

Calf Notes.com

Заметка о телятах №127. Что такое «уровень значимости P менее 0,05»?

Введение

Давайте рассмотрим такую ситуацию: вы только что закончили утренние дела на своей ферме по выращиванию телят и устроились в кресле перед компьютером, чтобы внести кое-какие данные и распечатать несколько отчетов. Звонит телефон. «Доброе утро! Это Сара Беллим из комбикормовой компании Cheatham Feed. Я хотела бы рассказать вам о нашей новейшей разработке — добавке к заменителю цельного молока для телят Calf Explosion. В ходе нашего последнего исследования было продемонстрировано *статистически значимое* улучшение показателей роста телят, получавших добавку Calf Explosion. Могу ли я заехать к вам завтра, чтобы рассказать больше об этом чудесном новом продукте?» И вы думаете: «Хmmm, статистически значимое улучшение? Звучит многообещающе. Может быть, стоит пообщаться с Сарой по поводу этого нового продукта...»

Но что на самом деле означает понятие «статистической значимости»? И что оно означает для вас как для фермера, выращивающего телят (или консультанта)? В данной заметке о телятах мы рассмотрим некоторые из этих вопросов.

Сегодня новые технологии появляются каждый день. Усовершенствования в области кормления, здоровья, размножения, ухода и информации возникают с неуклонно растущей скоростью. Когда животновод внедряет эти новые технологии в своем хозяйстве, он ожидает увидеть изменения, которые улучшат показатели животных, снизят производственные издержки, либо тем или иным образом улучшат экономику предприятия.

Но все мы также знаем о новых технологиях (продуктах и программах), которые не работают. Отсутствие эффекта возможно по многим причинам, но это не меняет того факта, что соответствующий продукт не работает на вашем предприятии (или, возможно, даже не действует вовсе). Как же отделить действительно эффективные продукты от тех, которые являются пустой тратой времени и денег? Первый шаг заключается в понимании научной основы продукта. И здесь как раз возникает термин «статистическая значимость».

Вариация — пример

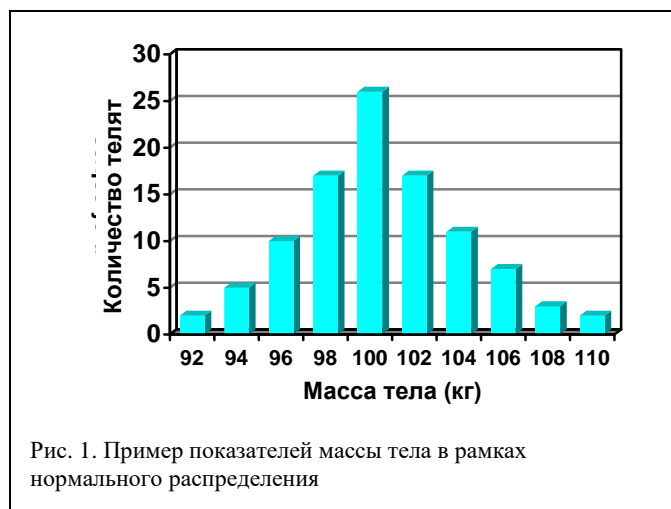
Давайте проясним значение термина «статистическая значимость», используя простой пример. Предположим, что у вас есть 2 группы по 100 телят. Первой группе вы скармливаете обычный рацион. Назовем ее «контрольная группа». Второй группе вы скармливаете обычный рацион плюс кормовую добавку X. Вы измеряете массу тела телят в начале, и затем еще раз в конце исследования.

Давайте сперва рассмотрим контрольную группу. Все телята в составе этой группы получали один и тот же рацион, одинаковый уход, и, теоретически, все они должны были иметь одинаковый вес в конце исследования. Конечно же, не все телята весили одинаково. Некоторые телята росли быстрее остальных в силу своих стандартных генетических характеристик. Другие телята в некоторые из дней потребляли чуть больше заменителя молока, чем остальные. У некоторых

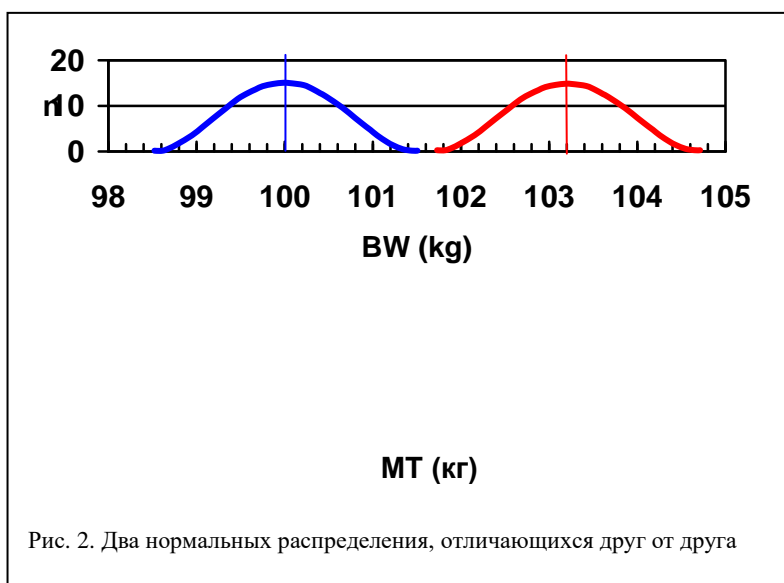
телят в течение срока исследования возникала диарея, которая замедлила их рост. Рассматривая средний показатель конечной массы тела всех телят в составе контрольной группы, давайте примем его как равный, скажем, 100 килограммам (220 фунтам). Однако в рамках этого среднего значения также присутствуют вариации — т. е. у некоторых телят вес в конце исследования превышал средний показатель (также именуемый «средним значением»), а у некоторых был ниже среднего значения. Допустим, что в рамках данного примера все телята весили от 92 до 110 килограмм в конце исследования. Это колебание около среднего значения можно рассчитать многими различными способами, но общепринятым методом является расчет *стандартного отклонения* от среднего значения. Допустим, что стандартное отклонение от среднего значения в контрольной группе составило 4 килограмма. Если мы построим график количества телят в каждой из категорий с разницей в 2 килограмма (скажем, 92–93,9 килограмма, 94–95,9 килограмма), то наш график будет аналогичен приведенному на рис. 1. Существует нормальная случайная вариация массы тела, и распределение около среднего значения (100 килограмм) именуется «нормальным распределением». Такая кривая распределения также называется «колоколообразной кривой», так как по форме она напоминает колокол.

Теперь давайте рассмотрим те же параметры у телят, получавших кормовую добавку. Средняя конечная масса тела у этой группы телят составила 103,2 килограмма при стандартном отклонении в 5 килограмм.

Мы можем использовать два значения (среднее значение и стандартное отклонение) для каждой группы, чтобы определить, насколько существенным (то есть, статистически значимым) является различие между двумя группами. Это можно сделать с помощью серии определенных статистических процедур. В основном эти процедуры направлены на сравнение средних значений по каждой группе наряду с отклонениями от этих средних значений, чтобы определить, являются ли эти два средних значения идентичными (т. е. кормовая добавка X не оказала никакого воздействия), или же эти средние значения различаются.



Взгляните на рис. 2. Здесь сразу же заметно, что различаются не только средние значения (прямые линии), но и колебания около среднего значения (колоколообразные линии). Это связано с тем, что две колоколообразные кривые не накладываются друг на друга. На рис. 3 мы видим, что хотя средние значения почти одинаковы (100 и 103 килограмма), колебания около среднего значения являются более существенными (т. е. колоколообразные кривые значительно шире) и происходит их наложение друг на друга. В этом



случае из статистического исследования можно заключить, что средние значения НЕ отличаются друг от друга, и что кормовая добавка «Х» не оказала никакого воздействия в рамках данного эксперимента.

Два типа ошибок

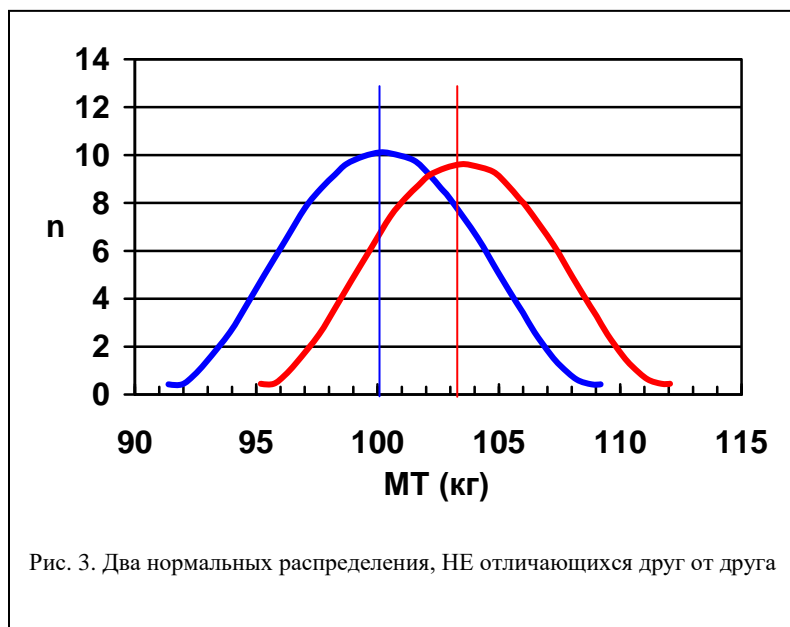
Когда мы проводим исследования данного типа, мы делаем заключение на основе средних значений и стандартных отклонений для каждой группы. Мы решаем, являются ли средние значения аналогичными или между ними есть различие. И, конечно, не исключено, что наше решение будет неверным. В этом случае существуют два пути совершения ошибки. Первый вид ошибки (именуемый ошибка типа I) возникает когда мы заключаем, что средние значения различаются (эффект от обработки имел место), когда как на самом деле различия не было. Это называется ложноположительным результатом — хорошим примером может служить ситуация, когда тест на беременность показывает положительный результат, в то время как беременность на самом деле отсутствует.

Второй вид ошибки (именуемый ошибка типа II) возникает тогда, когда мы заключаем, что средние значения не различаются (эффекта от обработки не было), когда как на самом деле средние значения различались. Такой результат называется ложноотрицательным — примером может служить ситуация, когда тест на беременность показывает отрицательный результат, в то время как женщина на самом деле беременна.

«Статистически значимо»

Принято считать, что по большей части статистические исследования обеспечивают контроль риска появления ложноположительных результатов, так как, согласно их условиям, заключение об отличии средних значений для экспериментальной группы должно быть верным в 95% случаев. Это также означает, что мы будем ошибаться (т. е. заключать, что существовало различие между группами в рамках исследования, когда на самом деле различия не было) только в 5% случаев. Отсюда и берется уровень значимости $P < 0,05$. Вероятность вынесения неверного заключения составляет 5% или менее.

Итак, когда кто-то заявляет, что два средних значения различаются с показателем уровня значимости (P) $< 0,05$, это означает, что если бы эксперимент был повторен 100 раз *при одинаковых условиях*, в 95% случаев результаты были бы аналогичны результатам первого эксперимента. При этом, даже если средние значения не будут полностью совпадать, относительные различия должны быть аналогичными.



Традиционно о «статистической значимости» или «значимости» результатов говорят при уровне значимости $P < 0,05$. Однако некоторые исследователи считают результат статистически значимым при $P < 0,10$ вместо $P < 0,05$. В целом продемонстрировать «значимость» гораздо легче, если вы используете более высокий показатель параметра P . Важно, чтобы вы понимали, что имеет в виду ваш собеседник, когда он говорит о «значимости». Важно прояснить следующее: если мы говорим о значимости, то из какого уровня вероятности мы исходим — 5, 10% или какого-то еще?

Что это может значить для вас?

В наши дни люди легко бросаются понятием «статистической значимости». Однако, как потребитель информации, вы должны быть осведомлены о ряде важных последствий и проявлять осмотрительность. Знакомясь с результатами исследований, вы должны учитывать следующие важные аспекты.

1. Целью большинства исследователей является установление статистически важного различия, если оно имеет место. Это означает, что для исследователя важно в максимально возможной степени снизить уровень различий и отклонений при проведении эксперимента (например, стараться вывести графики, подобные рис. 2 вместо рис. 3). Чтобы добиться этого, исследователи пытаются взять под контроль как можно больше аспектов эксперимента — животных, рационы, уход, содержание, среду и т. д. Такой контроль увеличивает возможности исследователя в части прослеживания изменений, но какой эффект это может принести вам? Уровень изменчивости в условиях вашего хозяйства может быть выше, потому что вы не в состоянии контролировать все переменные так, как это делает исследователь. Это наиболее распространенная причина, по которой животноводы, испытывающие новые продукты, не могут увидеть изменений, обещанных компаниями, предлагающими им новые технологии.
2. При каких условиях проводилось исследование? Исследователи могут искусственно контролировать условия эксперимента для повышения вероятности получения статистически значимых различий. Улучшение показателей роста от рождения до отъема можно увидеть в том случае, когда кормовая добавка вводится в заменитель молока, скармливаемый в объеме 454 г/день без дополнительного кормления стартером, но кто из животноводов на самом деле откажется от скармливания своим телятам стартера в первые восемь недель жизни? При изучении материалов исследования ищите признаки «искусственного» ухода за животными, которые должны насторожить вас и вызвать сомнения в истинной эффективности испытываемой технологии.
3. Обращайте внимание на популяции, в которых проводилось исследование. Исследователи, работающие с телятами, для своих экспериментов зачастую используют бычков (особенно до отъема), так как они отличаются более низкой стоимостью. Мы обычно исходим из того, что данные, собранные в отношении бычков, будут применимы и для телочек. Но всегда ли это так? Также всегда обращайте внимание на то, какие породы телят используют исследователи, ведь они могут отличаться от тех, с которыми работаете вы.
4. Остерегайтесь маломасштабных исследований! Согласно правилам статистики, если в вашем эксперименте используется небольшое число животных, заявить о его статистической значимости будет более затруднительно. Небольшое число животных также может означать их большую однородность, что уменьшает уровень изменчивости при проведении эксперимента. В связи с этим результаты исследования могут быть менее применимы в реальных условиях. При проведении экспериментов с маленькими телятами лично для меня предпочтительными являются группы численностью в 25–30 голов. Я всегда с настороженностью воспринимаю результаты исследований, в которых использовалось 10 или менее телят.

5. Какова степень различия между группами в рамках исследования? Многие исследователи испытывают большую радость по поводу статистической значимости, но при этом теряют из вида фактические показатели. Они могут увидеть «значимое отличие», составляющее 250 грамм прироста массы тела в день, но что это означает для животновода? Даже при наилучших условиях вы никогда не сможете измерить увеличение прироста массы тела на 250 грамм на своей ферме.
6. Какое значение P используется?
7. Что конкретно измеряется? Многие исследователи, ставящие эксперименты на телятах, измеряют такие параметры как прирост массы тела, увеличение роста, эффективность корма и т. д. Но измеряете ли вы эти параметры в своем хозяйстве? Являются ли они важными для вас — если телята при отъеме будут весить на 5 килограмм больше, что это может означать для вас с экономической точки зрения? Проанализируйте эти параметры и обдумайте, как их изменение может повлиять на ваше предприятие.
8. Какова потребительская ценность рассматриваемого предложения? Если вы вложите в новую технологию по 1 доллару из расчета на каждого теленка, что вы ожидаете получить взамен? Какова рентабельность инвестиций (ROI, return on investment) — экономическая выгода, поделенная на стоимость. Например, ежегодно вы вкладываете 10 000 долларов в технологию, снижающую уровень падежа телят до отъема на 5%, что экономит вам 50 000 долларов в год. Рентабельность инвестиций составляет $50\,000 / 10\,000 = 5:1$. Большинство животноводов считают минимально приемлемой рентабельность инвестиций 3:1.

Данные, полученные в результате хорошо организованных исследований, могут помочь животноводам понять потенциальную ценность новой технологии. Для того, чтобы полностью понять, насколько реальны различия между «старой» и «новой» технологиями, необходим надлежащий статистический анализ, который поможет вам отделить «зерна от плевел». Однако существует множество способов «солгать с помощью статистики». Тот факт, что кто-то рассказывает вам о продаваемой им технологии, результатом которой стали «значимые отличия», вовсе не означает, что применив ее в своем хозяйстве, вы обязательно получите тот же великолепный результат. Желаю удачи!

Автор: д-р Джим Кигли (30 сентября 2007 года).

© Д-р Джим Кигли, 2007

Calf Notes.com (<http://www.calfnotes.com>)