

КОЭФФИЦИЕНТ ИЗМЕНЧИВОСТИ – ЕГО ОБОСНОВАНИЕ И ЛИНЕАЛИЗАЦИЯ ШКАЛЫ ИЗМЕРЕНИЯ

Андрей БАБИЦКИЙ

Научно–производственное объединение ЗЕЯ, Молдова

Analizei teoretice a fost supus coeficientul variabilității indicelui cantitativ (IC) în biologie. S-a demonstrat că după sensul său acesta este tangenta unghiului de înclinare dintre pătratul mediu al IC și mărimea lui medie aritmetică. Este demonstrată gradația neliniară a măsurării IC. Pentru măsurarea IC s-a efectuat o nouă gradație. Sunt propuși noi indici statistici pentru aprecierea variabilității IC în investigațiile biologice. Acești factori sunt varianta relativă și neschim-barea IC. S-au propus formule noi pentru studierea analitică a IC în procesele biologice în locul celor descrise anterior.

The coefficient variability (CV) of a quantitative character (QC) in biology has been undergone to theoretical analysis. Its mathematical sense as tangent of a corner of an inclination between an average square of a QC and its average arithmetic quantity was proved. The non-linearity of scale calculated values of the CV was shown. The new linear scale for measurement of variability of a QC was created. There are offered new statistical factors for estimating of processes variation of QC in biological researches. They are the factors of relative variance and invariantness of QC. The new formula for analytical studying variability of QC in biology, as process, instead of as conditions as it was in earlier described formula has been given.

Для выражения степени изменчивости количественных признаков в антропологии Карл Пирсон [14], из эмпирических соображений, ввел более 100 лет тому назад в статистику безразмерный относительный критерий изменчивости количественного признака (КП) и назвал его коэффициентом изменчивости (КИ) (coefficient of variability – CV), вычисляемым в соответствии со следующим математическим выражением:

$$CV = \frac{\sigma}{\mu} 100\%, \quad (1)$$

где σ - среднее квадратическое отклонение КП,
 μ - среднее значение КП.

Из этой формулы явствует, что КИ представляет собой ту часть среднего квадратического отклонения, которая приходится на единицу среднего значения КП. Поскольку КП варьируют в широких пределах, то чтобы не иметь дело с малыми величинами шкала их измерения была зафиксирована при значении 100 единиц измерения путем умножения на 100, в результате КИ преобразован в величину среднего квадратического отклонения, приходящуюся на 100 единиц средней величины КП. Таким образом, из КИ были удалены наименования единиц измерения и КИ преобразовался в относительный и безразмерный признак в стандартизированной форме процентного выражения. В такой форме КИ стал широко использоваться экспериментаторами для оценки изменчивости КП, вначале в антропологии, а затем повсеместно в биологии и, наконец, распространился на все области прикладной статистики. Он дает количественную меру изменчивости, чаще всего как конечного состояния КП биологических объектов, но он не дает возможности для аналитического изучения их изменчивости как динамического процесса. Наряду с этим упущением Пирсон не дал и формулу для исчисления параметра, дополнительного к КИ, а именно – показателя неизменности КП. Величина, обратная КИ, названная однородностью КП, таким свойством не обладает [2].

Многие экспериментаторы пытались применить КИ в аналитических целях для исследования целого ряда биологических явлений. Однако сразу же выявилась масса случаев несоответствия биологического смысла показателя изменчивости математически получаемому КИ. Поскольку биологические КП взаимокоррелируемы между собой и не являются однородными, некоторые из них математически представляют одномерную функцию, например – длина тела, другие – двумерную, как его поверхность, третьи – трехмерную, как его объем, и далее существуют всевозможные комбинации таких

варьирующих признаков. Стало ясно, что КИ не может быть аналитическим критерием или инструментом исследования, вскрывающим биологические закономерности формирования КП. Некоторые из возникающих парадоксов применения КИ в биологии были рассмотрены А.А. Любишевым [10], который по этим и по ряду других причин предложил вовсе отказаться не только от его аналитического использования ввиду недостаточной обоснованности этого показателя, но также и от оценки изменчивости конечных состояний КП организмов. И все же за неимением другого критерия, этот показатель нашел свое применение в сугубо прикладных целях в качестве грубой оценки изменчивости КП. Переместившись в собственно прикладную область статистики, КИ уже не вызывал потребности в его математическом обосновании из-за кажущейся очевидности. На протяжении длительного периода времени математическая сторона этого критерия так и оставалась нерассмотренной. По прошествии еще 60 лет [11] при рассмотрении правомерности повсеместного использования в биологии КИ пришлось вновь отметить сохраняющуюся нерешенность задачи математической обоснованности показателя КИ, его очень слабую теоретическую обоснованность. Это вынуждает высказать множество возражений против его использования, поскольку в биологии есть области, где он совершенно не верен. Попытки заменить математический анализ обширными биологическими исследованиями изменчивости множества КП при широком выборе биологических объектов приводят лишь к накоплению данных, трудно поддающихся осмысленному анализу [12,13]. Все КП в биологии взаимокоррелируемы и управляются несколькими контурами отрицательной обратной связи [5, 6]. Без теоретического анализа КИ, его смыслового содержания и границ его применимости биологические данные не подлежат интерпретации. Ситуация несколько прояснилась, когда удалось выснить, что математически КИ имеет природу тангенса и является нелинейной величиной [2 -4], однако многие стороны теоретического обоснования продолжали оставаться нерешенными, выяснение которых и является предметом данной публикации.

Для теоретического анализа КИ обратимся к рассмотрению основных параметров статистического распределения варьирующего КП на воображаемой модели. Для построения такой модели представим, что количественный признак формируется в серии элементарных событий активации генов во время периодов роста и периодов покоя, что хорошо доказано биологами. Рассмотрим ряд моментов статистического процесса:

$$\mu_k = E [x^k] = \int_{-\infty}^{+\infty} x^k p(x)dx, \tag{2}$$

где $p(x)$ – плотность вероятности, μ_k – k - тый момент, $k=0,1,2..$ Нулевой момент при $k = 0$ равен площади под кривой нормального распределения и принят за единицу:

$$\mu_0 = E [x^0] = \int_{-\infty}^{+\infty} x^0 p(x)dx = 1. \tag{3}$$

Первый момент при $k=1$ дает среднее значение μ количественного признака:

$$\mu_1 = E [x^1] = \int_{-\infty}^{+\infty} x^1 p(x)dx = \mu. \tag{4}$$

Второй момент, при $k=2$, является средним квадратом количественного признака:

$$\mu_2 = E [x^2] = \int_{-\infty}^{+\infty} x^2 p(x)dx = \varphi^2. \tag{5}$$

Поскольку количественный признак растения – величина всегда положительная ($\mu > 0$), то для его учета используется и второй момент, вычисляемый относительно условного нуля, который перенесен в точку μ , являющуюся средним значением КП признака и представляющую центр нормального распределения. Этот второй момент именуется вторым центральным моментом μ_2^c , или дисперсией:

$$\mu_2^c = E [x - \mu]^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x)dx = \sigma^2. \tag{6}$$

При $\mu = 0$ выражение (6) превращается в (5), т.е. $\varphi^2 = \sigma^2$.

Из формулы (6) выводится важнейшее уравнение, связывающее воедино все три момента статистического распределения количественного признака:

$$\sigma^2 = E [(x - \mu)^2] = E [x^2 - 2x\mu + \mu^2]. \quad (7)$$

Теперь, с учетом аддитивности математического ожидания, имеем:

$$\sigma^2 = E [x^2] - 2E[x] \mu + \mu^2. \quad (8)$$

Подстановкой в уравнение (8) значений математического ожидания первого и второго моментов (4) и (5) получаем формулу (9), исходную для прояснения сущности коэффициента изменчивости – КИ:

$$\varphi^2 = \sigma^2 + \mu^2. \quad (9)$$

Уравнение (9) есть не что иное, как выражение теоремы Пифагора. Отсюда уже видно, что среднее значение количественного признака μ и среднее квадратическое отклонение σ являются катетами прямоугольного треугольника, гипотенуза которого суть среднее значение квадрата КП – φ , а используемая биологами формула (9) для определения КИ представляет собой тангенс угла наклона между гипотенузой φ и катетом μ . При сравнении смысловых значений формул (1) и (9) явствует, что в формуле (1) приведена функциональная зависимость между двумя величинами σ и μ , в свою очередь находящимися в функциональной зависимости от третьей величины φ , не представленной в этой формуле.

Следовательно, предложенный Карлом Пирсоном критерий изменчивости КП в виде КИ (1) страдает неполнотой и незавершенностью такого способа его математического выражения и не может применяться в аналитических целях при изучении биологических данных, а лишь в грубо оценочных прикладных целях. В биологических исследованиях очень важно следить за возрастной динамикой изменчивости КП, поскольку это позволяет выявить взаимные корреляции и обратные связи в биологических процессах роста и развития организмов [1].

Анализ функциональной зависимости КИ (1), представляющий тангенс угла, показывает, что все же четко определенной зависимости мы не имеем. Так, при неизменной величине среднего квадратического отклонения σ и варьирующего среднего значения признака μ , мы имеем гиперболическую связь между коэффициентом изменчивости КИ и средним значением признака μ . При возрастании среднего квадратического отклонения σ , при неизменной величине среднего значения признака μ , мы имеем линейную прямо пропорциональную зависимость. При одновременном изменении как σ , так и μ , мы имеем тангенциальную зависимость, уходящую в бесконечность. В двух случаях шкала для измерения КИ является нелинейной. Исключение составляет тот случай, когда среднее квадратическое отклонение σ линейно зависит от средней величины признака μ и КИ является постоянной величиной. Именно этот случай и считается [7,10,11] приемлемым для использования формулы Карла Пирсона (1). Количественные признаки, подчиняющиеся этому критерию А.А. Любищева, где есть основания принимать, что при одинаковой изменчивости среднее квадратическое отклонение возрастает пропорционально среднему арифметическому, названы Ю.С. Куршаковой [9] элементарными. Однако это редкий случай в биологических исследованиях, и настоятельно необходимо найти линейную шкалу для всех видов сложности КП. Такой шкалой может быть угловое представление величины КИ [2,3]. Однако представление углов в градусной форме крайне неудобно из-за дробной части, выражаемой минутами в 60-ричной шкале. Поэтому рационально углы измерять в десятиричной системе исчисления, представив их в радианной мере. В таком виде параметр относительной изменчивости приобретает линейную шкалу измерений и может применяться в прикладных целях. Вместе с тем, в аналитических целях нам очень важно следить за изменчивостью среднего значения квадрата количественного признака φ , который в формуле (1) находится в неявном виде и исключен из рассмотрения. Чтобы это обнаружить, нам необходимо обратиться к уравнению (9). Разделим все слагаемые этого уравнения на величину среднего значения квадрата КП – φ^2 . Тогда мы получим уравнение стандартизированной окружности

$$1 = \left(\frac{\sigma}{\varphi}\right)^2 + \left(\frac{\mu}{\varphi}\right)^2 \text{ или } 1 = \sin^2 \alpha + \cos^2 \alpha, \quad (10)$$

с диаметром, равным единице, на который опирается вписанный прямоугольный треугольник с прямым углом, образованным катетами $\frac{\sigma}{\varphi}$ и $\frac{\mu}{\varphi}$. Итак мы получили исходную формулу для создания

критерия как вариантности, так и инвариантности КП в биологии.

Показатель вариантности V (изменяемости) КП представляет собой квадрат относительной дисперсии:

$$\text{Var} = \left(\frac{\sigma}{\varphi}\right)^2, \text{ или в процентном отношении } \text{Var} = \left(\frac{\sigma}{\varphi}\right)^2 * 100\%. \quad (11)$$

Он является дополнительной величиной к показателю инвариантности - Inv (неизменности) КП, равному

$$\text{Inv} = \left(\frac{\mu}{\varphi}\right)^2, \text{ или } \text{Inv} = \left(\frac{\mu}{\varphi}\right)^2 * 100\%. \quad (12)$$

Разделив значение инвариантности, выраженное формулой (12), на вариантность, выраженную формулой (11), получим формулу квадрата однородности U^2 , эквивалентного количеству активаций генов $m = U^2$ при формировании данного КП [2]. Разделив катеты $\frac{\sigma}{\varphi}$ и $\frac{\mu}{\varphi}$ из формулы (10) между

собой, получим тангенс угла наклона большего катета к диаметру и придем к исходной формуле (1), в которой неявно подразумевается константность значения квадрата КП, представляющего в аналитических целях наибольший интерес. Для создания аналитического показателя изменчивости КП нам необходимо в математическое выражение КИ ввести и третий функциональный элемент, неявно связанный с изменчивостью, это среднее значение квадрата КП – φ . Это наиболее удобно сделать при переходе к полярной системе координат, при выражении функциональной зависимости радиус-вектора от угла α между радиус-вектором и полярной осью.

$$\rho = f(\alpha),$$

или в виде выражения функциональной зависимости среднего значения квадрата КП – φ от угла α между ним и средним значением КП – μ :

$$\varphi = k\alpha, k = \frac{\varphi}{\alpha}, \quad (13)$$

где α – угол в радианах, $\alpha = \arctg \frac{\sigma}{\mu}$, k – коэффициент пропорциональности, являющийся показателем обратной связи между процессами роста КП с его уже существующей величиной.

Этот показатель является именованным и зависит от масштаба измерения и не является относительным. Он предназначен для исследования кинетических процессов жизнедеятельности на конкретном объекте исследования, ценность которого нами будет продемонстрирована в последующей публикации по изучению кинетики роста листьев у кукурузы.

Выводы

Предпринят теоретический анализ параметра коэффициент изменчивости количественного признака в биологии. Показан его математический смысл как тангенса угла наклона между средним квадратом мерного признака и его средней арифметической величиной. Обнаружено, что в неявном виде величина тангенса угла, равная отношению среднего квадратического отклонения к средней величине признака, является также и функцией среднего квадрата мерной величины. Показана нелинейность шкалы измерения коэффициента изменчивости. Создана новая линейная шкала для измерения изменчивости количественного признака. Предложены новые параметры для оценки статистических процессов при оценке количественных признаков: вариантность и инвариантность количественного признака. Посредством ввода всех трёх статистических показателей функциональной зависимости значения количественного признака в виде единого выражения, предложена новая формула для аналитического изучения изменчивости количественных признаков в биологии как динамического процесса, а не как статического состояния, как это было согласно ранее существовавшим формулам.

Литература:

1. Бабицкий А.Ф. Онтогенетический градиент коэффициента изменчивости одномерного количественного признака кукурузы // Онтогенетика высших растений. - Кишинев, 1989, с.9-10.
2. Бабицкий А. Ф. Изменчивость длины листа при гетерозисе кукурузы // Известия АН Молдовы. Серия «Биол. и хим. науки». - 1991. - Том 6. - С.15-22.
3. Бабицкий А.Ф. Коэффициент изменчивости – его смысл и границы применимости // *Genetica și ameliorarea plantelor și animalelor în Moldova*. - Кишинев, 1992, с.11-12.
4. Бабицкий А.Ф. Стабильность варьирующего признака // *Probleme actuale ale Geneticii, biotehnologiei și ameliorării: Materialele Conferenței Naționale*. - Кишинев, 2005, с.211-216.
5. Бабицкий А.Ф. Коэффициент изменчивости и отрицательная обратная связь в биологических системах // *The 2-nd International Conference «Telecommunications, Electronics and Informatics»*. Proceedings. Volume 1. - Chishinau, 2008, p.509-514.
6. Бабицкий А.Ф., Симинел В.Д. Коэффициент изменчивости и эффект гетерозиса // *Генетические основы селекции с/х растений и животных // Генетические основы селекции с/х растений и животных: Тезисы Респ. конф.* - Кишинев, 1984, с.7.
7. Берлянд – Кожевников В.М. Возможности использования коэффициента вариации для анализа количественных признаков у растений // *Труды Ленинградского общества естествоиспытателей*. Том 72. Вып. 5. - Ленинград, 1975, с.34-45.
8. Гудвин Б. Аналитическая физиология клеток и развивающихся организмов. - Москва: Мир, 1979. - 288 с.
9. Куршакова Ю. С. О вариабельности размерных признаков и способах ее оценки // *Вопросы антропологии*. - 1965. - Вып. 21. - С.53-64.
10. Любищев А.А. О критерии изменчивости организмов // *Известия Биологического НИИ Пермского университета*. - 1923. - Т.1. - Вып.7-8. - С.121-128.
11. Любищев А.А. Дисперсионный анализ в биологии. - Москва: Издательство МГУ. 1986. - 200 с.
12. Филипченко Ю.А. Изменчивость и методы ее изучения. - Москва: Наука, 1978. - 238 с.
13. Яблоков А. В. Изменчивость млекопитающих. - Москва: Наука, 1966. - 363 с.
14. Pearson C. Contributions to the mathematical theory of evolution. 111. Regression, heredity and panmixia // *Phil. Trans. Roy. Soc. of London. Ser. A*. - 1896. - Vol.167. - P.253-318.
15. Бабицкий А.Ф. Коэффициент изменчивости и отрицательная обратная связь в биологических системах // *The 2-nd International Conference «Telecommunications, Electronics and Informatics»*. Proceedings. Volume 1. - Chishinau, 2008, p.509-514.

Prezentat la 03.04.2008