

**КРИТЕРИИ РАСПРЕДЕЛЕНИЙ ПРИЗНАКОВ В ВЕТЕРИНАРНОЙ МЕДИЦИНЕ****Борисевич М.Н.**

УО «Витебская государственная академия ветеринарной медицины»

В настоящее время в ветеринарной медицине все более актуальными становятся корректные применения статистических методов и научный подход к планированию экспериментов [1-3]. Это связано с развитием концепции доказательной ветеринарной медицины и постепенной интеграцией отечественной науки в науку мировую, а также расширением грантовой системы поддержки научных исследований, а следовательно, и с повышением требований к их методологическому качеству [4,5].

Основная цель применения статистических методов – сведение к минимуму случайных ошибок в научных исследованиях.

Для описания количественных признаков в ветеринарной медицине используются так называемые описательные статистики, т.е. параметры распределения [6]. Если исследование проведено с использованием малой выборки, то в его результатах необходимо отобразить таблицу данных. Однако зачастую необходимость в описательной статистике не требуется [7].

Чтобы выбрать способ описания количественного признака сначала нужно установить, соответствует ли вид распределения его значений нормальному (гауссовому) закону. Нормальное распределение признака характеризуется следующими свойствами [8]: оно симметрично относительно своего среднего значения; среднее значение, мода и медиана распределения совпадают; приблизительно 68% значений признака находятся в интервале  $M \pm s$ , 95% - в интервале  $M \pm 2s$ , 99% - в интервале  $M \pm 3s$ , где  $M$  – среднее значение,  $s$  - среднее квадратическое отклонение).

Реально ни в одной выборке не может быть строго нормального распределения признака. Тем не менее для исследователя важно знать, действительно ли отобранная им выборка принадлежит генеральной совокупности, где признак имеет нормальное распределение [9].

Для решения задачи соответствия можно предложить несколько способов. Рассмотрим их на при-

мере признаков, приведенных в табл.1 [1]. Первый признак описывает поголовье телят, ежегодно получаемых в Республике Беларусь, начиная с 1989 года и по 1998г, признак второй задает количество телят (из общего числа), заболевших колибактериозом, признак третий определяет количество павших от данного заболевания животных. Значения признаков являются объектами статистической ветеринарной отчетности, обобщающих по своему содержанию данные в масштабах республики.

Особый интерес представляют способы визуальной оценки близости распределений признаков к нормальному закону. Они базируются на вычислении стандартного нормального распределения  $\Phi(x)$  [10]:

$$\Phi(x) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(s-M)^2}{2}} dx, \quad (1)$$

(для стандартного нормального распределения  $s=1$ ,  $M=0$ ). Способы визуальной оценки близости признаков нормальному закону реализованы в компьютерной программе, специально разработанной для решения подобных задач кафедрой компьютерного образования ВГАВМ. В частности, полученные с ее помощью гистограммы для перечисленных выше признаков приведены на рис.1. Для сравнения здесь же показана и кривая нормального распределения. Хорошо видно, что распределения признаков неплохо согласуются с нормальным законом, для всех из них можно отметить отчетливо выраженную симметричность распределения относительно среднего значения (что, отметим, является главным свойством нормального распределения). Между тем способы визуальной оценки, хотя и обладают свойствами наглядности и выразительности, являются лишь качественными способами и по сути своей не позволяют осуществлять количественный анализ [11].

Таблица 1 – Рассматриваемые признаки

Годы	Признаки (тыс. голов)		
	получено телят	заболело телят	пало телят
1	2	3	4
1989г.	2000	316	15
1990г.	1700	347	23
1991г.	1570	422	31
1992г.	1400	392	34
1993г.	1400	356	30
1994г.	1270	314	24
1995г.	1200	254	16

1	2	3	4
1996г.	1220	256	15
1997г.	1250	281	15
1998г.	1180	162	6
<b>M</b>	<b>1419</b>	<b>310</b>	<b>20,9</b>
<b>M/2</b>	<b>709,5</b>	<b>155</b>	<b>10,5</b>
<b>s</b>	<b>265,7</b>	<b>75,7</b>	<b>8,8</b>
<b>сравнение</b>	<b>s&lt;M/2</b>	<b>s&lt;M/2</b>	<b>s&lt;M/2</b>
<b>A</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>E</b>	<b>-3</b>	<b>-3</b>	<b>-3</b>

Одним из количественных способов, реализованных в той же программе, является оценка симметричности распределения признаков, имеющих только положительные значения [12]. Если  $s < M/2$  (среднее квадратическое отклонение меньше половины среднего арифметического), то распределение можно считать симметричным. Расчетные данные способа, полученные с использованием той же программы, приведены в табл.1. Анализ данных показывает, что для всех трех признаков имеет место симметричное распределение, хотя для первого признака сделанное замечание справедливо в значительно меньшей степени, чем для остальных двух (для него соотношение между  $s$  и  $M$  наибольшее). Описанный подход не позволяет оценить степень крутизны графика или его плосковершинность (так называемый эксцесс). Решить эти задачи позволяют следующие методы [13].

Метод оценки показателя асимметрии  $A$  оценивает количественную меру симметричности распределения признака. Так, если распределение симметрично, то показатель асимметрии  $A=0$ , если склон кривой распределения расположен справа от оси симметрии, то  $A>0$ , если слева, то  $A<0$ .

Показатель эксцесса характеризует вершину кривой распределения. Для нормального распределения показатель эксцесса  $E=0$ . Если полученная кривая имеет более острую вершину, чем у нормального распределения, то  $E>0$ , если вершина более плоская, то  $E<0$ .

Вычислить показатель асимметрии можно по формуле:

$$A = \frac{\frac{1}{n} \sum (x_i - M)^3}{\sqrt{\left\{ \frac{1}{n} \sum (x_i - M)^2 \right\}^3}} \quad (2)$$

где  $n$  – количество данных в изучаемой выборке,  $x_i$  – отдельные показатели выборки.

Для удобства программирования эта формула может быть преобразована к следующему виду:

$$A = \frac{\sum x_i^3 + 3M(M \sum x_i - \sum x_i^2) - nM^3}{\sqrt{\frac{1}{n} (\sum x_i^2 - 2M \sum x_i + nM^2)^3}} \quad (3)$$

Показатель эксцесса находят по формуле:

$$E = \frac{\frac{1}{n} \sum (x_i - M)^4}{\left\{ \frac{1}{n} \sum (x_i - M)^2 \right\}^2} - 3 \quad (4)$$

Ее также можно преобразовать к более удобному виду:

$$E = \frac{\sum x_i^4 - 4M(\sum x_i^3 + M \sum x_i) + 6M^2 \sum x_i^2 + nM^4}{(\sum x_i^2 - 2M \sum x_i + nM^2)^2} - 3 \quad (5)$$

Рассчитанные значения  $A$  и  $E$  приведены в табл.1. В целом они подтверждают уже сформулированные положения.

Для получения более надёжной оценки соответствия изучаемого распределения признака закону нормального распределения следует пользоваться дополнительным способом – проверкой статистических гипотез о виде распределения, т.е. о том, включена ли выборка из генеральной совокупности, в которой изучаемый признак имеет нормальное (гауссово) распределение [14]. В конечном счете от вида распределения зависит выбор способа описания центральной тенденции и рассеяния значений признака, а также выбор методов дальнейшего анализа данных [15]. По результатам анализа принимается одна из следующих гипотез: нулевая гипотеза (распределение исследуемого признака в генеральной совокупности соответствует закону нормального распределения) и альтернативная гипотеза (распределение исследуемого признака в генеральной совокупности не соответствует закону нормального распределения).

В этой связи для принятия решения о виде распределения можно применять следующие критерии [16-18]: Колмогорова-Смирнова (в тех случаях, когда среднее значение и среднее квадратическое отклонение признака известны априори, а не вычисляются по выборке); Лиллиефорса (когда среднее значение и среднее квадратическое отклонение признака априори неизвестны и вычисляются по выборке); Шапиро-Уилка (когда также исходно неизвестны значения среднего и среднего квадратического отклонения; способ предпочтителен для применения, так как является наиболее мощным и универсальным критерием и при этом наиболее строгим из числа перечисленных). Названные критерии входят в состав разработанной программы в виде

# ОРГВЕТДЕЛО, ЭКОНОМИЧЕСКАЯ ТЕОРИЯ, ФИЛОСОФИЯ, ИСТОРИЯ

отдельного вычислительного блока, предназначенного для проверки статистических гипотез о виде распределения. Рассчитанные значения приведены на рис.1. Критерии нормальности проверяют нулевую гипотезу. Если полученное значение  $p$  для данного статистического критерия больше критического уровня значимости, принятого исследовате-

лем (например, 0,05), то нулевую гипотезу не отклоняют, т.е. распределение исследуемого признака приближенно можно считать нормальным. Из этого, в частности, следует, что для описания распределения признака можно использовать такие параметры, как среднее значение  $M$  и среднее квадратическое отклонение  $s$ .

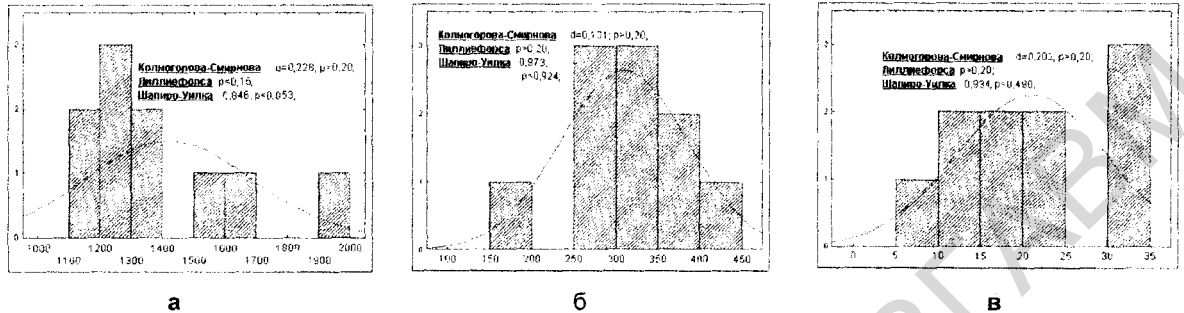


Рис. 1. Вид распределений: сплошная кривая – нормальное (гауссово) распределение; гистограммы - получено телят (а); количество заболевших телят (б); количество павших телят (в).

Если при использовании критериев нормальности  $p < 0,05$ , то следует отклонить нулевую гипотезу и принять альтернативную, т.е. распределение признака считать отличающимся от нормального. В этом случае для описания распределения исследуемого признака можно использовать исключительно непараметрические методы. Иногда критерии Лиллиефорса и Шапиро-Уилка дают разные результаты. В таком случае предпочтительнее опираться на результаты теста Шапиро-Уилка как на более надежные.

Проанализируем теперь данные, приведенные на рис.1, применительно к описанным критериям. Отсюда, в частности, следует, что критерий Колмогорова-Смирнова для всех трех признаков значительно больше общепринятого порогового уровня ( $p > 0,20$ ). Следовательно, согласно описанной выше методике все три признака имеют приближенно нормальное распределение.

Критерий Лиллиефорса также свидетельствует в пользу нормальности распределений (хотя для первого признака он и меньше 0,15, но все же больше 0,05; для второго  $p > 0,20$ ; для третьего -  $p > 0,20$ ).

Тот же вывод можно получить, анализируя и критерий Шапиро-Уилка. Примечательно, однако, что для первого признака здесь  $p < 0,053$ . Как видно, приведенное значение лишь немногим выше порогового уровня 0,05. Это значит, что для первого признака заключение о нормальности распределения является почти критическим. Точнее говоря, распределение признака условно можно признать нормальным и то лишь с малой долей вероятности, меньшей значения 0,053.

Для второго признака критерий Шапиро-Уилка дает значение  $p = 0,924$ , близкое к единице. Поэтому распределение второго признака можно с большей долей уверенности признать нормальным (с точностью 0,924). В свою очередь это означает, что для статистического описания второго признака можно использовать только два значения: среднее арифметическое  $M$  и среднее квадратическое отклоне-

ние  $s$  - в виде  $M(s)$  (широко распространенное представление  $M \pm s$  в настоящее время использовать не рекомендуется).

Для третьего признака критерий Шапиро-Уилка  $p = 0,490$  (что также значительно больше 0,05). Поэтому распределение третьего признака также можно считать нормальным. Для его описания можно использовать запись  $M(s)$ .

Дополнительно отметим следующее. Описательные статистики (параметры каждого распределения) необходимы для того, чтобы знать основные характеристики выборки – центральную тенденцию и рассеяние для заданного признака. И те, и другие показывают, в какой степени близка изучаемая выборка к исследуемому контингенту, иначе говоря, насколько обобщаемы результаты исследования [19, 20].

Меры центральной тенденции (меры локализации, меры положения) показывают наиболее типичное значение для данной выборки [21]. К мерам центральной тенденции относятся: среднее значение ( $M$ ) – среднее арифметическое; медиана ( $Me$ ) – значение, справа и слева от которого на оси значений признака располагаются равные количества значений признака данной выборки; мода ( $Mo$ ) – наиболее часто встречающееся значение признака в выборке; среднее геометрическое значение – антилогарифм среднего арифметического для логарифмированных данных.

Меры рассеяния показывают разброс значений признака в выборке [22]. К мерам рассеяния относятся: размах – разность максимального и минимального значений признака; интерпроцентильный размах (интервал) – значения каких-либо процентилей распределения, например, 10-го и 90-го; интерквартильный размах (интервал) – значения 25-го и 75-го процентилей (такой интервал независимо от вида распределения включает 50% значений признака в выборке); среднее квадратическое отклонение  $s$  (показывает разброс данных по интервалу значений признака относительно среднего значения).

Центральные тенденции и рассеяния количественных признаков, имеющих приближенно нормальное распределение (и только таких признаков) следует описывать средним значением  $M$  и средним квадратическим отклонением  $s$  в формате  $M(s)$ .

Центральные тенденции и дисперсии количественных признаков, не имеющих приближенно нормального распределения (а это подавляющее большинство ветеринарно-биологических признаков – около 80%) следует описывать медианой и интерквартильным размахом (25-й и 75-й процентиля) или другим интерквартильным размахом (например, 80% интерпроцентильным размахом – между 10-й и 90-й процентилями).

Обобщая приведенные результаты, отметим следующее. Если признак определен как количественный, то задача исследователя (прежде всего) заключается в том, чтобы установить, является ли его распределение нормальным (гауссовым) (при этом следует иметь в виду, что лишь около 20% распределений количественных признаков, встречающихся в ветеринарно-медицинских исследованиях, являются нормальными либо приближенно нормальными [23]). Статистическая гипотеза о близости распределения нормальному закону должна быть проверена в соответствии с описанной выше методикой. Необходимо это для того, чтобы оценить вариабельность значений признака, обусловленную влиянием множества причин, вносящих свой вклад в формирующееся распределение. В свою очередь для конкретного признака это означает установившуюся взаимосвязь с многими подсистемами, а не только с одной-двумя из них.

**Литература:** 1. Ломако Ю.В. Эпизоотический мониторинг колибактериоза новорожденных телят в Республике Беларусь / Ю.В. Ломако, Н.Н. Андросик // Ветеринарная медицина Беларуси. – 2002. – №2. – С.15-17. 2. Эпизоотологическое прогнозирование особо опасных болезней / В.М. Гуленкин [и др.] // Ветеринария. – 2001. – №12. – С.3-5. 3. Маринин Е.А. Прогнозирование сроков оздоровления хозяйств от лейкоза крупного рогатого скота / Е.А. Маринин // Ветеринария. – 1995. – №2. – С.15-17. 4. Планирование эксперимента в биологии и сельском хозяйстве / Под ред. В.Н. Максимова. – Москва: МГУ, 1991. – 317 с. 5. Власов В.В. Введение в доказательную медицину /

В.В. Власов. – Москва: МедиаСфера, 2001. – 492 с. 6. Лисенков А.Н. Математические методы планирования многофакторных медико-биологических экспериментов / А.Н. Лисенков. – Москва: Медицина, 1979. – 212 с. 7. Адлер, Ю.П. Планирование эксперимента при поиске оптимальных условий / Ю.П. Адлер Е.В. Маркова. – Москва: Наука, 1976. – 269 с. 8. Бащинский С.Е. Как следует представлять данные рандомизированных контролируемых исследований / С.Е. Бащинский // Международный журнал медицинской практики. – 1997. – Т.1. – С.7-10. 9. Новая форма представления результатов исследований, посвященных лечению // Международный журнал медицинской практики. – 1998. – Т.1. – С.7-8. 10. Поляков И.В. Практическое пособие по медицинской статистике / И.В. Поляков, Н.С. Соколова. – Ленинград: Медицина, 1975. – 315 с. 11. Айвазян С.А. Прикладная статистика: исследование зависимостей: справ. изд. / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин; под ред. С.А. Айвазяна. – Москва: Финансы и статистика, 1985. – с. 34-56. 12. Холлендер М. Непараметрические методы статистики / М. Холлендер, Д.А. Вулф; пер. с англ. – Москва: Финансы и статистика, 1983. – 358 с. 13. Афифи, А. Статистический анализ. Подход с помощью ЭВМ / А. Афифи, С. Эйзен. – Москва: Мир, 1982. – 217 с. 14. Гланц, С. Медико-биологическая статистика / С. Гланц; пер. с англ. – Москва: Практика, 1998. – 314 с. 15. Altman D.G. Practical Statistics for Medical Research / D.G. Altman. – London: Chapman&Hall, 1991. 16. Lang, T.A. How to report statistic in medicine / T.A. Lang, M. Sedic // Annotated guidelines for authors, editors, and reviewers. Copyright by American College of Physicians, Philadelphia, Pennsylvania, printing/binding by Port City Press, composition by Techsetters, 1997. 17. Власов В.В. Эффективность диагностических исследований / В.В. Власов. – Москва: Медицина, 1998. 18. Clarcе M. Cochrane Reviewers' Handbook 4.1.5. / M. Clarcе, A.D. Oxsmан // [updated April 2002]. In: The Cochrane Library, Issue 2, 2002. Oxford: Update Software. Updated quarterly. 19. Lau, J. Quantitative synthesis in systematic review / J. Lau J.P. Joannidis, C.H. Sebmид // Ann Intern Med. – 1997. – 127:820. – 6. 20. Mantel N. Statistical aspects of the analysis of data from retrospective studies of disease / N. Mantel, W. Haenszel // J Natl Cancer Inst. – 1959. – 22: 719. – 48. 21. Beta blockade during and after myocardial infarction: an overview of the randomized trials / S. Yusuf [and a. a.] // Prog Cardiovasc Dis. – 1985. – 27: 335. – 71. 22. DerSimonian R. Meta-analysis in clinical trials / R. DerSimonian, N. Laird // Control Chn Trials. – 1986. – 7: 177. – 88. 23. Cina C.S. Carotid endarterectomy for symptomatic carotid stenosis (Cochrane Review) / C.S. Cina C.M. Clase R.B. Haynes // In: The Cochrane Library, Issue 2. – Oxford: Update Software. 2002.

УДК 35357977978

**НЕЙРОСЕТЕВЫЕ КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ДИАГНОСТИКЕ ЗАБОЛЕВАНИЙ ЖИВОТНЫХ****Борисевич М.Н.**

УО «Витебская ордена «Знак Почета» государственная академия ветеринарной медицины»

**Введение.** Программные средства, базирующиеся на технологиях экспертных систем (ЭС), получили в последние годы распространение в самых разных областях науки и техники [1], в том числе и в ветеринарной медицине. Круг программируемых здесь вопросов чрезвычайно широк - от принятия решений и диагностики заболеваний животных до назначения лечебно-профилактических мероприятий и краткосрочно-долгосрочного прогнозирования.

Опыт создания ветеринарных ЭС показывает, что с их помощью можно уверенно проводить диагностику явно выраженных патологических процессов у животных (корректно диагностируются и самые ранние формы заболеваний и даже только предрасположенность к ним). Такую раннюю и точную диагностику невозможно провести никакими другими аппаратными способами (с помощью УЗИ, рентгена, ЯРМ, компьютерной томографии), кото-