A6** | **A12** | **A13** |
| **A4** |  | 1 | 1 | 1 | 1 |
| **A5** | 1 |  | 1 | 1 | 1 |
| **A6** | 1 | 1 |  | 1 | 1 |
| **A12** | 1 | 1 | 1 |  | 1 |
| **A13** | 1 | 1 | 1 | 1 |  |

Все значения парно взаимосвязаны друг с другом, гипотеза о том, что признаки не взаимосвязаны отвергается, так как Хи-квадрат больше, чем Хи-квадрат критическое

**Коэффициент сопряженности Пирсона**

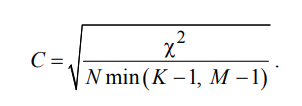
Необходимо проверить коэффициент сопряженности признаков на массиве количественных данных.



Коэффициент Пирсона показал такие результаты:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A2 | A3 | A11 | A14 | A15 |
| A2 |  | 0.31347251813739085 | 0.4676645036558632 | 0.30679243546680357 | 0.4057255353940894 |
| A3 | 0.31347251813739085 |  | 0.488098630668316 | 0.34680389453165544 | 0.4323228032877961 |
| A11 | 0.4676645036558632 | 0.488098630668316 |  | 0.4850623429663959 | 0.5349167467924574 |
| A14 | 0.30679243546680357 | 0.34680389453165544 | 0.4850623429663959 |  | 0.4284106020628657 |
| A15 | 0.4057255353940894 | 0.4323228032877961 | 0.5349167467924574 | 0.4284106020628657 |  |

**Коэффициент Крамера**

****

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Коэффициент Крамера | | | | | |
|  | A2 | A3 | A11 | A14 | A15 |
| A2 |  | 0.3301110135213661 | 0.5290884722675822 | 0.32233665482107593 | 0.4439034703865374 |
| A3 | 0.3301110135213661 |  | 0.5592404610745639 | 0.3697515393581051 | 0.47944298184211476 |
| A11 | 0.5290884722675822 | 0.5592404610745639 |  | 0.5546868766214221 | 0.6331089188093931 |
| A14 | 0.32233665482107593 | 0.3697515393581051 | 0.5546868766214221 |  | 0.4741236237847638 |
| A15 | 0.4439034703865374 | 0.47944298184211476 | 0.6331089188093931 | 0.4741236237847638 |  |

1. **Проверьте гипотезы о значимости взаимосвязей между переменными:**

**m=0 – по критерию хи-квадрат и с помощью перестановочного критерия**

# Текст программы

## Тесты к программе