Статистика для специалистов по бизнес-информатике Одномерный анализ распределений – Введение

Д-р Вольфганг Бёме

Октябрь 2010

2. Одномерный анализ распределений – Содержание

- **2.0 Введение**
- 2.1 Распределение частот / Функция распределения
- 2.2 Числовые характеристики распределения
- 2.3 Анализ концентрации

Собранные данные

Исходная точка: **Первичный список**

$$X_{(n,k)} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \vdots & \vdots & \ddots & \\ x_{n1} & x_{n2} & & x_{nk} \end{pmatrix}$$

где

x_{ij} – значение j-го признака у i-го носителя признака

j – индекс признака $(j = 1, 2, ..., k)$: $X_1, X_2, ..., X_k$

i – индекс носителя признака $(i = 1, 2, ..., n)$

Одномерная статистика

Систематика статистики по числу признаков

$k = 1$ – одномерная статистика

$k > 1$ – многомерная статистика

Первичный список для одномерной статистики

$$X_{(n,1)} = X_{(n)} \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

где

$x_{i1} = x_i$ – значение у i-го носителя признака (наблюдение i)

i – индекс носителя признака $(i = 1, 2, ..., n)$

$X_1 = X$ – признак (1)

Упорядоченные наблюдения

Пусть задан набор данных, измеренный хотя бы по порядковой шкале. Набор данных, упорядоченный по возрастанию, удовлетворяет условию:

$$x_1 \leq x_2 \leq \ ..., \ x_n$$

Обозначение: **Ранговый ряд**

Пример: Первичный список

$$X_{(n,k)} = \begin{pmatrix} 6861 & 73.0 & 6 & 21765.0 & 44.9 & 1.0 & 310.0 \\ 5150 & 72.5 & 7 & 28245.0 & 60.4 & 1.0 & 332.0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 3018 & 74.6 & 7 & 12201.0 & 37.1 & 4.2 & 104.0 \end{pmatrix}$$

где

x_{ij} — значение j-го признака у i-го носителя признака

j — индекс признака $(j = 1, 2, ..., 8)$: СТРАНА, ЭНЕРГИЯ, ..., ГАЗЕТА

i — индекс носителя признака $(i = B, DK, ..., E)$

Одномерная статистика – Первичный список

$$
\begin{array}{ll}
B: & 21765.0 \\
DK: & 28245.0 \\
D: & 25179.0 \\
FIN: & 19048.0 \\
F: & 24608.0 \\
EL: & 7465.0 \\
UK: & 17471.0 \\
IRL: & 14735.0 \\
I: & 17921.0 \\
L: & 27611.0 \\
NL: & 21536.0 \\
A: & 24823.0 \\
P: & 8822.0 \\
S: & 22499.0 \\
E: & 12201.0
\end{array}
$$

где $X_{i\mathbf{1}} = X_i$ — значение у i-го носителя признака (наблюдение i)

i — индекс носителя признака ($i = B, DK, \ldots, E$)

$X_{\mathbf{1}} = X$ — признак: ВВП

Одномерная статистика – Ранговый ряд

$$\begin{array}{ll} EL: & 7465.0 \\ P: & 8822.0 \\ E: & 12201.0 \\ IRL: & 14735.0 \\ UK: & 17471.0 \\ I: & 17921.0 \\ FIN: & 19048.0 \\ NL: & 21536.0 \\ B: & 21765.0 \\ S: & 22499.0 \\ F: & 24608.0 \\ A: & 24823.0 \\ D: & 25179.0 \\ L: & 27611.0 \\ DK: & 28245.0 \end{array}$$

$$x_{\mathbf{1}} \leq x_{\mathbf{2}} \leq \ \ldots, \ x_n$$

Пример из лекции 1

Набор данных: Vorlesungsbeispiel1.xls

Urliste

Merkmalsträger	Anzahl
1	4
2	4
3	3
4	5
5	5
6	5
7	3
8	4
9	5
10	4
11	3
12	4
13	3
14	3
15	3
16	6

Merkmalsträger	Anzahl
27	1
19	2
20	2
25	2
35	2
3	3
7	3
11	3
13	3
14	3
15	3
34	3
38	3
41	3
44	3
49	3

Statische Gesamtheit	5-Raum Wohnungen in einer Wohnungsbaugesellschaft
Merkmal	Anzahl der gemeldeten Personen

Одномерная статистика – Первичный список / Ранговый ряд

Первичный список	
Носитель признака	Кол-во
1	4
2	4
3	3
4	5
5	5
6	5
7	3
8	4
9	5
10	4
⋮	⋮

$\Longrightarrow$

Ранговый ряд	
Носитель признака	Кол-во
27	1
19	2
20	2
25	2
35	2
3	3
7	3
11	3
13	3
14	3
⋮	⋮