

ЛАБОРАТОРНАЯ РАБОТА № 9. ПОНЯТИЕ «КОРРЕЛЯЦИИ». МАТРИЦА КОРРЕЛЯЦИИ

Понятие «корреляция» у животных часто имеет место сопряженная (совместная) изменчивость признаков, например, удою и жирномолочности, яйценоскости и веса яйца и т.д. Совместную изменчивость разных признаков называют «корреляция» (co-relation -связь, соотношение) и обозначают символом «г»

Корреляционная связь является, во-первых, вероятностной - изменение одного признака у ряда особей на определенную величину сопровождается изменениями другого признака на различные (варьирующие) значения; во-вторых, статистической - проявляется лишь в среднем для всей выборки; в отношении отдельных наблюдений она очень неполная и неточная (см. также [116]). Корреляционную связь следует отличать от функциональной. При последней изменение одного показателя (аргумента) на определенную величину приводит к изменению другого показателя (функции) тоже на определенную величину (как, например, в формуле площади круга - $2R\pi S=$, здесь R - радиус круга;...14,3π=). Корреляция не вскрывает причины связи. Она дает лишь оценку силы, или тесноты связи между переменными.

Однако знать корреляции важно. Так, при селекции животных никогда не отбирают только по одному признаку. Более того, это невозможно, т.к. селекционируются особи. А особь - это десятки признаков, которые необходимо учитывать при отборе. Если бы корреляция между признаками отсутствовала, то селекция была бы проще. Отбор мог бы проводиться независимо и отдельно по каждому признаку. Корреляцию можно рассчитать для любой пары признаков. Однако должно быть «биологическое обоснование» взаимосвязи. Например, как биологически объяснить корреляцию между инвентарными номерами быков и удоем их дочерей?* Понятие «корреляция» в современном значении появилось в середине XIX века благодаря работам сэра Френсиса Гальтона (двоюродного брата Чарльза Дарвина) и Карла Пирсона.

Через 20 лет после того, как Френсис Гальтон впервые приступил к решению проблемы вероятностной взаимосвязи, К.Пирсон обнаружил, что эта задача была решена 50 лет назад французским астрономом А Бравэ в статье об ошибках в определении нахождения точки в пространстве.

Задачи корреляционного анализа

Корреляционный анализ призван решать следующие задачи:

1. Измерение степени связности двух и более переменных. Наши общие знания об объективно существующих причинных связях должны дополняться научно обоснованными знаниями о количественной мере зависимости между переменными. Данный пункт подразумевает верификацию уже известных связей.

2. Обнаружение неизвестных причинных связей. Корреляционный анализ непосредственно не выявляет причинных связей между переменными, но устанавливает силу этих связей и их значимость. Причинный характер выясняют с помощью логических рассуждений, раскрывающих механизм связей.

3. Отбор факторов, существенно влияющих на признак. Самые важные те факторы, которые сильнее всего коррелируют с изучаемыми признаками.

14.3. Характер и сила связи Коэффициенты корреляции могут варьировать от -1 до +1. При положительных корреляциях зависимость между признаками прямая: с увеличением одного увеличивается и другой признак. При отрицательных корреляциях зависимость обратная: увеличение одного признака приводит к уменьшению другого.

Нулевая корреляция свидетельствует о независимой изменчивости двух признаков - нет линейной связи между признаками. Однако вполне возможно, что при этом существует нелинейная связь. Коэффициент корреляции на уровне 0,5 представляется достаточно высоким. Можно даже полагать, что при такой корреляции совпадение вариации двух переменных должно быть в 50% случаев.

В действительности это не так. Степень линейной зависимости, «связности», в вариации двух переменных более точно измеряется квадратом коэффициента корреляции - коэффициентом детерминации ($2r$). Коэффициент детерминации изменяется от 0 до 1. В случае прямолинейной связи коэффициент детерминации указывает на долю изменчивости переменной Y , которая обусловлена изменчивостью переменной X (и наоборот). Тогда $1-2r$ - это остаточная доля изменчивости признака Y , обусловленная всеми другими, не учтенными в эксперименте причинами. Так, если коэффициент корреляции между двумя признаками равен 0,5, то только 25% изменчивости одного признака объясняется изменчивостью другого признака (степень связности). По остальной же части изменчивости соотношение между признаками чисто слу-

чайное. Таким образом, корреляция $\geq 0,7$ свидетельствует о тесной связи, порядка 0,5...0,6 – о средней и $< 0,5$ - указывает на слабую связь.

Виды корреляций

Корреляции могут быть: относительно характера проявления статистической связи - положительными и отрицательными; по форме связи - линейными и нелинейными; по числу переменных - простыми (парными), множественными (между более чем двумя переменными) и частными - между двумя переменными при «фиксированном» влиянии остальных переменных. Относительно природы источника совместной изменчивости различают корреляции фенотипические, паратипические (средовые) и генетические (см. также [101,131,137,143]).

Высокая паратипическая корреляция указывает на то, что значения признаков можно повысить, улучшая одни и те же условия среды (кормление, содержание). Для селекционера важно, в какой степени фенотипическая связь между признаками обусловлена средой и в какой – наследственностью. При высокой генетической корреляции отбор животных можно ограничить только одним из признаков, как правило тем, который проще измерять.

В этом случае можно сократить затраты на контроль других признаков. Их улучшение пойдет «само собой» при отборе по контролируемому признаку. Высокие генетические и паратипические корреляции при высокой фенотипической свидетельствуют о том, что на совместную фенотипическую изменчивость признаков одновременно и очень заметно влияют как средовые, так и генетические факторы. Часто бывает, что высокая фенотипическая корреляция не сопровождается столь же высокой генетической. В этом случае отбор по фенотипу одного признака приведет только к незначительному одновременному улучшению другого признака. Генетические и паратипические корреляции могут различаться и по знаку. Различие в знаках означает, что источники изменчивости, обусловленные наследственностью и средой, влияют на признаки посредством различных физиологических механизмов.

Простая линейная корреляция

Выше отмечалось, что корреляция измеряет совместную изменчивость двух (и более) признаков. Однако непосредственно сравнить изменчивости признаков нельзя, т.к. они, как правило, выражаются в разных единицах измерения.

Проблема с ковариансой, как меры связи между переменными, состоит в том, что ее размерность зависит от шкалы измерения. Например, если бы скорость роста измерялась в кг/сутки вместо г/сутки, то коэффициент корреляции был бы 0,05167. Следовательно, размер ковариансы без единиц измерения не имеет никакого значения.

На величину коэффициента парной корреляции могут оказывать влияние другие переменные. Интенсивность связи в «чистой» форме определяют с помощью коэффициента частной корреляции. В этом случае сопряженную вариацию между двумя переменными оценивают при фиксировании (исключении) влияния остальных переменных.

Множественная корреляция. В биозоотехнических исследованиях чаще всего встречаются сложные взаимосвязи между переменными. Для определения интенсивности или тесноты связи одной из переменных с совокупностью переменных используют коэффициент множественной корреляции. Например, коэффициент корреляции $R_{1.2}$ указывает интенсивность связи при условии, что переменная Y одновременно зависит от переменных X_1, X_2, \dots, X_n .

Коэффициенты множественной корреляции варьируют от 0 до 1. По их значениям нельзя сделать вывод о характере взаимосвязи, т.е. «+» или «-» корреляции между переменными. Только если все коэффициенты парной корреляции имеют одинаковый знак, то этот знак можно отнести также к коэффициенту множественной корреляции и утверждать о соответствующем характере множественной связи. Используя матричную форму записи, выражение коэффициента множественной корреляции для любого числа объясняющих переменных можно получить из уравнения.

Ранговая корреляция Для вычисления парных корреляций необходимо, чтобы исходные данные были выражены достаточно точно и имели

Из кн.: Кузнецов В.М. Основы научных исследований в животноводстве. Киров: Зональный НИИСХ Северо-Востока, 2006.- 568 с. 236 нормальное распределение. Это не всегда возможно. Существуют признаки, которые с трудом поддаются точной оценке, например, балл за

экстерьер. Кроме того, распределение одного или обоих признаков может быть очень неравномерным и неправильным. В таких случаях для количественной оценки связи между признаками используют метод ранговой корреляции Спирмена (этот и последующий метод относятся к непараметрической статистике; см. главу 12). В данном методе необходимы не точные значения количественных признаков, а их ранги (порядковые номера животных по соответствующему признаку). Коэффициент ранговой корреляции является парным. Поэтому оценивается соответствие между двумя рядами порядковых номеров. Ранги присваивают по нисходящей: от большего значения к меньшему. Если встречаются два или более животных с одинаковыми (связными) значениями, то используют метод «средних рангов». Например, присвоили по живой массе ранги пяти лучшим животным. У следующих трех животных живая масса была по 420 кг. Необходимо усреднить ранги, которые имели бы эти животные, если бы их значения различались: $(6+7+8)/3=7$. Таким образом, всем трем животным присваивают ранг 7. Последующему, с живой массой ниже 420 кг, но выше, чем у остальных не ранжированных, присваивают ранг 9.