

РЕГРЕССИОННЫЙ АНАЛИЗ В ПРОГНОЗИРОВАНИИ ПРИРОДНО-ОЧАГОВЫХ
ИНФЕКЦИЙ

В.А. ЕВСТЕГНЕЕВА, Т.В. ЧЕСТНОВА, О.Л. СМОЛЯНИНОВА

Медицинский институт, Тульский государственный университет, ул. Болдина, 128, Тула, Россия, 300012

Аннотация. Регрессионный анализ – совокупность статистических методов обработки экспериментальных данных, позволяющих в условиях стохастической зависимости исследуемой величины от неслучайных или случайных переменных определять данную зависимость. Постановка задачи регрессионного анализа формулируется следующим образом. Имеется совокупность результатов наблюдений. Требуется установить количественную взаимосвязь между показателем и факторами.

В данной работе мы попытаемся установить количественную взаимосвязь между показателями заболеваемости природно-очаговыми инфекциями и биотическими и абиотическими факторами природной среды. К биотическим факторам относятся: численность и инфицированность основных носителей и переносчиков инфекций, к абиотическим факторам – метеорологические (средняя месячная температура воздуха, среднее месячное количество осадков, глубина снежного покрова в декабре, январе, феврале, марте).

При изучении влияния 22 факторов на заболеваемость лептоспирозом с помощью множественной регрессии получена математическая модель, которая имеет низкий уровень доверия, а при использовании метода пошаговой регрессии установлено влияние одного фактора – инфицированность обыкновенной полевки из 22 факторов. Уровень доверия модели и коэффициентов модели значимы. Данный метод позволяет определить только линейную связь между заболеваемостью и природными факторами, а в случае нелинейной связи тесноту не устанавливает.

Природный очаг инфекций представляет сложную экологическую систему. Исходя из условий моделирования сложной системы, к которым могут быть отнесены: возможные нелинейные влияния элементов на выходной параметр, синергетика и реципрокность при совместном влиянии отдельных факторов, необходимость учета в отдельных случаях категориальных факторов и нескольких выходных параметров сложной системы, необходимо выбрать искусственные нейронные сети. Они позволяют реализовать указанные условия при получении математической модели системы.

Ключевые слова: природно-очаговые инфекции, лептоспироз, показатель заболеваемости, инфицированность, обыкновенная полевка, регрессионный анализ.

REGRESSION ANALYSIS FORECASTING PRIRIRODNO FOCAL INFECTIONS

V.A. EVSTEGNEEVA, T.V. CHESTNOVA, O.L. SMOLYANINOVA

Medical Institute, Tula State University, str. Boldin, 128, Tula, Russia, 300012

Abstract. Regression analysis - a set of statistical methods for processing of experimental data to a condition of stochastic dependence study of the value of non-random or random variables to define this relationship. Statement of the problem of regression analysis is formulated as follows. There is a set of observational results. Requires a quantitative relationship between the index and factors.

In this paper, we try to establish a quantitative relationship between the incidence of natural - focal infections and biotic and abiotic factors of the environment. By biotic factors include: the number of infection and the major carriers and vectors to abiotic factors - weather (average monthly air temperature, the monthly average rainfall, snow depth in December, January, February, March).

When studying the effect of 22 factors on the incidence of leptospirosis using multiple regression the mathematical model, which has a low level of trust, and when using the stepwise regression established the influence of one factor - infection of the common vole of the 22 factors. Level of trust models and model coefficients are significant. This method allows to determine only the linear relationship between the incidence and natural factors, as in the case of the nonlinear coupling tightness does not install.

Natural foci of infection is a complex ecological system. Based on the terms of modeling complex systems, which may include: the possible impact of non-linear elements in the output parameter, synergy and reciprocity under the joint influence of individual factors, the need to address in some cases categorical factors and multiple output parameters of a complex system, it is necessary to choose an artificial neural network (ANN), allowing to realize these conditions in the preparation of a mathematical model of the system.

Keywords: natural focal infection, leptospirosis, morbidity, infection, common vole, regression analysis.

Библиографическая ссылка:

Евстегнеева В.А., Честнова Т.В., Смольянинова О.Л. Регрессионный анализ в прогнозировании природно-очаговых инфекций // Вестник новых медицинских технологий. Электронное издание. 2015. №4. Публикация 1-8. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2015-4/5024.pdf> (дата обращения: 30.11.2015). DOI: 10.12737/17086

Регрессионный анализ – совокупность статистических методов обработки экспериментальных данных, позволяющих в условиях стохастической зависимости исследуемой величины от неслучайных или случайных переменных определять данную зависимость. Постановка задачи регрессионного анализа формулируется следующим образом. Имеется совокупность результатов наблюдений. Требуется установить количественную взаимосвязь между показателем и факторами. В таком случае задача регрессионного анализа понимается как задача выявления такой функциональной зависимости $y^* = f(x_1, x_2, \dots, x_k)$, которая наилучшим образом описывает имеющиеся экспериментальные данные. В нашем случае требуется установить взаимосвязь между показателями заболеваемости лептоспирозом и ГЛПС и биотическими и абиотическими факторами природной среды.

Основные допущения регрессионного анализа:

- количество наблюдений достаточно для проявления статистических закономерностей относительно факторов и их взаимосвязей;
- обрабатываемые экспериментальные данные содержат некоторые ошибки (помехи), обусловленные погрешностями измерений, воздействием неучтенных случайных факторов;
- матрица результатов наблюдений является единственной информацией об изучаемом объекте, имеющейся в распоряжении перед началом исследования.

В зависимости от количества факторов, включенных в уравнение регрессии, принято различать простую (парную) и множественную регрессии.

Простая регрессия представляет собой регрессию между двумя переменными y и x , т.е. модель вида $y = f(x)$, где y – зависимая переменная (результативный признак); x – независимая, или объясняющая, переменная, (признак – фактор).

Строится простая (парная) регрессия в случае, когда на результативный показатель влияет единственный фактор.

В нашем случае на y -зависимую переменную (показатели заболеваемости, показатели численности) влияют несколько факторов от 13 и более, поэтому будем использовать множественную регрессию).

Множественная регрессия соответственно представляет собой модель вида: $y = f(x_1, x_2, \dots, x_k)$, где x_k – признак – факторы.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели (выбор факторов, вида уравнения и др.)

Факторы, включаемые в модель множественной регрессии, должны отвечать следующим требованиям:

- должны быть количественно измеримы;
- не должны быть интеркоррелированы или находится в функциональной зависимости;
- в одну модель нельзя включать совокупный фактор и, образующие его, частные факторы, что может привести к неоправданному увеличенному их влиянию на зависимый показатель, к искажению реальной действительности;
- количество включаемых в модель факторов не должно превышать одной трети числа наблюдений в выборке.

Отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы, исходя из сущности проблемы; на второй – на основе матрицы показателей корреляции определяют t -статистики для параметров регрессии.

При отборе влияющих факторов используются статистические методы отбора. Так, существенного сокращения числа влияющих факторов можно достичь с помощью пошаговых процедур отбора переменных. Ни одна из этих процедур не гарантирует получения оптимального набора переменных. Однако при практическом применении они позволяют получить достаточно хорошие наборы существенно влияющих факторов.

В данной работе мы попытаемся установить количественную взаимосвязь между показателями заболеваемости природно-очаговых инфекций и биотическими и абиотическими факторами природной среды.

При прогнозировании заболеваемости лептоспирозом мы использовали множественную регрессию и пошаговый регрессионный анализ. В качестве выходных параметров модели (y) был представлен показатель заболеваемость на 100 тыс. населения лептоспирозом – за период с 1985 по 2015 годы. В качестве входных параметров модели (x) были представлены следующие 22 фактора: – инфицированность обыкновенной полевки лептоспирами (X_1), численность обыкновенной полевки (X_2), средняя температура воздуха октября (X_3), средняя температура воздуха ноября (X_4), средняя температура воздуха декабря (X_5), средняя температура воздуха января (X_6), средняя температура февраля (X_7), средняя температура

Библиографическая ссылка:

Евстегнеева В.А., Честнова Т.В., Смольянинова О.Л. Регрессионный анализ в прогнозировании природно-очаговых инфекций // Вестник новых медицинских технологий. Электронное издание. 2015. №4. Публикация 1-8. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2015-4/5024.pdf> (дата обращения: 30.11.2015). DOI: 10.12737/17086

марта (X8), средняя температура апреля (X9), средняя температура мая (X10), количество осадков октября (X11), количество осадков ноября (X12), количество осадков декабря (X13), количество осадков января (X14), среднее количество осадков февраля (X15), среднее количество осадков марта (X16), среднее количество осадков апреля (X17), среднее количество осадков мая (X18), глубина снежного покрова декабря (X19), глубина снежного покрова января (X20), глубина снежного покрова февраля (X21), глубина снежного покрова марта (X22). Применение стандартных пакетов для определения параметров модели выперечисленных факторов с помощью множественной регрессии позволило получить следующие результаты.

Регрессионная модель представлена зависимостью:

$$Y=359,46+3,39X1+1,29X2-26,95X3+3,34 X4+14,99X5+4,12X6+5,50X7-5,62X8-12,52X9-3,52X10-0,07X11-2,03X12-1,34X13-2,55X14+2,71X15-1,81X16+1,91X17+1,46X18-1,37X19+6,65X20+1,19X21-0,89X22$$

Уровень доверия модели 43%, коэффициенты модели также имеют низкий уровень доверия.

При использовании метода пошаговой регрессии, получена регрессионная модель: $Y=13,43+4,18X1$.

Из уравнения видно, что с помощью метода пошаговой регрессии из 22 факторов выделен один основной – инфицированность обыкновенной полевки лептоспирами, который влияет на заболеваемость лептоспирозом в Тульской области. Уровень доверия – 99,9%. Коэффициенты модели значимы. Влияние инфицированности обыкновенной полевки на заболеваемость лептоспирозом отражено на рис.

Component+Residual Plot for Col_1

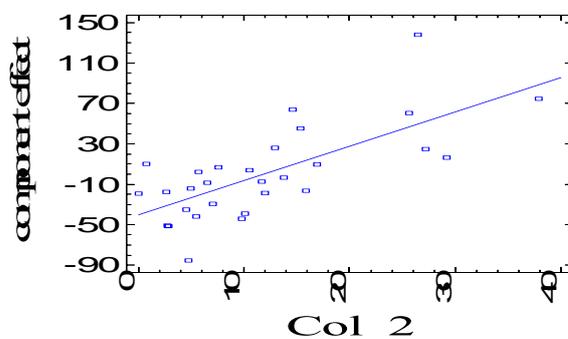


Рис. Влияние инфицированности обыкновенной полевки на заболеваемость лептоспирозом

Таким образом, с помощью метода пошаговой регрессии нам удалось получить модель и установить влияние на заболеваемость лептоспирозом из 22 факторов, одного – инфицированность обыкновенной полевки. Данный метод позволяет определить только линейную связь между заболеваемостью и природными факторами, а в случае нелинейной связи тесноту не устанавливает [1,2].

Природный очаг инфекций представляет сложную экологическую систему. Система «Заболеваемость природно-очаговыми инфекциями – Численность мелких млекопитающих – Метеорологические факторы» является сложной, не поддающейся точному описанию, созданной природой, динамической системой. По продолжительности функционирования она является долгосрочной, по режиму функционирования – непрерывной, предсказуемости поведения – стохастической, т.е. носит вероятностный характер. Это открытая, взаимодействующая с внешней средой система.

К основным элементам системы относятся:

- биотические факторы: численность основных носителей инфекций, их инфицированность, численность иксодовых клещей, их инфицированность;
- абиотические факторы: среднемесячная величина снежного покрова, среднемесячная температура воздуха, среднемесячное количество осадков и др.

Как показал анализ данной системы, она относится к сложной, структура и характер взаимодействия элементов заранее не известны, возможны совместное и нелинейное влияние факторов. Причем состав факторов, характеризующих каждый элемент системы также заранее не известен.

При моделировании сложной системы необходимо учитывать возможные нелинейные влияния элементов на выходной параметр, синергетику и реципрокность при совместном влиянии отдельных факторов, необходимость учета в отдельных случаях не только количественных, но и категориальных факторов, нескольких выходных параметров. В данном случае нами были выбраны искусственные нейронные сети, позволяющие реализовать указанные условия при получении математической модели системы [3-5].

Библиографическая ссылка:

Евстегнеева В.А., Честнова Т.В., Смольянинова О.Л. Регрессионный анализ в прогнозировании природно-очаговых инфекций // Вестник новых медицинских технологий. Электронное издание. 2015. №4. Публикация 1-8. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2015-4/5024.pdf> (дата обращения: 30.11.2015). DOI: 10.12737/17086

Литература

- 1.Новохатка А.Д., Смольянинова О.Л., Честнова Т.В. Влияние инфицированности и численности мелких млекопитающих по ландшафтно-географическим зонам Тульской области на заболеваемость лептоспирозом серогруппы гриппотифоза // Вестник новых медицинских технологий. 2005. №1. С. 122–124.
- 2.Честнова Т.В., Смольянинова О.Л., Логвинов С.И. К вопросу о выборе метода математического анализа с целью прогнозирования заболеваемости лептоспирозом // Вестник новых медицинских технологий. 2011. №4. С.18–21.
- 3.Честнова Т.В., Смольянинова О.Л., Смольянинова В.А. К вопросу прогнозирования численности иксодовых клещей *I. ricinus* в природных биотопах с помощью искусственных нейронных сетей // Вестник новых медицинских технологий. 2012. №1. С. 231–232.
- 4.Евстегнеева В.А., Честнова Т.В., Смольянинова О.Л. О нейросетевом моделировании и прогнозировании эпизоотий туляремии на территории Тульской области // Вестник новых медицинских технологий. Электронное издание. 2014. № 1. Публикация 1-9. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2014-1/5022>. (Дата обращения: 1.12.2014). DOI: 10.12737/7240
5. Евстегнеева В.А. К вопросу о математических методах прогнозирования заболеваемости природно - очаговыми инфекциями // Вестник новых медицинских технологий. Электронное издание. 2014. № 1. Публикация 1-10. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2014-1/5023>. (Дата обращения: 1.12.2014). DOI: 10.12737/7241

References

- 1.Novokhatka AD, Smolyaninova OL, Chestnova TV. Vliyanie infitsirovannosti i chislennosti melkikh mlecopitayushchikh po landshaftno-geograficheskim zonam Tulskey oblasti na zaboлеваemost leptospirosom serogruppygrippotifoza. Vestnic novykh meditsinskikh tekhnologiy. 2005;1:122-4. Russian.
- 2.Chestnova TV, Smolyaninova OL, Logvinov SI. K voprosu o vybore metoda prognosirovaniya zaboлеваemosti leptospirosom [The question of choosing the method of mathematical analysis for solving the problem of medical forecasting leptospirosom morbidity]. Vestnic novykh meditsinskikh tekhnologiy. 2011;4:18-21. Russian.
- 3.Chestnova TV, Smolyaninova OL, Smolyaninova VA. K voprosu prognozirovaniya chislennosti ikso-dovykh kleshchey *I. ricinus* v prirodnykh biotopakh s pomoshchyu iskusstvennykh neyronnykh setey [Forecasting the number of ixododae ix ricinus in natural biotopes by means of artificial neural networks]. Vestnic novykh meditsinskikh tekhnologiy. 2012;1:231-2. Russian.
- 4.Evstegneeva VA, Chestnova TV, Smol'yaninova OL. O neyrosetevom modelirovanii i prognozi-rovanii epizootiy tulyaremii na territorii Tul'skoy oblasti. Vestnik novykh meditsinskikh tekhnolo-giy. Elektronnoe iz-danie [internet]. 2014[cited 2014 Dec 1];1[about 7 p.]. Russian. Available from: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2014-1/5022>. DOI: 10.12737/7240
5. Evstegneeva VA. K voprosu o matematicheskikh metodakh prognozirovaniya zaboлеваemosti pri-rodno - ochagovymi infektsiyami. Vestnik novykh meditsinskikh tekhnologiy. Elektronnoe izdanie[internet]. 2014[cited 2014 Dec 1];1:[about 7 p.]. Russian. Available from: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2014-1/5023>. DOI: 10.12737/7241

Библиографическая ссылка:

Евстегнеева В.А., Честнова Т.В., Смольянинова О.Л. Регрессионный анализ в прогнозировании природно-очаговых инфекций // Вестник новых медицинских технологий. Электронное издание. 2015. №4. Публикация 1-8. URL: <http://www.medtsu.tula.ru/VNMT/Bulletin/E2015-4/5024.pdf> (дата обращения: 30.11.2015). DOI: 10.12737/17086