

М. А. Д р а н и ц ы н а (Москва, ВМК МГУ). **Алгоритм выполнения недостающих данных для однородных сигналов.**

Возникновение недостающих данных в целом характерно для биологических и клинических исследований. Недостающие данные являются потенциальным источником систематических ошибок в клинических исследованиях, а также оказывают влияние на регуляторные решения по результатам таких исследований (см. [1]). В исследованиях с целью доказательства эквивалентности подходы к обработке недостающих данных зачастую приводят к возникновению смещений и некорректным выводам (см. [2]). В частности, при анализе фармакокинетических кривых, которые отражают зависимость концентрации соединения в биологических матрицах человека или животного после его введения в разных формуляциях от времени, в качестве замены пропущенного значения концентрации используют 0 или $\frac{\text{НПКО}}{2}$ (НПКО — нижний предел количественного определения) даже в тех случаях, когда в следующей временной точке значение концентрации ненулевое.

Предположим, что сигнал однороден в некоторой группе, состоящей из N наблюдений (например, субъектов в исследовании или эпох записи сигнала), где объем выборки N — небольшое число. Пусть поведение сигнала можно описать параболой. Введем обозначение $y_{n,k}$ — измеряемое с некоторой нормально распределенной ошибкой значение сигнала (например, концентрации) для n -го субъекта в k -й временной точке, $n = \overline{1, N}$, $k = \overline{1, K}$. Тогда можно записать $y_{n,k} = C_n x_{n,k}^2$, где нам известен только вид зависимости, а C_n — индивидуальная для каждого n характеристика наблюдений.

Пусть отсутствует некоторое значение $y_{m,l} > 0$, $m = \overline{1, N}$, $l = \overline{2, K-1}$, то есть пропущено значение для некоторого наблюдения из группы в некоторой временной точке измерения сигнала, при этом эта временная точка не является ни первой, ни последней точкой наблюдения. Кроме того, пусть значения $y_{m,l-1}$ и $y_{m,l+1}$ наблюдаются и больше 0. Тогда, для каждого наблюдения из N , за исключением m , определим степень влияния d_n значения в точке $y_{n,k-1}$ на значение $y_{n,k}$ следующим образом:

$$\begin{aligned} d_1 &= \sqrt{\frac{y_{1,k}}{y_{1,k-1}}} = \frac{x_{1,k}}{x_{1,k-1}}, & (1) \\ &\dots \\ d_{l-1} &= \frac{x_{l-1,k}}{x_{l-1,k-1}}, \\ d_{l+1} &= \frac{x_{l+1,k}}{x_{l+1,k-1}}, \\ &\dots \\ d_N &= \frac{x_{N,k}}{x_{N,k-1}}. \end{aligned}$$

Далее вычислим среднюю степень влияния

$$\bar{d} = \frac{1}{N-1} \sum_{n=1, n \neq l}^N \frac{x_{n,k}}{x_{n,k-1}}.$$

Используем полученное усредненное значение для вычисления $y_{m,l} = C_m x_{m,l}^2 = C_m (\bar{d} x_{m,l-1})^2 = \bar{d}^2 y_{m,l-1}$. В дальнейшем планируется обобщение и изучение свойств более сложных моделей, в частности, с использованием сегментирования сигнала для более гибкого определения временных интервалов с преобладанием сходных подлежащих процессов, формирующих сигнал (см. [3]).

Работа выполнена при поддержке РФФИ (проект 18-07-00252).

СПИСОК ЛИТЕРАТУРЫ

1. *EMA Committee for Medicinal Products for Human Use (CHMP) Guideline on Missing Data in Confirmatory Clinical Trials*, EMA/CPMP/EWP/1776/99 Rev. 1.
2. *Rabe B. A., Day S., Fiero M. H., Bell M. L.* Missing data handling in non-inferiority and equivalence trials: A systematic review. — *Pharmaceutical Statistics*, 2018, v. 17, is. 5, p. 77–488.
3. *Dranitsyna M. A., Zakharova T. V.* Investigation of the window variance noise component of multicomponent signals. — *J. of Mathematical Sciences*, 2019, v. 237, is. 5, p. 639–645.