

Методы статистического анализа

Полученные в результате обследования 1350 больных фактические материалы в виде качественных и количественных клинических, лабораторно-инструментальных признаков (519 показателей) регистрировались в тематической карте и составили компьютерную базу данных.

Статистический анализ проводился в центре «Биостатистика» (руководитель – доцент факультета информатики Томского государственного университета, канд. техн. наук В.П. Леонов). Биометрический анализ осуществлялся с использованием пакетов SAS 6.12, SPSS 10, S-PLUS 2000 и STATISTICA-6.

В исследовании применялись методы дисперсионного, дискриминантного и кластерного анализов с построением классификационных дендрограмм, анализ таблиц сопряженности, корреляционный анализ. При анализе таблиц сопряженности оценивались значения статистики Пирсона Хи-квадрат (χ^2), достигнутый уровень значимости (p) и фи-коэффициент (ϕ) - показатель силы связи. Вклады в связь дискретных признаков изучали по величине $\Delta\chi_{ij}^2$ отдельных клеток таблиц сопряженности. Во всех процедурах статистического анализа критический уровень значимости p принимался равным 0,05. Средние выборочные значения количественных признаков приведены в тексте в виде $M \pm SD$, где M - среднее выборочное, SD – стандартное отклонение. Для проверки статистических гипотез применяли непараметрические методы - корреляция по Спирмэну, тест Манна-Уитни, медианный тест Краскэла-Валлиса и непараметрический дисперсионный анализ Фридмана. Проверка нормальности распределения производилась с использованием трех методов: Колмогорова-Смирнова, Лиллиефорса и Шапиро-Уилки, проверка гипотез о равенстве генеральных дисперсий - с помощью критериев Кохрэна и Левене. При построении прогностических вероятностных моделей использовали логистический регрессионный анализ. Построение порядка 100 уравнений логистической регрессии осуществлялось с использованием 80-90 признаков. В каждом из вариантов уравнения логистической регрессии пошаговый алгоритм отбирал предикторы с указанием процента верного предсказания - значения (Concordant) и величины коэффициента связи (Somers'D). Достигнутые уровни значимости теста согласия Hosmer and Lemeshow во всех итоговых уравнениях составляли 0,7-0,9, что свидетельствовало о высокой степени адекватности созданных моделей реальным данным.