

Визначення статистичних характеристик вибірки за допомогою програмного пакету STATISTICA

Л.О. Власенко, В.В. Хомяков

Національний університет харчових технологій

Програма Statistica від Statsoft суттєво полегшує статистичне керування технологічними процесами, оскільки призначена для всіх можливих варіантів роботи з масивами даних, через що отримала широке застосування як в спеціалізованих сферах діяльності (наукові дослідження, стратегічний маркетинг), так і рядовими користувачами (наприклад, підприємцями).

Для початкового аналізу і виведення первинних описових статистик передбачені операції по обчисленню наступних характеристик: дисперсії, моди, медіани, математичного очікування, екстремумів та інших статистик. Для визначення характеристик взаємозв'язку між технологічними змінними доцільно використовувати набір кореляційних методів, проводити обчислення всіх значимих коефіцієнтів (Пірсона, Крамер, Кендалл і т. д.), отримання різних видів кореляцій (приватна, автокореляція), виведення табличних результатів.

Для аналізу вибірки на відповідність нормальному закону розподілу можна використовувати Імовірнісний калькулятор; наприклад, якщо задати z-значень (значення випадкової величини, що має стандартний нормальний розподіл) рівним 3, відповідний імовірнісний рівень, обчислений STATISTICA буде менше ніж 0.001, тому що при нормальному розподілі більше 99.99% спостережень потраплять в діапазон ± 3 стандартних відхилень.

Якщо аналізується менше 60 спостережень, доцільно використовувати критерій Шапіро-Уїлка, якщо більше 60, то критерій Колмогорова-Смірнова.

Одновибірковий критерій перевірки нормальності Колмогорова-Смірнова заснований на максимумі різниці між кумулятивними розподілом вибірки і передбачуваним кумулятивним розподілом:

$$D_n = \sup_n |F_n(x) - F(x)|$$

де $F_n(x)$ – кумулятивний розподіл вибірки;

$F(x)$ – очікуване кумулятивне розподіл (з відомими параметрами).

Критерій Колмогорова-Смірнова та критерій Шапіро-Уїлка в програмі STATISTICA. Для перевірки вибірки на нормальність в програмі STATISTICA можна обрати пункт Distribution Fitting в меню Statistics. Крім того, здійснювати візуальну перевірку закону розподілу на нормальність можна за допомогою гістограми «Plot of observed and expected distribution».

В меню «Статистичні модулі» (Statistics) за допомогою функції «Основні статистики» і таблиці (Basic Statistics/Tables) можна знайти Описові статистики (Descriptive Statistics). У вікні (рис. 1) було проведено завантаження значень досліджуваної вибірки по температурі дифузійного соку в четвертій секції дифузійного апарату за допомогою кнопки «Змінні» (Variables).

В даному вікні є можливість розрахувати у вкладці «Advanced» наступні показники: мінімум і максимум (Minimum & maximum), верхній і нижній квартилі (Lower and upper quartiles), середнє (Mean), вибіркочу дисперсію (Variance), медіану (Median), моду (Mode), розмах (Range), стандартне відхилення (Standard Deviation), вибірковий коефіцієнт асиметрії (Skewness), вибірковий коефіцієнт ексцесу (Kurtosis). Після натиснення кнопки «Summary» з'явиться таблиця з усіма розрахованими вибраними значеннями (табл. I).

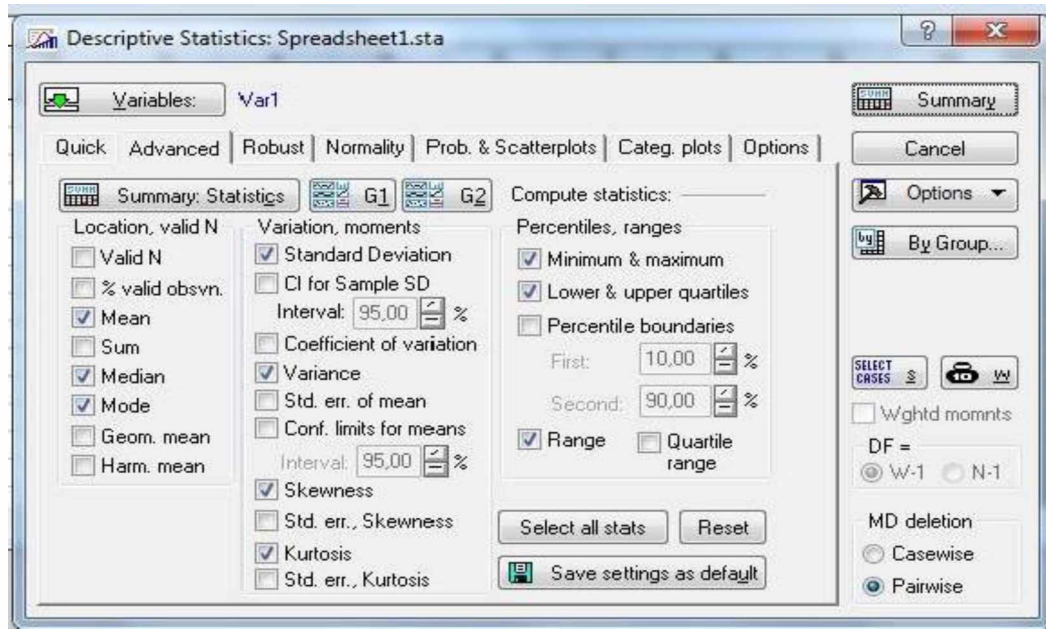


Рис. 1. Вікно знаходження основних статистичних характеристик вибірки по температурі дифузійного соку в четвертій секції дифузійного апарату в пакеті STATISTICA

Таблиця I

Значення основних статистичних характеристик вибірки

| Variable | Mean | Median | Mode | Frequency of Mode | Minimum | Maximum | Lower Quartile | Upper Quartile | Range | Variance | Std.Dev. | Skewness | Kurtosis |
|----------|----------|----------|----------|-------------------|---------|----------|----------------|----------------|----------|----------|----------|----------|----------|
| Var1 | 69.88790 | 70.00000 | 71.00000 | 2329 | 0.00 | 77.00000 | 69.00000 | 71.00000 | 77.00000 | 4.169839 | 2.042018 | -8.46379 | 265.5278 |

Визначені основні статистичні характеристики технологічної змінної є основою для подальшої роботи з ними. На основі них можна зробити аналіз про необхідність проведення відсіювання грубих похибок, проведення згладжування одним з відомих методів тощо.

Література

1. Бочаров П.П. Теория вероятностей. Математическая статистика: учеб. пособие / П.П. Бочаров, А.В. Печинкин. – М., 1998. – 326 с.
2. Статистика. Навчальні матеріали онлайн. [Електронний ресурс]: Режим доступу – <http://pidruchniki.com/statistika/>