

# **МЕТОДИКА РЕШЕНИЯ ПРОБЛЕМЫ МАТЕМАТИЧЕСКОЙ ВОСПРОИЗВОДИМОСТИ СТАТИСТИЧЕСКИХ ДАННЫХ НАУЧНЫХ ИССЛЕДОВАНИЙ В СЕЛЬСКОХОЗЯЙСТВЕННЫХ ОТРАСЛЯХ НАУКИ**

С. В. Соляник, аспирант  
РУП «Научно-практический центр Национальной академии наук  
Беларуси по животноводству», г. Жодино, Республика Беларусь

Согласно «Номенклатуре специальностей научных работников» к сельскохозяйственной отрасли науки относятся такие специальности как агрономия, ветеринария, зоотехния, лесное и рыбное хозяйства. При этом никого из чиновников аттестационных органов стран СНГ не смущает, что лесное хозяйство базируется в первую очередь на

агрономии, а рыбное хозяйство – это одно из направлений частной зоотехнии, ну а ветеринария относится к самостоятельным отраслям науки – ветеринарным.

Однако проблема не в формальном делении групп специальностей и отнесении конкретных специальностей к сельскохозяйственной отрасли науки, а в сложности разработки компьютерных программ для развития этих специальностей. К слову, эта же проблема существует и в биологических, и медицинских отраслях науки. Основным путем для разработки компьютерных программ в биологии и сельском хозяйстве стало создание баз данных с функциями выбора, запроса, сортировки и др. Для хранения огромных массивов информации требуются значительные компьютерно-технические ресурсы.

По общему правилу исследователи в проводимых ими экспериментах выявляют те или иные зависимости. Но прежде, чем приступить к установлению влияния одного фактора на другой, или группы фактором между собой, полученные данные подвергаются статистической обработке. В большинстве случаев исследователи в области агрономии, зоотехнии, ветеринарии, биологии ограничиваются определением «стандартного» набора статистических параметров: средняя арифметическая, среднеквадратическое отклонение, ошибка средней арифметической, коэффициент вариации, коэффициент корреляции и пр.

При этом два последних коэффициента (вариации и корреляции), чаще всего встречаются у исследователей, занимающихся селекционными вопросами в растениеводстве и животноводстве. Остальные ученые сельскохозяйственной отрасли науки указывают среднее арифметическое и его ошибку, лишь иногда приводится значения среднеквадратического отклонения.

После расчета «базовых» статистических показателей подопытных групп происходит сравнение их между собой с установлением уровня достоверности различий. В итоге в научных публикациях приводятся таблицы, содержащие статистическое выражение  $M \pm m^*$  для конкретного исследованного параметра, а в текстовом описании к ним указывается, на сколько процентов отличается один анализируемый параметр от другого. Таким образом, исследователь формулирует механизм влияния технологического фактора на биологический объект (растение, животное).

Проводя многочисленные исследования, ученые группируют полученные статистические данные и заносят их в базы данных, к

которым обращаются для подтверждения тех или иных установленных тенденций развития биологических или технологических процессов.

На наш взгляд, необходимо не столько создавать базы данных, сколько пытаться заменить средние значения параметров на математические формулы, описывающие или изменения параметра во времени, или зависимость одного параметра от другого. Для этой цели используются программные продукты, позволяющие строить аппроксимационные кривые от одной или двух переменных (CurveExpert, TableCurve 2D и 3D, Mathematica и др.). По сути, доказательством установленной исследователем зависимости  $y = f(x)$  или  $y = f(x, z)$  может быть лишь вариант, когда созданная аппроксимационная функция позволяет воспроизводить исходные данные с минимальной погрешностью (не более  $\pm 5\%$ ).

Анализ полученных и статистически обработанных исследователями первичных данных показал, что даже при наличии высокого уровня корреляции между ними не всегда имеется возможность спроектировать функцию аппроксимации надлежащего качества, т. е. способную адекватно воспроизвести исходные параметры. При этом, чем больший объем выборки конкретного параметра, например, количества проб крови от животных различного возраста, тем сложнее определить аппроксимационную модель взаимосвязи показателей гематологического профиля.

Для решения этой проблемы при исследовании естественной резистентности свиней, а также морфологических и биохимических параметров крови (в общей сложности 50 показателей), мы поступали следующим образом.

Вначале пробы крови, отбираемые от животных в конкретном возрасте мы статистически обрабатывали, группируя тем самым базу данных гематологических параметров для конкретного возраста. К слову, нами было взято примерно 360 образцов крови свиней от рождения до достижения ими возраста полутора лет. Показатели образцов крови группировались по мере взросления животных – у нас получилось 19 временных периодов, примерно с одинаковым количеством проб, отнесенных к каждому периоду.

Затем усредненные групповые параметры были подвержены статистической обработке, что позволило нам установить так называемые зоотехнические (зооигиенические) тенденции формирования гематологического профиля свиней. При этом из полученных группировок нами с помощью программного продукта CurveExpert 1.3 по средним значениям в группах были определены

модели аппроксимации, максимально удачно описывающие взаимосвязь между гематологическими параметрами.

Далее мы провели поиск аппроксимационных кривых для гематологических показателей всех 360 проб крови. Полученные математические модели гематологических показателей мы назвали биологическими тенденциями их формирования. По сути, аппроксимационные кривые являются стохастическими зависимостями, описывающими биохимические взаимосвязи между морфологическими, биохимическими и иммунологическими показателями крови свиней без привязки к возрасту, полу, физиологическому состоянию и др.

Необходимо отметить, что для зоотехнических тенденций формирования гематологического профиля свиней характерны, в основном, прямые и криволинейные зависимости, и лишь треть нелинейные, в то время, как для биологических тенденций нелинейных, больше половины. К нелинейным функциям относятся: Sinussidal Fit, MMF Model, Weibull Model, Vapor Pressure, Modified Exponential, Exponential Association и др.

Для того чтобы в процессе исследований биохимических и технологических параметров в сельскохозяйственных отраслях науки, а также в ходе статистической обработки определить, возможно ли подобрать модель аппроксимации, наиболее надежно описывающую зависимости между показателями, необходимо обязательно определить коэффициент вариации. Нами установлено, что ошибка аппроксимационной кривой (процент по модулю от фактических значений) составляет  $0,46 \cdot C_v$ . Например, если коэффициент вариации составляет 5 %, то ошибка модели аппроксимации составит  $\pm 2,3$  %, 10,9 % –  $\pm 5$  %, 50 % –  $\pm 23$  % и т. д. При значениях коэффициента вариации выше 70 %, ошибка аппроксимационной кривой в отдельных случаях больше, чем в предлагаемой нами формуле. Таким образом, чем выше коэффициент вариации для конкретного параметра, тем аппроксимационная модель будет менее адекватно описывать исходные данные.

В базах данных каждое значение (число, функция, логическое выражение и др.) занимает отдельную ячейку, при этом ее максимальный объем, например в MS Excel, составляет несколько сотен знаков. В наших исследованиях, описывающих зоотехнические и биологические тенденции, массив биохимических показателей крови свиней занимал диапазон почти в 20 тысяч ячеек, а замена формулами позволила его уменьшить в 4 раза. При этом появилась возможность определять промежуточные значения, а главное моделировать течение

обменных процессов в организме свиней, изменяя значения морфологических, биохимических и иммунологических параметров крови.

Таким образом, исследователи, разработав зависимости (формулы), которые надлежащим образом описывают исходные параметры не статистически, а математически, доказывают наличие взаимосвязи между показателями. Это позволяет гарантировать многократное воспроизведение полученных результатов независимыми экспериментаторами. Следовательно, выявленные математические зависимости являются доказательством того, что исследователем получены новые знания, являющиеся достоверными с научной и практической точки зрения.

В случае, когда невозможно использовать программные продукты для подбора надлежащей модели аппроксимационной кривой, с минимальным процентом ошибки, можно поступить следующим образом. Данные одного из исследуемых показателей отсортировать по возрастанию, а затем разбить на несколько участков и для каждого написать формулу взаимосвязи. Затем полученные формулы, через функцию «ЕСЛИ», превратить в единую модель. В табличном процессоре MS Excel версии 2003 и ранее, количество функций ЕСЛИ в одной модели не должно превышать 7, а в более поздних – 64.

Если известен механизм формирования того или иного показателя в граничных диапазонах, то отпадет необходимость проведения научных исследований, чтобы установить промежуточные значения. Ведь в последние четверть века в зоотехнической науке представление в диссертационной работе морфологических, биохимических показателей крови стало основой «биологического раскрытия механизма» влияния того или иного технологического фактора на организм сельскохозяйственных животных. Однако проведенные исследования и защищенные диссертации не стали доказательством того, что полученные в эксперименте результаты с той же эффективностью воспроизводятся в реальных производственных условиях.

Создание аппроксимационных моделей на основе статистического анализа первичных данных, позволяет исследователям в области агрономии, зоотехнии и зоогигиены, математически обосновать реальный механизм формирования того или иного технологического или биологического показателя. Создание математической модели дает возможность направить материально-финансовые усилия не на постановку и проведение очередного

эксперимента в рамках образовательной дисциплины или области исследований согласно паспорту специальности, а на реальное решение научно-производственных проблем.

В итоге сельскохозяйственная наука, которая является прикладной по своей сути, получит специалиста не столько по предмету, сколько специалиста по проблеме, т. е. специалиста (ученого), владеющего всеми вопросами технологии производства продуктов растительного и животного происхождения.