

МЕТОДИКА БОТАНИЧЕСКИХ ИССЛЕДОВАНИЙ

УДК 581.55

© В. С. Ипатов

**ОСТОРОЖНО — БИОМЕТРИКА. (ОБ ИСПОЛЬЗОВАНИИ ОЦЕНОК
«ДОСТОВЕРНОСТИ» ПРИ ИССЛЕДОВАНИИ
КОЛИЧЕСТВЕННЫХ ЗАКОНОМЕРНОСТЕЙ)**V. S. IPATOV. BE CAREFUL OF BIOMETRICS. (ON USING «SIGNIFICANCE»
ESTIMATIONS IN RESEARCHES OF QUANTITATIVE REGULARITIES)

Санкт-Петербургский государственный университет

199034 С.-Петербург, Университетская наб., 7/9

Тел. (812) 328-14-72

E-mail: vsipatov@mail.ru

Поступила 12.05.2010

Использование вероятностных оценок достоверности данных требует соблюдения условия нормальности распределения признаков. Обычно же в ботанических материалах распределения существенно отличаются от нормального или же малое число наблюдений не позволяет установить характер распределения. В этих случаях вероятностная оценка достоверности не имеет смысла, что часто не принимается во внимание ботаниками.

Ключевые слова: статистические методы, критерии, вероятность, распределение, достоверность.

В ботанике широко используются статистические методы и критерии. Например, такие критерии (параметры, показатели), как средняя арифметическая, разница между средними, дисперсия, коэффициент корреляции, регрессия и т. д. При этом обычно применяется оценка достоверности (существенности значимости) критериев.

Абсолютной достоверностью могут обладать только критерии, вычисленные в генеральной совокупности объектов, т. е. во всех объектах (выборках) данного рода (сходных по существенным признакам) в одних и тех же условиях.

Вместе с тем провести исследование всей генеральной совокупности практически (чаще всего) невозможно, прежде всего из-за большого множества объектов. Кроме того, необходима оценка сходства объектов для определения правомерности отнесения их к одной генеральной совокупности.

По этой причине исследователь вынужден ограничиваться только частью генеральной совокупности — выборочной совокупностью (выборкой), вычисляя для нее необходимые критерии и по ним уже судить о том, что выявившиеся критерии, закономерности свойственны генеральной совокупности, любым выборкам из нее. Насколько правдоподобно такое суждение, определяется вероятностью, с которой значение критерия может находиться в определенных пределах (интервале). Для определения пределов вычисляется, в частности, средняя квадратическая ошибка. Критические значения вероятности устанавливаются исследователем, но обычно ими считается вероятность 0.95, 0.99. Если вероятность оказывается меньше критической, то значение критерия (параметра) признается недостоверным.

Проблема заключается в том, что применение вероятностной оценки оправдано только в тех случаях, когда исследуемые варьирующие признаки в выборке соответствуют нормальному случайному распределению (рис. 1, Д) или не слишком от него отличаются. При этом насколько допустимо отклонение от нормального распределения не установлено. Но очевидно, что чем больше выборочное распределение отличается от нормального, тем менее надежна оценка вероятности и тем самым и достоверность.

По этой причине некоторые журналы вполне обоснованно требуют при использовании вероятностных оценок в указанном выше смысле подтверждения нормальности распределения. Перечень требований при использовании статистических методов (включая проверку нормальности распределения) приведен в электронном журнале «Биометрика» (www.biometrika.tomsk.ru). Отмечу, что в ботанических работах обычно не только не проверяется выборка на нормальность распределения, но часто нарушаются приемы определения достоверности. Прежде всего, это касается сравнения средних арифметических. Традиционно вычисляются квадратические ошибки средних и нередко при сравнении средних считают, что если разница между средними превышает сумму их ошибок, различие принимается как достоверное. При этом забывают, что использовать следует ошибку, умноженную на нормированное отклонение.

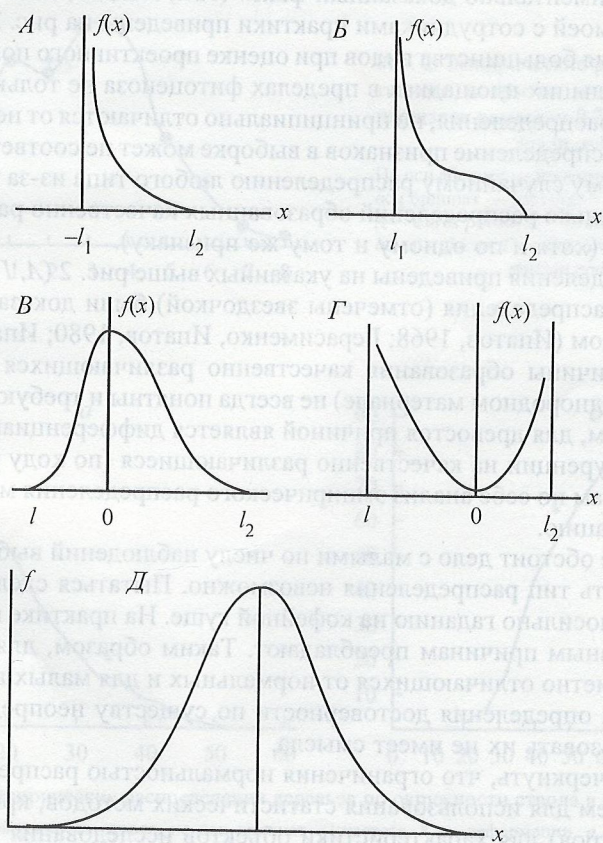


Рис. 1. Примеры кривых теоретического случайного распределения.

По оси абсцисс — значения признака; по оси ординат — частота. А—Г — распределения Пирсона типа I; Д — нормальное распределение.

Нормальное распределение в качественно однородной совокупности образуется, когда вероятность появления элементарного события (исследуемого признака) составляет 0.5 (рис. 1, Д). По определению нормальное распределение является случайным. Очевидно, случайных распределений в качественно однородных совокупностях может быть много при вероятности элементарного события, отличающегося от 0.5. Более того, сама эта вероятность может зависеть от значения признака.

С разнообразием кривых случайных распределений можно ознакомиться в книге А. К. Митропольского (1971, см. кривые Пирсона). На рис. 1 (А—Г) приведены некоторые из них.

Реально на практике нормальные распределения встречаются не чаще, чем иные типы распределений.

Приведу цитату из книги Дж. Эдни Юла и М. Дж. Кендэла. «Читатель может с полным основанием спросить, — говорил Карл Пирсон, — не является ли возможным найти материал, подчиняющийся в пределах вероятных границ нормальному закону? Да, отвечу я, но этот закон не есть всеобщий закон природы. Мы должны охотиться за подобными случаями». И далее: «Как однажды сказал мне М. Липпманн (Lippmann), — пишет Пуанкаре в своем «Исчислении вероятностей» (*Calcul des Probabilites*), — все верят в закон ошибок: экспериментаторы — потому что принимают его за математическую теорему, а математики — потому что принимают его за экспериментально доказанный факт» (Юл, Кендэл, 1960 : 225).

Примеры из моей с сотрудниками практики приведены на рис. 2 и 3. Полагаю, что распределения большинства видов при оценке проективного покрытия или фитомассы на небольших площадках в пределах фитоценоза не только отклоняются от нормального распределения, но принципиально отличаются от него. Кроме того, эмпирическое распределение признаков в выборке может не соответствовать вообще теоретическому случайному распределению любого типа из-за того, что включает в себя несколько распределений образованных качественно различающимися совокупностями (хотя и по одному к тому же признаку).

Такие распределения приведены на указанных выше рис. 2 (А, Г, Д) и 3. Деформации кривых распределения (отмечены звездочкой) были доказаны предложенным мной методом (Ипатов, 1968; Герасименко, Ипатов, 1980; Ипатов, Кирикова, 1997, 1999). Причины образования качественно различающихся совокупностей (казалось бы, в однородном материале) не всегда понятны и требуют специального анализа. Впрочем, для древостоя причиной является дифференциация древостоя в результате конкуренции на качественно различающиеся (по ходу роста) совокупности. Вообще сам по себе анализ эмпирического распределения может дать интересную информацию.

Еще сложнее обстоит дело с малыми по числу наблюдений выборками, для которых определить тип распределения невозможно. Пытаться сделать это логическим путем равносильно гаданию на кофейной гуще. На практике малые выборки, пожалуй, по разным причинам преобладают. Таким образом, для ненормальных выборок или заметно отличающихся от нормальных и для малых выборок вероятностные оценки определения достоверности по существу неопределенны, не надежны и использовать их не имеет смысла.

Следует подчеркнуть, что ограничения нормальностью распределения не служат препятствием для использования статистических методов, критериев (вычисленных параметров) для характеристики объектов исследования при любом распределении и в малых выборках.

В этих случаях надежность результатов, полученных в одной выборке, их правдоподобность — могут подтверждаться при использовании фундаментального на-

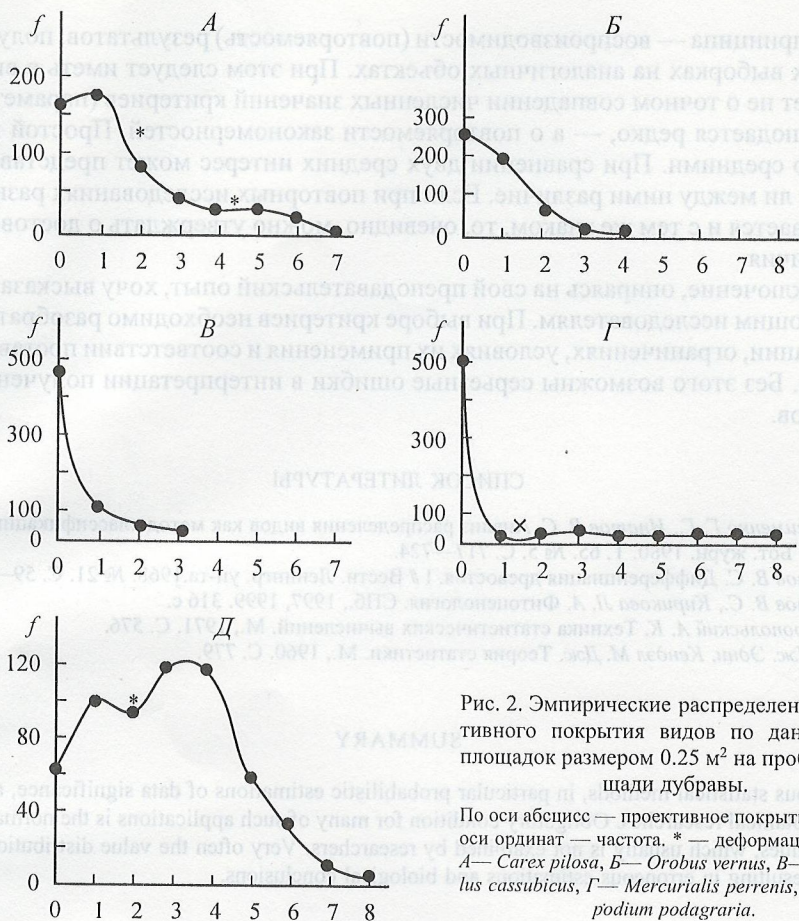


Рис. 2. Эмпирические распределения проективного покрытия видов по данным 580 площадок размером 0.25 м² на пробной площади дубравы.

По оси абсцисс — проективное покрытие вида, по оси ординат — частота. * — деформация кривой. А — *Carex pilosa*, Б — *Orobus vernus*, В — *Ranunculus cassubicus*, Г — *Mercurialis perrenis*, Д — *Aegopodium podagraria*.

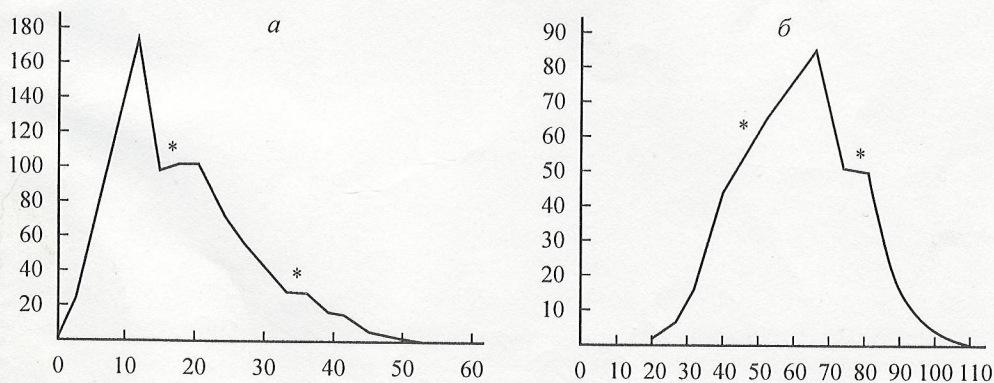


Рис. 3. Эмпирические распределения деревьев по окружности ствола в древостоях.

По оси абсцисс — окружность ствола, по оси ординат — частота. * — деформации. а — сосняк, 30—35 лет; б — сосняк, 85—90 лет.

учного принципа — воспроизводимости (повторяемость) результатов, получаемых в других выборках на аналогичных объектах. При этом следует иметь в виду, что речь идет не о точном совпадении численных значений критериев (параметров) — это наблюдается редко, — а о повторяемости закономерностей. Простой пример, опять со средними. При сравнении двух средних интерес может представлять то, имеется ли между ними различие. Если при повторных исследованиях разница обнаруживается и с тем же знаком, то, очевидно, можно утверждать о достоверности их различия.

В заключение, опираясь на свой преподавательский опыт, хочу высказать совет начинающим исследователям. При выборе критериев необходимо разобраться в их содержании, ограничениях, условиях их применения и соответствии поставленным задачам. Без этого возможны серьезные ошибки в интерпретации полученных результатов.

СПИСОК ЛИТЕРАТУРЫ

- Герасименко Г. Г., Ипатов В. С.* Анализ распределения видов как метод классификации растительности // Бот. журн. 1980. Т. 65. № 5. С. 717—724.
Ипатов В. С. Дифференциация древостоя. I // Вестн. Ленингр. ун-та. 1968. № 21. С. 59—68.
Ипатов В. С., Кирикова Л. А. Фитоценология. СПб., 1997. 316 с.
Митропольский А. К. Техника статистических вычислений. М., 1971. С. 576.
Юл Дж. Эдн, Кендэл М. Дж. Теория статистики. М., 1960. С. 779.

SUMMARY

Various statistical methods, in particular probabilistic estimations of data significance, are widely used in botanical researches. Obligatory condition for many of such applications is the normal distribution of values, which usually is not examined by researchers. Very often the value distributions are not normal, resulting in erroneous estimations and biological conclusions.