

Огородник Е.А.

ОСОБЕННОСТИ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ И РАСЧЕТА РИСКОВ ОБЛУЧЕНИЯ В СРЕДЕ ПРОГРАММИРОВАНИЯ R

Научный руководитель: канд. мед. наук, доц. Аветисов А.Р.

Кафедра радиационной медицины и экологии

Белорусский государственный медицинский университет, г. Минск

Актуальность. Среда программирования R – язык программирования, который был разработан для выполнения статистической обработки и анализа данных, построения разнообразных моделей среди эпидемиологов, гигиенистов, математиков, экономистов, генетиков и многих других ученых в связи с такими преимуществами, как бесплатность системы, простота синтаксиса и функций, доступность для большинства операционных систем большого количества готовых пакетов.

Цель: на основании данных научной литературы и эксперимента оценить эффективность работы среды программирования R, оценить пакеты среды: «stats4», «LinERR», «rERR».

Материалы и методы. Использованы данные о дозах, полученных жителями Лунинецкого района при проживании на загрязненных радионуклидами территориях для последующей статистической обработки данных и оценки рисков облучения при помощи среды программирования R.

Результаты и их обсуждение. Пакет «rERR» в радиационной эпидемиологии используется для анализа взаимосвязи доза-эффект для каждого отдельного случая в зависимости от времени экспозиции. Обычные подходы к анализу данных о когорте и конкретных случаях контроля часто основываются на схемах выборки, составленных на основе набора рисков, где в каждом случае определяется новый набор рисков, включая индекс и все меры контроля, которые в то время подвергались риску. Такого рода схемы связаны с моделью пропорциональных рисков Кокса, $\Phi(z, \beta) = e^{\beta_1 z_1 + \dots + \beta_k z_k}$, где $z_1, \dots, z_k$ – векторы переменных (ковариаты). Эта модель доступна в большинстве стандартных статистических пакетов, но ограничивается логарифмической линейной моделью.

Одной из моделей, представляющих особый интерес, особенно в области медицинской экологии и радиационной эпидемиологии, является линейная модель ERR (Excess Relative Risk) – пакет «LinERR» (от англ. linear excess relative risk) в среде программирования R. Формула линейной зависимости доза-эффект выглядит так: $\Phi(z, \beta) = 1 + \beta * f(\text{dose})$, где z – вектор ковариат. Оценка тренда доза-эффект в рамках линейной модели позволяет сделать вывод о том, что при каждом увеличении воздействия на 1 единицу время возникновения заболевания увеличивается (или уменьшается) аддитивно.

Изменение эффекта воздействия в линейных моделях путем исследования ковариат z можно оценить путем включения логарифмического линейного компонента для линейного воздействия, подразумевающего модель формы $\Phi(z, \beta) = e^{\beta_1 z_1 + \dots + \beta_k z_k} * (1 + \beta * f(\text{dose}))$, где $z_1, \dots, z_k$ – ковариаты.

Этот пакет представляет собой R-имплементацию моделей, которые редко реализуются за пределами программного обеспечения Эпикур: модель линейного избыточного относительного риска, позволяющая учитывать фактор времени, корректировать в зависимости от времени, замедлять воздействие и подвергать данные стратификации.

Выводы: при помощи пакетов «stats4», «LinERR», «rERR» среды программирования R можно быстро и достаточно эффективно производить статистическую обработку данных и оценку рисков облучения жителей Лунинецкого района. При этом результаты оказались простыми для понимания, репрезентативными и не требовали денежных затрат для приобретения программы.