МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВПО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ
АГРАРНЫЙ УНИВЕРСИТЕТ»

Факультет экологии
Кафедра прикладной экологии

ЭКСПЕРИМЕНТАЛЬНАЯ ЭКОЛОГИЯ

Учебно-методическое пособие для практических занятий

Краснодар КубГАУ 2015

Составители: Горковенко Н.Е.

Пособие предназначено для оказания методической помощи при подготовке к семинарам по дисциплине «**Экспериментальная экология**», содержит тематику, цели и задачи каждого занятия, методику его проведения, задания для подготовки к семинарам, список рекомендуемой литературы.

Издание предназначено для обучающихся по направлению подготовки: 05.06.01 – Науки о Земле.

Рассмотрено и одобрено методической комиссией факультета экологии 29.06.2015 г., протокол № 10.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 1

Тема: Построение вариационного ряда.

Цель занятия: Сформировать у аспирантов представление о статистическом анализе результатов исследований.

Задачи: Освоить методику построения вариационных рядов.

Аспирант должен знать:

1) до изучения темы:
- Понятие о статистических методах анализа;
- Понятие статистической совокупности;

2) после изучения темы:
- Технику построения вариационных рядов дискретных признаков;
- Технику построения вариационных рядов непрерывных признаков.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Любое статистическое исследование должно начинаться с установления характера распределения изучаемых признаков. Распределение – это соотношение между значениями случайной величины и частотой их встречаемости. Бóльшая повторяемость одних значений по сравнению с другими заставляет задумываться о причинах наблюдаемых процессов. Если значения признака откладывать по оси абсцисс, а частоты их встречаемости – по оси ординат, то можно построить гистограмму, частотную диаграмму, удобную для целей иллюстрации и исследования.

Основой для построения гистограммы служит вариационный ряд – представленный в виде таблицы ряд значений изучаемого признака, расположенных в порядке возрастания с соответствующими им частотами их встречаемости в выборке.

Вариационным рядом называют двойной ряд чисел, показывающий, каким образом числовые значения признака связаны с их повторяемостью в данной статистической совокупности. Например, из урожая картофеля, собранного на одной из опытных делянок, случайным способом, т. е. наугад, отобрано 25 клубней, в которых подсчитывали число глазков. Результаты подсчета оказались следующие: 6, 9, 5, 7, 10, 8,9, 10,8, 11, 9, 12,9,8, 10, 11,9, 10,8, 10,7,9, 11,9, 10. Чтобы разобраться в этих данных, расположим их в ряд (в порядке регистрации результатов наблюдений) с учетом повторяемости вариант в этой совокупности:

Вариаиты x_i ….. 6 9 5 7 10 8 11 12

Число иариант f_j ….. 1 7 1 2 6 4 3 1

Это и есть вариациониый ряд. Числа, показывающие, сколько раз отдельные варианты встречаются в данной совокупности, называются частотами или весами вариаит и обозначаются строчной буквой латинского алфавита *f*. Общая сумма частот вариационного ряда равна объему данной совокупности, т. е.

$$\sum_{i=1}^{n} f_i = n,$$

где ∑ (греческая буква сигма прописная) обозначает действие суммирования, в данном случае суммирование частот вариационного ряда от первого (*i*= 1) до *k*-го класса, а *n* – общее число наблюдений, или объем совокупности.

Частоты (веса) выражают не только абсолютными, но и относительными числами – в долях единицы или в процентах от общей численности вариант, составляющих данную совокупность.

В таких случаях веса называют относительными частотами или частостями. Общая сумма частостей равна единице, т. е. $\sum f_i / n = 1$, или $\sum(f_i / n) \cdot 100 = 100$ %, если частоты выражены в процентах от общего числа наблюдений *n*. Замена частот частостями не обязательна, но иногда оказывается полезной и даже необходимой в тех случаях, когда приходится сопоставлять друг с другом вариационные ряды, сильно отличающиеся по их объемам.

Распределение исходных данных в вариационный ряд преследует определенные цели. Одна из них – ускорение работы при вычислении по вариационному ряду обобщающих числовых характеристик - средней величины и показателей вариации. Другая сводится к выявлению закономерности варьирования учитываемого признака, Приведенный ряд удовлетворяет первой, но не удовлетворяет достижению второй цели. Чтобы

ряд распределения полностью удовлетворял предъявляемым к нему требованиям, его нужно строить по ранжированным значениям признака.

Под ранжированием (от франц. *ranger* – выстраивать в ряд по ранжиру, т. е. по росту) понимают расположение членов ряда в возрастающем (или убывающем) порядке. Так, в данном случае результаты наблюдений следует распределить так:

Вариаиты x_i …..5 6 7 8 9 10 11 12

Частоты f_i . …… 1 1 2 4 7 6 3 1

Этот упорядоченный ряд распределения в равной мере удовлетворяет достижению и первой, и второй целей. Он хорошо обозрим и наилучшим образом иллюстрирует закономерность варьирования признака.

Рассмотрим пример изучения плодовитости серебристо-черных лисиц, которое дало следующие результаты (число щенков на самку): 5 5 6 5 5 6 4 4 4 5 6 4 6 6 4 6 4 5 5 8 5 3 6 5 5 5 5 5 6 3 6 4 6 4 6 2 5 6 5 3 7 6 3 4 6 8 6 3 5 5 6 5 4 3 8 4 7 5 4 3 1 6 5 3 4 5 6 7 4 4 6 5 6 4 6 5.

Для дискретного признака (такова плодовитость) построение вариационного ряда обычно не представляет сложности, достаточно подсчитать встречаемость конкретных значений.

Плодовитость, x	Частота, a
1	1
2	1
3	8
4	16
5	23
6	21
7	3
8	3

Гистограмма, построенная по данным о плодовитости лисиц (рис. 1), сразу же обнаруживает характерное поведение случайной величины – высокие частоты встречаемости значений в центре распределения и низкие – по периферии.

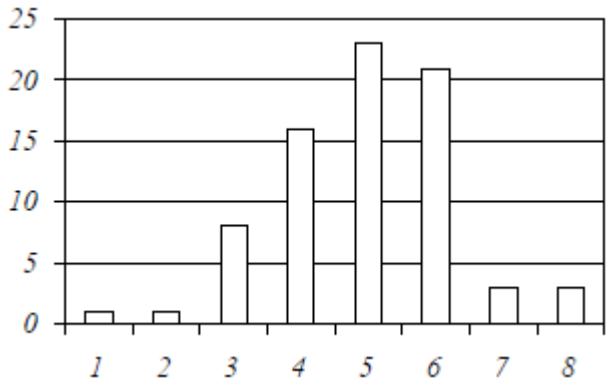


Рис. 1. *Распределение плодовитости лисиц*

Если же изучаемый признак непрерывен (таковы размерно-весовые характеристики), то для построения вариационного ряда сначала весь диапазон изменчивости признака разбивается на серию равных интервалов (классов вариант), затем подсчитывают, сколько вариант попало в каждый интервал.

Число классов для больших выборок (n > 100) должно быть не менее 7 и не более 12, их оптимальное число можно приблизительно определить по эмпирической формуле: $k = 1 + 3.32 \cdot \lg(n)$, где n – объем выборки (число вариант в выборке).

Составим для примера вариационный ряд для непрерывного признака – по данным о весе 63 взрослых землероек (г):

9.2 11.6 8.1 9.1 10.1 9.6 9.3 9.7 9.9 9.9 9.6
7.6 10.0 9.7 8.4 8.6 9.0 8.8 8.6 9.3 11.9 9.3
9.2 10.2 11.2 8.1 10.3 9.2 9.8 9.9 9.3 9.1 9.4
9.6 7.3 8.3 8.8 9.2 8.0 8.6 8.8 9.0 9.5 9.1
8.5 8.8 9.7 11.5 10.5 9.8 10.0 9.4 8.7 10.0 7.9
8.6 8.7 9.1 8.2 9.2 9.4 8.8 9.8

1) Все операции могут быть выполнены вручную. Вначале следует определить объем выборки n = 63.

2) Рассчитать пределы размаха изменчивости значений, лимит – разность между максимальным и минимальным значениями:

$Lim = x_{max} - x_{min} = 11.9 - 7.3 = 4.6$.

3) Найти число классов вариационного ряда по формуле:

$k = 1 + 3.32 \cdot \lg(63) = 6.973811 \approx 7$.

4) Найти длину интервала dx (допустимо округление):

$dx = Lim / k = 4.6 / 7 \approx 0.7$.

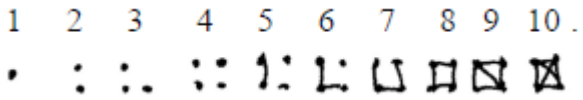
5) Установить границы классов; в качестве первой границы имеет смысл взять округленное минимальное значение: $x_{min} = 7$.

6) Вычислить центральное значение признака в каждом классе; исходным берется значение центра первого интервала; для первого класса 7–7.7, для второго – 7.8–8.4…

7) Произвести разноску вариант в соответствующие классы с подсчетом их числа методом конверта (табл. 1):

1 2 3 4 5 6 7 8 9 10 .

Теперь данные можно представить графически, в виде полигона частот (ломаной кривой) или гистограммы (столбиками) (рис. 2).



Теперь данные можно представить графически, в виде полигона частот (ломаной кривой) или гистограммы (столбиками) (рис. 2).

Таблица 2

Классы	Центр классового интервала	Подсчет частот	Частоты, a
7–7.7	7.35	:	2
7.8–8.4	8.05	⊔	7
8.5–9.1	8.75	⊠⊓	18
9.2–9.8	9.45	⊠⊠:	22
9.9–10.5	10.15	⊠	10
10.6–11.2	10.85	.	1
11.3–11.9	11.55	:.	3
Сумма			63

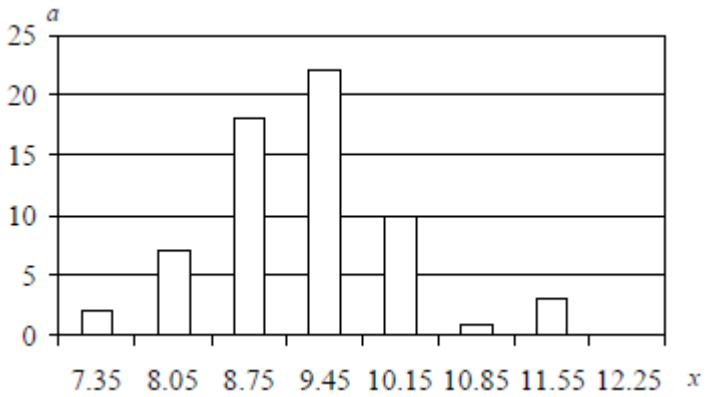


Рис. 2. *Распределение бурозубок по весу тела*

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

1. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
2. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
3. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

1. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
2. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 2

Тема: Вычисление параметров выборок.

Цель занятия: Сформировать у аспирантов общее представление об ошибках измерений и их математическом вычислении.

Задачи: Освоить методы вычисления средней арифметической величины и статистических ошибок измерений.

Аспирант должен знать:

1) до изучения темы:
   - Понятие о статистических методах обработки данных;
   - Общие правила вычислений;

2) после изучения темы:
   - Понятие о систематических, случайных и статистических ошибках измерения;
   - Способы вычисления статистических ошибок измерения.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Средняя арифметическая

Одной из важнейших обобщающих характеристик вариационного ряда является средняя величина признака (обычно обозначается буквой M). Существует несколько видов средних (средняя арифметическая – простая и взвешенная, средняя гармоническая, средняя квадратичная), но в практике биологических исследований наибольшее значение имеет средняя арифметическая – величина, вокруг которой «концентрируются» варианты.

Общая формула для определения величины средней арифметической – это отношение суммы значений всех вариант (x_i) выборки к их числу (объему выборки, n):

$$M = \frac{\sum x_i}{n}.$$

В нашем примере с определением массы бурозубок средняя величина равна M = 9.298412698 г. При расчетах статистических параметров на ЭВМ следует помнить, что большое количество значащих цифр обычно не имеет никакого биологического смысла. Записывая такие статистические параметры, как средняя и стандартное отклонение, следует оставлять в лучшем случае на одну значащую цифру больше, чем имели значения вариант, а оценки ошибок – на две значащих цифры. Масса тела бурозубок колебалась от 7.3 до 11.9 г, отсюда средняя с учетом округления должна иметь вид M = 9.3 г.

Средняя арифметическая характеризует действие только систематических факторов, поскольку сумма случайных отклонений влево и вправо от средней в силу симметричности нормального распределения обращается в нуль. Поэтому модель варианты меняется: xi = M ± хслуч.

В биологических исследованиях зачастую встречается ситуация, когда требуется первичная статистическая обработка большого числа выборок, но необязательно с большой точностью. Это может понадобиться для предварительного рассмотрения и оценки материала, в частности для оперативного выявления общих тенденций его изменчивости, с тем, чтобы в дальнейшем перейти к специальным методам статистического анализа. Для этих случаев предложен простой экспресс-метод с использованием полученного для данной выборки размаха значений (Lim). В случае нормального распределения средняя арифметическая находится точно по центру (совпадает со значением медианы), т. е. левая и правая границы

распределения находятся на одинаковом расстоянии от средней. Исходя из этих соображений, среднюю арифметическую можно рассчитать по формуле медианы:

$$M = Me = \frac{x_{\min} + x_{\max}}{2}.$$

Для бурозубок эта средняя составит M = (7.3 + 11.9) / 2 = 9.6 г, что вполне соответствует первой точной оценке.

В случаях, когда необходимо объединить результаты расчетов по нескольким выборкам и на этой основе найти общую среднюю, характеризующую весь изученный материал, пользуются взвешенной средней, которая учитывает объемы частных выборок:

$$M = \frac{\sum n_j \cdot M_j}{\sum n_j},$$

где *Mj* – значение частной средней,

nj – условные «веса» частного значения, объемы выборок.

Чтобы рассчитать среднюю взвешенную, необходимо значения всех частных средних арифметических помножить на свои «веса», все эти произведения сложить и сумму разделить на сумму весов (общий объем всех выборок). Пусть получены результаты определения средней величины выводка у рыжих полевок (экз. / самку) по месяцам: май 5.0, июнь 5.4, июль 6.2, август 6.0, сентябрь 4.5, причем известно число определений (самок) для каждого месяца: 22, 43, 103, 33 и 5. Взвешенная средняя арифметическая составит:

M = (5·22 + 5.4·43 + 6.2·103 + 6·33 + 4.5·5) / (22 + 43 + 103 + 33 + 5) = 5.8.

Средняя, рассчитанная обычным способом, оказалась заниженной:

M = (5 + 5.4 + 6.2 + 6 + 4.5) / 5 = 5.4.

В число прочих констант вариационного ряда входит медиана (Me) – значение, делящее размах выборки пополам, и мода (Mo) – класс (или значение), представленный наибольшим числом вариант.

Общее представление об ошибках измерений.

Различают ошибки систематические, случайные и промахи.

Систематические ошибки появляются постоянно при повторных измерениях и возникают, как правило, ввиду неисправности измерительных приборов или недостатков самого метода измерения. Систематические ошибки всегда односторонне влияют на результаты измерения, увеличивая или уменьшая их, и могут быть всегда устранимы путем совершенствования радиометрической аппаратуры.

Случайные ошибки возникают при изменении напряжения электротока в сети во время проведения радиометрических измерений, недостаточной чувствительности аппаратуры и т.п. Все это приводит к тому, что несколько измерений одной и той же величины дают различные результаты.

Исключить случайные ошибки, возникающие при измерении, нельзя, но, использую закон теории вероятностей, можно оценить их и сделать правильный вывод о достоверности полученных результатов.

Необходимо помнить, что увеличение числа измерений уменьшает влияние случайных ошибок. Если же случайная ошибка больше систематической, то именно случайную ошибку нужно уменьшить в первую очередь.

Промахи. Источником подобных ошибок является невнимательность исследователя. Для устранения промахов нужно соблюдать аккуратность, тщательность в работе и в записях результатов измерений. Многократное измерение одной и той же величины в одних и тех же условиях, как правило, позволяет обнаружить и устранить промахи.

Статистические ошибки.

Абсолютной ошибкой данного измерения называется разность между средним арифметическим значением и отдельным измерением. Абсолютная ошибка обозначается греческой буквой Δ.

Точность каждого значения данного ряда измерений характеризуется средней квадратической ошибкой, которую обозначают греческой буквой ς (сигма малая) и определяют по формуле:

$$\varsigma = \sqrt{\frac{\Sigma(a-M)^2}{n-1}},$$

где $\Sigma(a\text{-}M)^2$ – сумма квадратов отклонений всех измерений ряда от среднего арифметического значения;

n-1 – число членов вариационного ряда, уменьшенное на единицу.

Средний результат определения радиоактивности записывают в виде формулы: A=M±ς. Здесь средняя квадратическая ошибка (ς) показывает усредненную величину отклонения каждого измерения от своей средней арифметической.

В так называемых нормальных вариационных рядах весь размах изменчивости (варьирования), ограниченный максимальным и минимальным значением изучаемого показателя, заключает в себе шестикратную величину среднего квадратического отклонения. При этом максимальный вариант отличен от средней арифметической на значение +3ς, а минимальный вариант – на значение -3ς. Поэтому принято весь размах изменчивости выражать такой записью: M±3ς.

На основании этой особенности изменчивости в нормальных рядах можно осуществлять некоторые ориентировочные расчеты. Например, по величине средней арифметической и значению ς можно определить величину максимального и минимального членов в данной совокупности измерений.

Как будет показано в дальнейшем, по размаху изменчивости можно рассчитать приближенное значение среднего квадратического отклонения.

Среднее квадратическое отклонение (ς) – величина именованная, она выражается в тех же единицах, что и среднее арифметическое значение: кюри/км^2, мкюри/кг, пкюри/л и т.д.

Пример. Для определения ς воспользуемся величинами радиоактивности зернофуража по стронцию-90, приведенными в таблице:

Радиоактивность объекта, пКи/кг	a-M	$(a-M)^2$
41	4	16
38	1	1
35	-2	4
40	3	9
39	2	4
38	1	1
37	0	0
36	-1	1
32	-5	25
34	-3	9
Сумма 370		70
Среднее 37		

Подставив полученные величины в формулу $\varsigma=\sqrt{\frac{\Sigma(a-M)^2}{n-1}}$, определим среднюю квадратичную ошибку отдельного измерения данного вариационного ряда:

$$\varsigma=\sqrt{\frac{70}{9}}=\sqrt{7,77}=2,8\text{пк/кг}.$$

Итоговый результат определения радиоактивности объекта после проведенной математической обработки записывают в виде: A=M±ς= 37±2,8пк/кг.

Методы статистической обработки, требующие при определении усредненных ошибок возведения в квадрат каждого отклонения от средней арифметической величины, называются квадратическими. Эти методы при практическом применении, как видно, довольно громоздки.

При большом числе измерений количество арифметических действ достигает иногда десятков, что не редко ведет к так называемым ошибкам внимания.

Поэтому приходится пользоваться более простыми методами, которые хотя и менее точны, однако дают вполне удовлетворительные результаты (ошибка вычисления статистических показателей при этом обычно не превышает 5%).

Так, среднюю квадратическую ошибку (ς) можно определить по разности максимального и минимального значения измерений в соответствии с формулой:

$$\varsigma = \frac{a_{макс.} - a_{мин.}}{K_{\varsigma}},$$

где: K_{ς} – величина коэффициента, определяемого по таблице №2 приложения.

При большом количестве измерений (например, более сотни) коэффициент – «K_{ς}» можно брать равным 6, поскольку максимальный размах вариабильности ряда близок к M±3ς, т.е. укладывается в пределы 6ς.

Подставив в приведенную формулу данные из предшествующего примера, получим:

$$\varsigma = \frac{41 - 32}{3{,}08} \approx 3{,}0 \text{ пк/кг}.$$

Как видно, результаты достоверно близки, а бóльшая простата второго способа вычислений очевидна.

Процентное отношение величины средней квадратичной ошибки отдельного измерения к своей средней арифметической называется коэффициентом вариации. Он определяется по формуле:

$$K_{в} = \frac{\varsigma}{M} \times 100\%.$$

3) Практическая работа:

Решение предложенных задач.

Задание 1. Определить стандартную ошибку средней арифметической величины, исходя из данных, приведенных в таблице (концентрация стронция-90 в свекле, пк/кг):

a	a – M	$(a - M)^2$
13,3	-6,1	37,21
14,7	-4,7	22,09
16,2	-3,2	10,24
16,8	2,6	6,76
19,0	0,4	0,16
19,6	0,2	0,04
20,1	0,7	0,49
21,4	2,0	4,00
21,4	2,0	4,00
22,5	3,1	9,61
23,0	3,6	12,96
24,8	5,4	29,16

Задание 2. Вычислить среднюю арифметическую величину и ее стандартную ошибку следующего вариационного ряда (округление до десятков произведены для упрощения расчетов).

Варианты (a): 130, 130, 120, 140, 130, 120, 110, 110, 150, 160.
Отклонения (a-M): 0, 0, 10, 10, 0, 10, 20, 20, 20, 30.

Рекомендуемая литература:

Основная литература:

1. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
2. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
3. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

1. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
2. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 3

Тема: Статистическая оценка генеральных параметров.

Цель занятия: Сформировать у аспирантов представление о способах определения диапазона возможной изменчивости изучаемых биологических признаков..

Задачи: Освоить методику построения вариационных рядов.

Аспирант должен знать:

1) до изучения темы:
   - Понятие распределения величин;
   - Понятие генеральной совокупности;

2) после изучения темы:
   - Свойства нормального распределения;
   - Понятие доверительного интервала.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Приблизительный прогноз всегда можно дать в виде интервала между конкретными минимальными и максимальными значениями, в пределах которого будет находиться интересующая нас величина. Ясно, например, что рост очередного встречного взрослого человека вряд ли превысит два метра или будет меньше полутора метров. Более точный (вероятностный) прогноз можно дать, ориентируясь на распределение случайных величин. Распределение – это соотношение между значениями случайной величины и частотой их встречаемости. Как мы видели на примере веса тела землероек, числовые значения вариант располагаются в некоторой ограниченной зоне, в центре которой их особенно много, а по краям мало. Ключом к получению вероятностного прогноза служит знание законов распределения случайных величин. Очень большое число случайных величин, распространенных в природе, может быть описано с помощью закона нормального распределения, который задается уравнением:

$$p = \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2},$$

$$t = \frac{(x - M)^2}{S}$$

где – нормированное отклонение;

M, S – параметры нормального распределения.

Эта модель лежит в основе многих статистических методов.

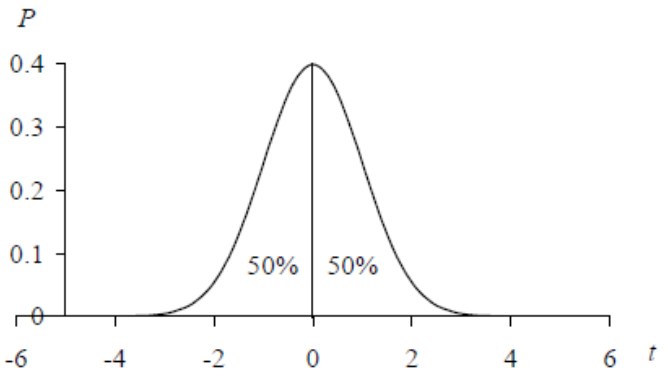
Свойства нормального распределения

Приведенное уравнение определяет ход кривой линии, имеющей характерную колоколообразную форму, и позволяет вычислить ординаты нормальной кривой, или «плотность вероятности» (p). Вероятность (статистическая, или частость) – численная мера возможного, определяется как отношение числа вариант (исходов испытаний) определенного вида к общему числу вариант (опытов). Поскольку нормальное распределение

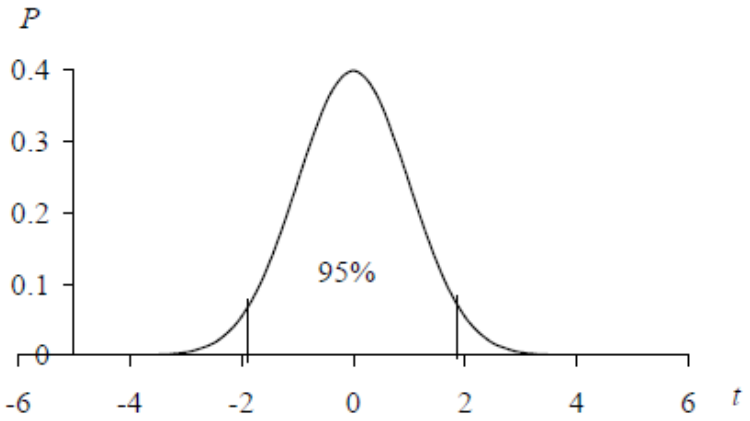
характерно для непрерывных случайных величин, говорят не о вероятности какого-то определенного значения варианты, но о «плотности вероятности», отражая тем самым плавность изменения вероятности значений для разных значений t, чем ближе к центру распределения, тем плотность вероятности выше.

С помощью представленного выше уравнения можно рассчитать вероятность появления нового значения случайной величины t в интервале той или иной ширины и дать статистическую оценку – найти интервал значений признака, в котором с той или иной вероятностью заключено значение генерального параметра. Формула количественно выражает вполне определенные свойства поведения случайной величины, из которых можно назвать следующие практически важные следствия:

1. Все варианты лежат в интервале плюс-минус бесконечность. Иными словами, с вероятностью P = 1 (P = 100%) мы вправе ожидать появление новой варианты в пределах от −∞ до +∞. Слева и справа от средней арифметической лежит по 50% вариант (свойство симметрии нормального распределения), т. е. с вероятностью P = 0.5 (50%) можно предсказать появление новой варианты в интервалах M − ∞ и M + ∞.



2. Между M − 1.96S и M + 1.96S лежит 95% вариант. Это позволяет с 95 %-ой вероятностью предполагать, что новая варианта окажется в интервале M ± 1.96S (округленно M ± 2S – так называемое правило двух стандартных отклонений).



3. С вероятностью P = 0.99 значение новой варианты будет заключено в пределах M ± 2.58S и с вероятностью P = 0.999 – в интервале M ± 3.3S.

Исходя из сказанного можно оценить вероятность появления новых значений признака. В отношении непрерывных случайных величин (метрических признаков) эта процедура сводится к так называемой интервальной оценке. Для полученных ранее характеристик, массы бурозубок, средней M = 9.26 и стандартного отклонения S = 0.79 (г), находим прогнозный интервал: M ± 1.96S = 9.26 ± 1.53. Новое значение признака с вероятностью P = 0.95 между 7.68 и 10.82 г. Предсказание веса землероек, конечно, не имеет большого практического значения. Гораздо важнее может быть прогноз численности ценных промысловых видов, сельскохозяйственных вредителей, вспышек болезней, урожая культурных растений и т. п.

Важнейшее значение для практического применения имеет «соглашение о 95%». В соответствии с ним совокупности, состоящей из 95% особей (объектов), мы доверяем так же, как и 100%-й. Термин «доверительная вероятность P = 0.95» означает, что, согласно принятому допущению, 95% вариант достаточно полно характеризуют изучаемое явление (в данном случае изменчивость веса землероек), что позволяет ограничиться рассмотрением вариант в области M ± 1.96S, охватывающей эту 95%-ю совокупность. Так, мы принимаем, что нормальный вес землероек данного вида может изменяться в пределах 7.7–10.8 г, не больше и не меньше. За этими пределами мы обнаруживаем животных иного вида или статуса.

При этом в биометрии обычно довольствуются доверительной вероятностью P = 0.95 (уровень значимости α = 0.05), хотя в наиболее ответственных исследованиях принимают и более строгие уровни – P = 0.99 и P = 0.999.

Однако это имеет смысл лишь при очень больших выборках исходных данных, точно описывающих закономерности изменчивости признаков. Обычно же выборки не очень велики, что позволяет ограничиться меньшей степенью доверительной вероятности P = 0.95.

Уровень значимости – понятие, альтернативное доверительной вероятности ($\alpha = 1 - P$). Для доверительной вероятности 0.95 уровень значимости составляет 0.05, а для 0.99 и 0.999 – соответственно 0.01 и 0.001. Уровень значимости, равный 0.05 (5%), можно интерпретировать так: имеется всего 5% шансов, что полученная величина не будет соответствовать изучаемой совокупности. Уровень значимости – это тот теоретический процент значений нормального распределения, который можно отбросить, не учитывать, дабы с меньшими усилиями получить основную информацию об изучаемом явлении. Можно целую жизнь положить на попытки отловить обыкновенную землеройку-бурозубку весом 2.5 г, но так и не собрать выборку, достаточную по объему, чтобы это реализовать (миллионы особей). Для практического понимания достаточно знать, что уровень значимости – это вероятность ожидаемой ошибки наших выводов, вероятность того, что данный статистический вывод не верен. И с этой позиции 5% – достаточно мало. Использование доверительной вероятности и уровня значимости можно назвать теоретической базой разумного ограничения времени и масштабов исследования, позволяющей получить достоверную общую информацию за счет исключения ничтожной доли частной.

Генеральная совокупность

Генеральная совокупность – все варианты одного типа. В предметной биологии это понятие можно интерпретировать как мыслимое множество вариант, сформированных при одинаковых (внешних и внутренних) условиях.

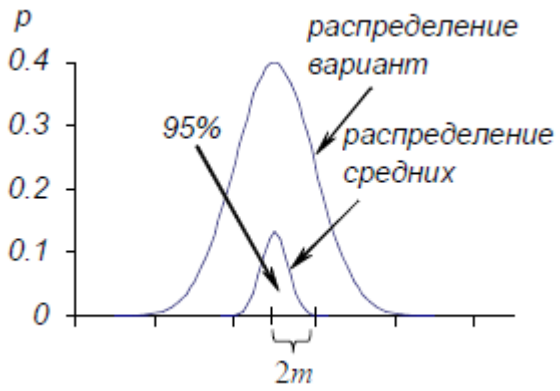
Теоретическая бесконечность генеральной совокупности означает, что ее никогда нельзя познать до конца, в действительности мы всегда имеем дело с выборками. Выборочная совокупность, выборка – это множество вариант одного типа, ограниченное способом отбора (методами получения вариант) из генеральной совокупности. Отличие выборок от генеральной совокупности состоит в том, что действующие в генеральной совокупности факторы не могут проявиться в полной мере в любой отдельной выборке. Каждая новая выборка обязательно будет отличаться от предыдущей в силу случайности, варианты новой выборки будут нести одинаковый отпечаток действия доминирующих факторов, но разные следы действия случайных факторов.

По этой причине параметры (средняя M и стандартное отклонение S) разных выборок из одной генеральной совокупности никогда не совпадут ни друг с другом, ни со значениями генеральных параметров (обычно обозначаемых буквами μ, σ), они будут немного отличаться, смещаясь относительно друг друга и варьируя вокруг генеральных значений.

Отличие генеральных параметров от их оценок по выборкам состоит еще и в том, что в первом случае они рассчитаны по всем вариантам, а во втором – по ограниченному их числу. Интуитивно понятно, что чем меньше объем выборок, тем менее точным будут выборочные оценки генеральных параметров, и, напротив, чем больше выборка, тем ближе выборочные средние и дисперсии лежат к генеральным значениям. Это явление называется законом больших чисел – с ростом числа наблюдений значения выборочных параметров стремятся воспроизвести генеральные.

Доверительный интервал

Параметры генеральной совокупности практически всегда остаются неизвестными, о них судят по выборочным оценкам, используя для этого значения ошибок репрезентативности. Теоретические исследования поведения выборочных средних (как случайных величин) показали, что они подчиняются нормальному закону, большинство из них (95%) находится поблизости от генеральной средней – в диапазоне Мген. ± 1.96m (приближенно ±2m). Это обстоятельство позволяет делать обратное заключение – генеральная средняя находится в диапазоне Мвыбор. ± 1.96m, т. е. предсказывать ширину интервала, в котором заключен генеральный параметр, давать интервальную оценку генеральному параметру.



В соответствии с законом нормального распределения можно ожидать, что генеральный параметр (истинное значение) окажется в интервале от M − tm до M + tm,

где m – ошибка средней арифметической,

t – квантиль распределения Стьюдента (табл. Прил.) при данном числе степеней свободы (df) и уровне значимости (обычно α = 0.05).

Сказанное можно перефразировать так: с вероятностью P можно ожидать, что генеральная средняя находится в доверительном интервале M ± tm, построенном вокруг выборочной средней арифметической M.

Доверительный интервал – интервал значений изучаемого признака, в котором с той или иной вероятностью P находится значение генерального параметра.

Возвращаясь к примеру о весе землероек-бурозубок, мы теперь можем записать доверительные интервалы при разных уровнях вероятности (граничные значения t взяты для случая n = ∞):

для P = 0.95 M ± tт = 9.3 ± 1.96 · 0.11 = 9.3 ± 0.21 г;

для P = 0.99 M ± tт = 9.3 ± 2.58 · 0.11 = 9.3 ± 0.28 г.

Здесь искомая генеральная средняя величина веса землероек с вероятностью P = 95% находится в пределах 9.11-9.53 г, а при P = 99% – 9.04–9.6 г.

Если объем выборки, для которой были получены параметры и ошибка репрезентативности m, был невелик (n < 50), то необходимо вводить поправки на объем выборки, расширяя область возможного пребывания генерального параметра. Это понятно, поскольку при дефиците информации любые заключения не могут быть очень точными. Так, для выборки объемом n = 20 экз. ошибка средней составит

$$m_M = \frac{0.89}{\sqrt{20}} = 0.19901 \text{ г},$$

а доверительный интервал M ± tт = 9.3 ± 2.09·0.2 = 9.3 ± 0.41 г – от 8.9 до 9.7 г (при уровне значимости α = 0.05 и числе степеней свободы df = n − 1 = 20 − 1 = 19 табличная величина статистики Стьюдента равна t = 2.09).

Аналогичным образом можно построить доверительный интервал для стандартного отклонения (S ± tmS), коэффициента вариации (CV ± tmCV), а также других статистических параметров (коэффициентов асимметрии, эксцесса, регрессии, корреляции).

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

1. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
2. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.

3. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

1. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
2. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 4

Тема: Определение точности опыта. Оптимальный объем выборки.

Цель занятия: Сформировать у аспирантов представление об оптимальном количестве экспериментов, достаточном для получения репрезентативных оценок.

Задачи: Освоить методику определения оптимального объема выборки при планировании экспериментов.

Аспирант должен знать:

1) до изучения темы:
   - Понятие выборки;
   - Понятие планирования эксперимента;

2) после изучения темы:
   - Способы определения точности опыта;
   - Способы установления необходимого объема выборки для проведения эксперимента.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Определение точности опыта

В практике биометрического анализа используется относительная ошибка измерений – «показатель точности опыта» – отношение ошибки средней к самой средней арифметической, выраженное в процентах:

$$\varepsilon = \frac{m}{M} \cdot 100\% .$$

Чем точнее определена средняя, тем меньше будет ε, и наоборот. Точность считается хорошей, если ε меньше 3%, и удовлетворительной при 3 < ε < 5%. Иначе приходится собирать дополнительный материал. В примере показатель точности составил ε = (0.11 / 9.3) · 100 = 1.2%, что говорит о достаточной надежности выборочной оценки.

Оптимальный объем выборки

В биологических исследованиях часто заранее требуется установить число наблюдений, достаточное для получения репрезентативных оценок генеральной совокупности.

Для непрерывных признаков метод состоит в том, чтобы, используя известные соотношения между средней, стандартным отклонением, ошибкой средней, плотностью вероятности распределения Стьюдента, найти число степеней свободы, соответствующее доверительному интервалу для средней при уровне значимости α = 0.05. Объем выборки, достаточной для получения результата заданной точности, находят по формуле:

$$n = \left(\frac{t \cdot CV}{\varepsilon} \right)^2 ,$$

где n – объем выборки,

t – граничное значение из таблицы распределения Стьюдента (табл. Прил.), соответствующее принятому уровню значимости при планируемом объеме выборки,

CV – приблизительное значение коэффициента вариации (%),

ε – планируемая точность оценки (погрешности) (%).

Рассчитаем необходимый объем условной выборки, обеспечивающий хорошую точность ε = 3%, для уровня значимости α = 0.05 (t = 1.98, для df ≈ 100) и для коэффициента вариации CV = 12% (такова относительная изменчивость многих размерно-весовых признаков животных):

$$n = \left(\frac{1.98 \cdot 12}{3} \right)^2 = 62.726 \approx 63 \text{ экз.}$$

Если исследуется фенотипическое (видовое) разнообразие (дискретный признак), может возникнуть задача определения минимального объема выборки, в которой будет присутствовать хотя бы один экземпляр с

определенным фенотипом (Животовский, 1991). С позиций теории вероятности задача ставится так: определить объем выборки, в которой с вероятностью P можно ожидать присутствие особи с признаком, частота которого в генеральной совокупности составляет π. Предлагается следующая формула:

$$N = \frac{\ln(1-P)}{\ln(1-\pi)}.$$

В первом приближении значение π можно определить приблизительно по имеющимся данным. Что же касается вероятности P, то ее уровень довольно сильно влияет на величину необходимого объема выборки. Для большей надежности следует брать P = 0.99, но тогда возрастет объем работ; не столь высокие требования (P = 0.95) могут и не позволить найти искомый фенотип. В частности, при уровне вероятности P = 0.95 и предположительной частоте фенотипа в популяции π = 0.05 потребуется

$$N = \frac{\ln(1-0.95)}{\ln(1-0.05)} = 58.4 \approx 59 \text{ экз.},$$

чтобы отловить хотя бы одну особь с этим дискретным признаком.

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

1. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
2. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
3. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

1. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
2. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 5

Тема: Оценка принадлежности варианты к выборке.

Цель занятия: Углубить знания аспирантов в области статистического анализа; сформировать понятие о «выскакивающих» значениях.

Задачи: Освоить методику определения нормированного отклонения.

Аспирант должен знать:

1) до изучения темы:
   - Понятие выборки;
   - Понятие варианты;

2) после изучения темы:
   - Методику оценки принадлежности вариант к той или иной совокупности.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Иногда встречается ситуация, когда одна из полученных вариант сильно отличается от остальных. Можно ли такие резко выделяющиеся значения использовать при дальнейших расчетах? В терминах математической статистики поставленный вопрос звучит так: относится ли данная варианта вместе с другими вариантами изучаемой выборки к одной и той же генеральной совокупности или к разным? Его можно сформулировать и по-другому: сформировано ли данное значение варианты под действием тех же доминирующих и случайных факторов, что и все остальные варианты данной выборки, или это были иные факторы? Здесь возможны два ответа.

1. Факторы те же, т. е. все варианты взяты из одной и той же генеральной совокупности.

2. Факторы иные, т. е. особенная варианта и выборка порознь взяты из разных генеральных совокупностей.

Ответ на этот вопрос можно получить с использованием рассмотренных выше свойств нормального распределения. Так, если все варианты были взяты из одной генеральной совокупности, значит, они должны отличаться друг от друга только в силу случайных причин и (с вероятностью P = 0.95) находиться в диапазоне M ± 2 · S. Иными словами, по случайным причинам варианты достаточно большой выборки будут отклоняться влево или вправо от средней не более чем на 2 · S: x−M < 2 · S или (x −M)/S < 2.

Эта величина, нормированное отклонение, и служит безразмерной характеристикой отклонения отдельной варианты от средней арифметической:

$$t = \frac{x - M}{S} \sim t_{табл.}$$

где t – критерий выпада (исключения),

x – выделяющееся значение признака,

M – средняя величина для группы вариант,

tтабл. – стандартные значения критерия выпадов, определяемые свойствами нормального распределения, их можно найти по табл. 5П для трех уровней вероятности (для больших выборок обычно пользуются значением tтабл. = 2 при P = 0.95, или α = 0.05).

Для вариант, принадлежащих к изучаемой достаточно большой выборке, нормированное отклонение меньше двух (с вероятностью P = 0.95): t < 2.

В случае действия на варианту некоего необычного фактора, она окажется за пределами указанного диапазона M ± 2S и ее нормированное отклонение будет равно или больше двух: t ≥ 2.

Нормированное отклонение есть простейший статистический критерий, который помогает определять так называемые «выскакивающие» варианты и решать вопрос о возможности их отбрасывания как артефактов (исключать из дальнейшей обработки). После такой «чистки» параметры выборки должны быть рассчитаны заново. К оценке чужеродности вариант, как и к другим методам статистики, нельзя подходить формально; цель биометрического исследования всегда состоит в том, чтобы понять специфику явления.

В частности, «отскакивающая» варианта может быть следствием того, что признак имеет иное, не-нормальное распределение.

Рассмотрим работу критерия на примере. При измерении длины черепа взрослых самцов обыкновенной землеройки-бурозубки получены выборки с такими параметрами: M = 18.8, S = 0.3 мм. Общее число животных n = 85.

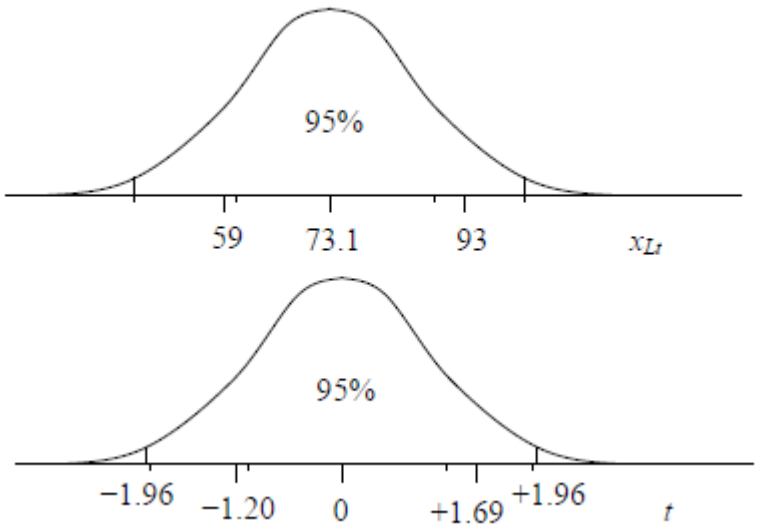
Среди прочих вариант два больших значения (19.2 и 21.0) вызывали сомнения. Определим для них критерии выпада:

$$t_1 = \frac{19.2 - 18.8}{0.3} = 1.3 < 2\,,\; t_2 = \frac{21.0 - 18.8}{0.3} = 7.3 > 2\,.$$

Согласно таблице (Прил.), критическое значение нормированного отклонения для уровня значимости α = 0.05 и n = 85 равно t = 2.0. Поскольку первое полученное значение (1.3) меньше табличного (2), первый из сомнительных результатов исключать не следует, а второй должен быть отброшен – критерий выпада (7.3) превышает табличное значение (2).

Понятие нормированного отклонения позволяет ввести важнейшее понятие статистики. Статистика – безразмерная случайная величина, которая имеет известный закон распределения и используется в качестве критерия для проверки статистических гипотез.

В этом смысле нормированное отклонение есть статистика. Во-первых, это безразмерная величина, поскольку единицы измерения числителя (xi −M) и знаменателя (S) взаимно уничтожаются. Во-вторых, нормированное отклонение имеет вполне определенное распределение (в случае непрерывных признаков – нормальное) со своими параметрами (рис.). Его средняя равна нулю Mt = tM = (M − M) / S = 0, а стандартное отклонение равно единице St = tS = (S − M) / S = (S − 0) / S = S / S = 1.



Переход от реального признака x к нормированному отклонению t

Нормированное отклонение – универсальная величина. Какой бы признак (имеющий нормальное распределение) мы ни брали, его значения можно выразить в виде расстояния от центра в единицах стандартного отклонения, т. е. на сколько S данное значение x отклонилось от M. При

этом, как следует из свойств нормального распределения, крайние значения в 95% случаев не будут принимать значения меньше −2 и больше 2.

С помощью нормированного отклонения можно, например, оценивать отличия разнокачественных объектов (пород и сортов, видов, популяций, генераций и пр.), причем даже по разным признакам.

Нормированное отклонение можно использовать и для сравнительной оценки разных индивидов по одному и тому же признаку. Например, если сопоставляемые по относительному весу сердца молодая и взрослая землеройки-бурозубки демонстрируют одинаковые показатели (10.5 мг%), то это, тем не менее, не означает их сходства по изучаемому признаку. Используя известную информацию (у молодых средний индекс сердца равен M = 10.0 при стандартном отклонении S = 1.3, у взрослых – M = 11.8, S =1.1), рассчитаем нормированное отклонение для молодого зверька

$$t_1 = \frac{10.5 - 10}{1.3} = 0.3$$

и для взрослого –

$$t_2 = \frac{10.5 - 11.8}{1.1} = -1.2 .$$

Налицо существенное различие: взрослый зверек имеет относительно низкий показатель сердечного индекса, а молодой близок по этому признаку к видовой норме.

Наибольшее развитие такой подход получает в процедурах обработки многомерных данных, при исследовании объектов, охарактеризованных по многим признакам, методом корреляций, главных компонент, при их кластеризации и т. п. Во многих случаях обработка многомерного массива начинается с нормирования данных по формуле нормированного отклонения.

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

1. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
2. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
3. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

1. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
2. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 6

Тема: Частная и ранговая корреляция.

Цель занятия: Углубить знания аспирантов в области корреляционного анализа.

Задачи: Освоить методику определения коэффициентов частной, ранговой корреляции.

Аспирант должен знать:

1) до изучения темы:
   - Понятие корреляционного анализа;

2) после изучения темы:
   - Виды корреляции признаков.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Коэффициент частной корреляции позволяет оценить связь между первым и вторым признаками при постоянных значениях третьего и вычисляемый по формуле:

$$r_{A(BC)} = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{(1 - r_{AC}^2) \cdot (1 - r_{BC}^2)}},$$

где A и B – факторы, связь которых требуется изучить;
C – фактор, влияние которого необходимо исключить из корреляционной зависимости между A и B (реперный признак);
r_{AB}, r_{AC}, r_{BC} – соответствующие парные коэффициенты корреляции, вычисляемые обычным способом;
$r_{A(BC)}$ – искомый коэффициент частной корреляции, показывающий связь между двумя признаками при исключении влияния третьего.

Этот же метод можно применить и для элиминации двух факторов при четырех переменных и т. д. Формула для расчетов примет в этом случае следующий вид:

$$r_{AB(BD)} = \frac{r_{AB(C)} - r_{AC(B)} \cdot r_{BC(D)}}{\sqrt{(1 - r^2_{AC(D)}) \cdot (1 - r^2_{BC(D)})}}.$$

Рассмотрим нахождение коэффициента частной корреляции на упрощенном примере. Получены данные о корреляции между давлением крови (A), содержанием в ней холестерина (B) и возрастом (C) у 142 женщин. Соответствующие коэффициенты корреляции таковы: r_{AB} = +0.25; r_{AC} = +0.33; r_{BC} = 0.51.

Известно, что повышенное артериальное давление может быть связано с высоким содержанием холестерина в стенках кровеносных сосудов, однако и давление крови, и концентрации холестерина увеличиваются с возрастом.

Поэтому возникает вопрос, создается ли корреляция между давлением крови и содержанием в ней холестерина за счет их общей связи с возрастом или же она реально существует для каждого возраста (и независимо от него). Элиминируя эффект возраста по приведенной выше формуле, получим:

$$r_{A(BC)} = \frac{0.25 - 0.33 \cdot 0.51}{\sqrt{(1 - 0.33^2) \cdot (1 - 0.5^2)}} = 0.12.$$

По таблице (Прил.) можно установить, что при n = 150 для достоверности коэффициента корреляции даже при уровне значимости α = 0.05 его величина должна быть не меньше 0.159. В данном же случае полученное значение меньше табличного и, следовательно, коэффициент корреляции от нуля достоверно не отличается. Таким образом, внутри отдельных возрастных групп корреляционной связи между давлением крови и содержанием холестерина, по крайней мере на изученном материале, не обнаруживается. Пока нет оснований отбрасывать нулевую гипотезу.

Второй пример демонстрирует использование коэффициента частной корреляции для более глубокого проникновения в структуру нескольких факторов наведения. Рассмотрим выборку объектов разного статуса (11

видов мелких млекопитающих), взяв в качестве признаков их численность в семи биотопах прибайкальской равнины. Реперным признаком послужила суммарная численность вида во всех биотопах. Здесь коэффициент корреляции отражает сходство между биотопами по соотношениям численности 11 видов. Например, оказалось, что между березняком и экотоном (граница между березняком и коренными лесами) и общая корреляция (r = 0.92), и частная (r = 0.64) высока и положительна. Можно утверждать, что население животных этих биотопов почти идентично.

В свою очередь, корреляция между кедровником и лугом не проявилась (r = −0.08), но коэффициент частной корреляции был велик и отрицателен (r = −0.43). Этим оттеняется тот факт, что виды, отсутствующие на лугу, многочисленны в кедровнике (красная полевка, мышь), а обычные в агроценозе – крайне редки в тайге (серые полевки).

Частная корреляция показала, что население этих биотопов во многом диаметрально противоположно. Она выявила два вида факторов наведения.

Один из них хорошо известен – это сезонное расселение видов в другие биотопы. В течение периода размножения видовой состав тайги и луга меняется несогласованно (одни виды идут из тайги в агроценозы, другие – в противоположном направлении) и численность всех видов относительно выравнивается, r = −0.08. Частная корреляция устраняет эффект прироста численности за счет иммигрантов и выдвигает на первый план контраст «базовой» численности, которую формируют характерные обитатели биотопов: в тайге это лесные полевки, на лугу – серые. Так проявляется второй фактор: отличие качества среды в разных биотопах. Он обеспечивает формирование принципиально несходных зооценозов, что и показывает высокой частной корреляцией r = −0.43.

Ранговая корреляция

Помимо рассмотренных выше параметрических показателей связи в биометрии применяются и непараметрические. Обычно их используют при сильных отклонениях изучаемого распределения от нормального (или сомнениях на этот счет), а также в тех случаях, когда требуется оценить зависимость между качественными или полуколичественными признаками, точное количественное измерение которых затруднено (оценки в баллах или других условных единицах). Если варианты выборки могут быть упорядочены по степени выраженности их свойств, для измерения степени сопряженности между ними можно воспользоваться непараметрическим показателем связи – ранговым коэффициентом корреляции Спирмена:

$$r_S = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)},$$

где d – разность между рангами сопряженных значений признаков x и y;

n – объем выборки.

Этой формулой следует пользоваться в тех случаях, когда выборки не содержат повторяющихся вариант, когда все ранги выражены разными целыми числами. Если же исходные ряды содержат одинаковые значения, расчет корреляции приходится вести по другой формуле, включающей поправку на повторы (при этом одинаковым вариантам присваивается средний ранг):

$$r_S = \frac{\frac{(n^3 - n)}{6} - (T_x + T_y) - \sum d^2}{\sqrt{\left(\frac{(n^3 - n)}{6} - 2 \cdot T_x\right)\left(\frac{(n^3 - n)}{6} - 2 \cdot T_y\right)}},$$

где Tx, Ty – поправки на серии повторов для каждой выборки:

$$T_x = \frac{\sum^{k} (t_x^{\,3} - t_x)}{12},$$

где t – число членов в каждой группе одинаковых вариант.

Поправки Tx, Ty учитывают k групп повторяющихся вариант.

Рассмотрим технику вычислений на примере изучения связи между оцененными в баллах численностью лисицы (x) и обилием мышевидных грызунов (y) (по годам наблюдений):

	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966
x	2.6	2.1	2.3	2.3	1.6	2.2	3.0	2.1	1.5	2.2
y	3.0	2.4	3.6	2.9	3.7	3.3	4.0	2.1	1.0	3.5

Чтобы проверить наличие и определить силу этой связи, нужно упорядочить значения сопряженных признаков по степени их выраженности, затем присвоить им ранги, обозначив значения порядковыми числами натурального ряда, и рассчитать коэффициент корреляции. Техника вычислений показана в таблице:

Численность лисицы в баллах, x	Обилие грызунов в баллах, y	Ранги вариант		Разность между рангами, d	d^2
		R_x	R_y		
1.5	1.0	1	1	0	0
1.6	3.7	2	6	−4.0	16.00
2.1	2.4	3.5	3	+0.5	0.25
2.1	2.1	3.5	2	+1.5	2.25
2.2	3.3	5.5	7	−1.5	2.25
2.2	3.6	5.5	8.5	−3.0	9.00
2.3	3.6	7.5	8.5	−1.0	1.00
2.3	2.9	7.5	4	+3.5	12.25
2.6	3.0	9	5	+4.0	16.00
3.0	4.0	10	10	0	0
					Σ = 59

В ряду значений признака x есть три пары одинаковых вариант, поэтому поправка будет равна:

$$T_x = \frac{(2^3-2)+(2^3-2)+(2^3-2)}{12} = 1.5.$$

В ряду признака y всего одна пара одинаковых значений; поправка составит:

$$T_y = \frac{(2^3-2)}{12} = 0.5.$$

Находим величину – $\frac{(n^3-n)}{6} = \frac{(10^3-10)}{6} = 165.$

Коэффициент ранговой корреляции составит:

$$r_S = \frac{165-(1.5+0.5)-59}{\sqrt{(165-2\cdot1.5)(165-2\cdot0.5)}} = 0.638.$$

Если воспользоваться формулой без поправок, результат будет несколько иным:

$$r_S = 1-\frac{6\cdot\sum d^2}{n\cdot(n^2-1)} = 1-\frac{6\cdot59}{10\cdot(10^2-1)} = 0.642.$$

Статистическая ошибка и критерий достоверности отличия коэффициента корреляции от нуля вычисляются по формулам:

$$m_r = \sqrt{\frac{1-r_S^2}{n-2}} = \sqrt{\frac{1-0.638^2}{10-2}} = 0.272,$$

$$t_r = r_S / m_r = 0.638 / 0.272 = 2.34.$$

Величина критерия (2.34) несколько выше критического значения (2.31) для уровня значимости α = 0.05 и числа степеней свободы df = n − 2 = 8 (табл. Прил). Казалось бы, это дает основание отвергнуть нулевую гипотезу (rS = 0) и с вероятностью P = 95% констатировать достоверность

установленной связи. Однако при небольших выборках статистические свойства коэффициента Спирмена не очень «хороши» и для оценки значимости корреляции лучше воспользоваться специальной таблицей (16 Прил.).

Чтобы полученный коэффициент можно было считать достоверно отличным от нуля, он должен превышать табличное значение при данном n.

В нашем случае (n = 10, α = 0.05) коэффициент r = 0.638 ниже табличного r = 0.64, следовательно, значимо от нуля не отличается. Зависимость численности лисицы и грызунов по приведенным данным достоверно не прослеживается.

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

4. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
5. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
6. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

3. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
4. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов

по дисциплине Б1.В.ДВ.1 Экспериментальная экология

к практическому занятию № 7

Тема: Определение достоверности изменения величин.

Цель занятия: Сформировать у аспирантов общее представление о достоверности различий между показателями.

Задачи: Освоить методы математической обработки цифровых данных.

Аспирант должен знать:

1) до изучения темы:

– Понятие об ошибках измерений;
– Общие правила вычислений;

2) после изучения темы:

– Методы определения степени достоверности между изучаемыми величинами.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Определение достоверности разности между средними арифметическими двух вариационных рядов.

Для сравнения уровней радиоактивной загрязненности продуктов сельского хозяйства в разных зонах в пределах одного промежутка времени или в одном районе (области) в разные годы требуется определить не только разницу средних арифметических двух рядов измерений, но и установить достоверность этого различия, т.е. подтвердить его неслучайный характер. Методика оценки достоверности разности средних арифметических двух статистических рядов изложена в следующем примере.

Предположим, что при анализе проб молока из пяти хозяйств было установлено, что радиоактивность этого продукта по стронцию-90 в 1976г. в данной местности была равна $M_1 \pm m_1 = 14.5 \pm 1,8$ пКи/л и в 1977г. – $M_2 \pm m_2 = 10 \pm 1,4$ пКи/л (всего было проведено десять радиохимических анализов). Снижение радиоактивности молока за год составило 14,5-10=4,5 пк/л, т.е. радиоактивность уменьшилась примерно на одну треть. Насколько достоверно это уменьшение радиоактивности? Ответить на это можно только после определения коэффициента достоверности (t – критерий Стьюдента) по формуле:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

Подставив вместо буквенных обозначений соответствующие цифровые данные, получим: $t = \frac{14,5 - 10}{\sqrt{3,24 + 1,96}} = \frac{4,5}{\sqrt{5,2}} = \frac{4,5}{2,28} = 1,97$.

Сравнивая полученный результат с табличным значением критерия Стьюдента для десяти измерений по двум строкам (число степеней свободы f= $n_1+n_2-2=5+5-2=8$) видим, что наш результат (1,97) меньше указанного в таблице №5 (см. приложение) не только для высокой (99%), но и для обычной достоверности (95%), ибо табличные коэффициенты соответственно равны 3,36 и 2,31. Поэтому следует сделать вывод об отсутствии достоверной разницы в уровнях радиоактивности молока между 1976 и 1977гг. Разница считается действительной, если ее достоверность равна или больше 95% (т.е. P≥95% или P≥0,95).

Таблицей №5 пользуются также при определении уровня достоверности отдельно взятой средней арифметической. Например, если средняя арифметическая величина и ее стандартная ошибка равны 11±2, а число вариант 5, то, определив критерий достоверности $t=\frac{11}{2}=5.5$ и количество степеней свободы (f=5-1=4), находим по таблице, что уровень достоверности средней арифметической более 99% (но меньше 99,9%), ибо соответствующие табличные значения критерия Стьюдента при указанном количестве степеней свободы равны 4,60 и 8,61, а найденное нами значение t – критерия занимает промежуточное положение.

По формуле: $t=\frac{M_1-M_2}{\sqrt{m_1^2+m_2^2}}$ достоверность различий между средними арифметическими определяют лишь в том случае, если варианты одной совокупности изменяются независимо от вариант другой совокупности. Если же между вариантами обоих совокупностей имеется взаимосвязь (сопряженность), то достоверность различий между рядами целесообразно вычислять методом прямой разности, который заключается в следующем. Варианты сравниваемых совокупностей (a_1 и a_2) выписывают в рядом расположенные строки (вертикальные или горизонтальные) и в каждой паре вариант определяют разность (a_1-a_2). Затем суммируют полученные разности и находят их среднее арифметическое значение. После этого по размаху варьирования вычисляют стандартную ошибку этого среднего арифметического, делят на нее значение самой средней арифметической и получают величину t – критерия, характеризующую достоверность различий между сравниваемыми рядами.

Пример. Содержание цезия-137 в пробах свеклы, отобранных с одних и тех же участков в семи контрольных пунктах области в 1975 и 1976 гг., характеризуется данными, приведенными в таблице (графы 2 и 3).

Номера контрольных пунктов	Уровни радиоактивности свеклы (пКи/кг)		Разность (a_1-a_2)
	1975г (a_1)	1976г (a_2)	
1	19,5	16,9	2,6
2	20,2	16,2	4,0
3	22,9	19,4	3,5
4	27,7	25,6	2,1
5	30,5	29,3	1,2
6	35,6	31,4	4,2
7	31,9	28,5	3,4
Сумма	188,3	167,3	21,0
Среднее	26,9	23,9	3,0

Поскольку пробы свеклы в 1975 и 1976гг. брали на одних и тех же участках, можно предположить наличие зависимости между значениями из обоих сравниваемых рядов. Это дает основание провести вычисление достоверности разности между средними арифметическими этих рядов методом «сопряженных пар». Для этого среднюю арифметическую разностей (a_1-a_2) делят на ее стандартную ошибку, определяемую по формуле:

$$m = \frac{a_{макс.} - a_{мин.}}{n} = \frac{4{,}2 - 1{,}2}{7} = 0{,}43 .$$

В итоге получаем величину t – критерия: $\frac{3{,}0}{0{,}43} = 7{,}0$.

В данном случае число степеней свободы f=n-1=7-1=6.

При этом значении числа степеней свободы в таблице №5 для уровня достоверности 99,9% имеет величину 5,96, а наша величина t – критерия, равно 7, значительно превосходит табличный показатель. Следовательно, различие между рядами является высокодостоверным. Оно свидетельствует о существенном снижении радиоактивности исследуемого объекта в 1976г. по сравнению с 1975г.

Если бы в данном примере применялся обычный критерий (т.е. не учитывающий сопряженность пар), то различие не было бы обнаружено, ибо большой разброс величин радиоактивности в пределах каждого ряда из-за резкого различия в характере почв привел бы к завышенным значениям стандартной ошибки средних арифметических и, следовательно, к заниженной величине t – критерия.

Для подтверждения этого положения расчетами, находим величину стандартных ошибок m_1 и m_2 по размаху варьирования:

$$m_1 = \frac{35{,}6 - 19{,}5}{7} = \frac{16.1}{7} = 2.3$$

$$m_2 = \frac{31{,}4 - 16{,}2}{7} = \frac{15{,}2}{7} = 2{,}17$$

и определим t – критерий разности между средними арифметическими по формуле:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{26{,}9 - 23{,}9}{\sqrt{2{,}3^2 + 2{,}17^2}} = \frac{3}{\sqrt{9{,}9989}} = \frac{3}{3{,}16} = 0{,}95,$$

что значительно ниже табличного 95-процентного уровня достоверности (при $f=n_1+n_2-2=7+7-2=12$ в таблице №5 указана величина t – критерия, равная 2,18), несмотря на то, что в этом случае используют вдвое большее число степеней свободы.

Поэтому методом «сопряженных пар» пользуются только тогда, когда взаимосвязь пар несомненна (в этом случае влияние уменьшения числа степеней свободы перекрывается влиянием уменьшением стандартной ошибки разности) и в случаях одинакового количества вариант в сравниваемых рядах.

Сравнивание двух альтернативных распределений.

Для оценки разницы между признаками используют критерий различия «хи-квадрат» (χ^2). Его можно применять для изучения многих биологических процессов (при анализе влияния на организм различных факторов).

Данные наблюдений (опытов) группируют в таблицы, состоящие из нескольких полей. Наиболее простой является четырехпольная таблица. Значение χ^2 в этом случае определяют по формуле:

$$X^2 = \frac{\left(\left[p_1 \times p_4 - p_2 \times p_3\right] - 1/2n\right)^2 \times n}{(p_1 + p_2) \times (p_3 + p_4) \times (p_1 + p_3) \times (p_2 + p_4)},$$

где p – частоты (их размещение показано в нижеследующей таблице, в которой в качестве примера приведены результаты эксперимента по профилактике лучевой болезни у овец с помощью химического радиопротектора);

n – общее число наблюдений.

Прямые скобки, в которых находятся $p_1 \times p_4 - p_2 \times p_3$, показывают, что берется только абсолютное, т.е. положительное численное значение этой разности.

Пример. При проверке эффективности противолучевого препарата при введении его подкожно до облучения были получены следующие данные:

Группа овец	Исход лучевой болезни		Сумма
	Выздоровело животных	Пало животных	
Леченные	p_1=25	p_2=23	p_1+p_2=48
Контрольные	p_3=7	p_4=29	p_3+p_4=36
Сумма	p_1+p_3=32	p_2+p_4=52	n=84

Для выяснения значимости действия радиопротектора использована четырехпольная таблица. В ней буквами p_1 и p_2 обозначены частоты для двух верхних полей таблицы, а буквами p_3 и p_4 – для двух нижних. В самой нижней строке таблицы – суммы частот по вертикальным столбцам (p_1+p_3) и (p_2+p_4), в правом крайнем столбце таблицы – суммы частот по горизонтальным строчкам (p_1+p_2) и (p_3+p_4). Общая сумма $n= p_1+p_2+ p_3+p_4$. В нашем примере n=84 овцам.

Подставив в вышеприведенную формулу соответствующие числовые значения и произведя необходимые вычисления, получим величину χ^2:

$$X^2 = \frac{\left([25\times 29 - 23\times 7] - 84/2\right)^2 \times 84}{48\times 36\times 32\times 52} = \frac{22888656{,}0}{2875392{,}0} \approx 8{,}0.$$

По таблице критических значений X^2 (таблица №6 приложения) для уровня достоверности 95% и 99% находим граничные величины этого критерия, применительно к нашему случаю. Они соответственно равны 3,85 и 6,63 (при одной степени свободы для четырехпольной таблицы). Поскольку полученная нами величина χ^2=8,0 намного превосходит оба табличных значения, то можно считать эффективность противолучевого препарата значимой (P>0,99).

Доверительный интервал.

Зная среднюю арифметическую величину (M) ряда измерений и ее стандартную ошибку (m), можно установить с определенной точностью те границы, в которых находится истинная средняя арифметическая ($M_{ист.}$). Для этого используют формулу:

$$M-tm \le M_{ист.} \le M+tm,$$

где t – критерий по Стьюденту.

Эти границы получили название доверительных; интервал, т.е. разница между максимумом и минимумом, также называемые доверительным.

Заранее устанавливают тот или иной уровень достоверности (Р), при котором желают получить доверительные границы для истинного значения средней арифметической величины ($M_{ист.}$).

В интервале М±м истинная величина среднего арифметического значения содержится в 68% случаев. Это имеет следующий смысл: пусть взято 100 выборок по «n» измерений каждая и, следовательно, получен 100 доверительных интервалов. Все они будут несколько различаться между собой как величиной средних арифметических, так величиной стандартных ошибок, но только 68 из этих интервалов покроют $M_{ист.}$ (или, иными словами, 68 этих интервалов будут содержать $M_{ист.}$).

Однако такая вероятность (68%) недостаточна для надежного определения истинного значения средней арифметической величины. В биологических применениях статистики достаточно надежным считается 95%-ный доверительный интервал.

Пример. Определяли концентрацию стронция-90 в молоке (в пКи/л). При девяти измерениях был получен результат А=М±м=31±2. По таблице находим значение t – критерия для уровня достоверности Р=0,95. При числе степеней свободы f=9-1=8 величина t – критерия равна 2,31. Доверительный интервал определяем по вышеприведенной формуле: $31-2x2{,}31 \leq M_{ист.} \leq 31+2x2{,}31$ или $26{,}38 \leq M_{ист.} \leq 35{,}62$. Таким образом, вероятность того, что истинное значение средней арифметической радиоактивности молока находится в пределах от 26,38 до 35,62 пКи/л, равна или больше 95%. Результат записывают так: $M_{ист.}=31\pm4{,}62$ пКи/л ($P \geq 0{,}95$ или $P \leq 0{,}05$).

Обычно используют следующие три уровня достоверности: Р=0,95 (95%), 0,99 (99%) и 0,999 (99,9%), что соответствует уровням значимости Р=0,05 (5%), 0,01 (1%) и 0,001 (0,1%).

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

7. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
8. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
9. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

5. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
6. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов

по дисциплине Б1.В.ДВ.1 Экспериментальная экология

к практическому занятию № 8

Тема: Регрессионный анализ.

Цель занятия: Сформировать у аспирантов представление о зависимости между показателями и ее количественной оценке.

Задачи: Освоить методы регрессионного анализа.

Аспирант должен знать:

1) до изучения темы:

- Понятие о статистических методах обработки данных;
- Понятие об оценке достоверности различия между показателями;

2) после изучения темы:

- Сущность регрессионного анализа.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Регрессионный анализ (метод наименьших квадратов).

Коэффициент корреляции указывает лишь на степень связи в вариации двух попеременных величин, но не дает возможности судить о том, как количественно меняется одна величина по мере изменения другой. Этот вопрос решают с помощью регрессии. При простой корреляции изучают зависимость между изменчивостью двух признаков. С помощью регрессии ставят задачу установить количественную взаимосвязь между признаками. Одним из приемов решений этой задачи при наличии прямолинейной зависимости является метод наименьших квадратов.

Зависимость между величинами в этом случае выражается уравнением регрессии, которое имеет вид обычного уравнения прямой линии:

$$У=аХ+в,$$

где У и Х – коррелирующие величины;

а – коэффициент пропорциональности, который показывает степень зависимости У от Х;

в – первоначальное значение У при Х=0.

Для того, чтобы определить численное значение «а» и «в» в уравнении У=аХ+в, надо решить систему двух уравнений:

1) $У=ахХ+вхn$
2) $ХУ=ахХ^2+вхХ$

Составление этих уравнений основано на методе наименьших квадратов, который позволяет вычислить такие параметры для уравнений регрессии, при которых сумма квадратов отклонений эмпирических значений «У» от критически вычисленных окажется наименьшей.

Пример. В результате проведенных испытаний ядерного оружия загрязненность почвы по стронцию-90 постепенно возросла до 7,8 мКи/км^2. По мере увеличения радиоактивности почвы повышалось содержание стронция-90 в фураже и продуктах животноводства, в частности в молоке. Динамика этого процесса представлена в таблице (Х – радиоактивность почвы в десятках мКи/км^2, У – радиоактивность молока в десятках пКи/л):

$Х^2$	Х	У	ХУ
1	*2*	*3*	*4*
0,81	0,9	0,43	0,4
0,81	0,9	0,42	0,4
1,7	1,3	0,51	0,7
5,3	2,3	0,87	2,0
9,0	3,0	1,06	3,2
12,2	3,5	1,17	4,1
19,4	4,4	1,37	6,0
17,6	4,2	1,37	5,7
23,0	4,8	1,44	6,9
23,0	4,8	1,47	7,0
1	*2*	*3*	*4*

30,2	5,5	1,54	8,5
28,1	5,3	1,50	7,9
42,2	6,5	1,76	11,4
42,2	6,5	1,62	10,5
51,8	7,2	1,93	13,9
49,0	7,0	1,82	12,7
64,0	8,0	1,94	15,5
44,9	6,7	2,48	16,6
60,8	7,8	2,79	21,8
X^2=540	X=90	У=27,5	ХУ=155

В левом и правом столбцах таблицы, а так же в нижней строке, приведены данные, полученные в результате обработке двух средних рядов. После подстановки итоговых значений этой таблицы в уравнения 1 и 2 они приобретут следующий вид:

1) 27,5=90а+19б
2) 155=540а+90б

Для решения их обычными алгебраическими методами надо умножить коэффициенты уравнения 1 на 6 и вычесть уравнение 2 из уравнения 1.

165=540а+114б
155=540а+90б
10=24б

Отсюда «б» равно $\frac{10}{24}=0{,}42$. После подстановки значения «б» в уравнение 1 получим значение «а»: 27,5=90а+8, отсюда, а=0,22.

В окончательном виде уравнение регрессии будет следующим:

У=0,22Х+0,42

Если подставить в это уравнение различные значения уровней загрязненности почвы, можно получить соответствующие этим уровням количества стронция-90, содержащегося в молоке. Предположим, что загрязнение почвы составляет 30 (три десятка) пКи остронция-90 на 1км^2. Тогда содержание радиоизотопа в молоке будет:

У=0,22х3+0,42=1,08 (т.е. 10,8 пКи/л).

Расхождение этой величины с приведенным выше в таблице значением не превышает нескольких процентов: $\frac{1{,}08-1{,}06}{1{,}08}\times 100\%=1{,}85\%$, что подтверждает приемлемость пользования полученным уравнением для определения количественных значений одной величины при измерении другой.

Определение количественной зависимости между варьирующими величинами методом наименьших квадратов может быть использовано для ориентировочного прогнозирования радиоактивной загрязненности продукции сельского хозяйства на следе радиоактивного облака.

Пример. После глобального выпадения радиоактивных веществ загрязненность почвы по стронцию-90 составила 50 (5 десятков) мКи/км^2. Требуется определить возможное содержание этого нуклида в молоке у коров на данной территории.

Пользуясь уже полученными нами уравнением регрессии, находим величину радиоактивной загрязненности молока:

У=0,22х5+0,42=1,1+0,42=1,52 (т.е. 15,2 пКи/л).

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

10. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
11. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
12. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

7. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
8. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов

по дисциплине Б1.В.ДВ.1 Экспериментальная экология

к практическому занятию № 9

Тема: Анализ временных рядов.

Цель занятия: Углубить знания аспирантов в области статистического анализа результатов исследований.

Задачи: Получить представление об особенностях статистического анализа, связанного с прогнозированием.

Аспирант должен знать:

1) до изучения темы:

- Понятие регрессионного анализа;

2) после изучения темы:

- Методы сглаживания.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Существуют две основные цели анализа временных рядов:

1. Определение природы ряда. Определение закономерностей, которые можно выделить посредством исследования графика.

2. Прогнозирование. Предсказание будущих значений временного ряда по настоящим и прошлым значениям.

Обе эти цели требуют, чтобы модель ряда была идентифицирована и, более или менее, формально описана. Как только модель определена, вы можете с ее помощью интерпретировать рассматриваемые данные (например, использовать в вашей теории для понимания сезонного изменения цен на товары, если занимаетесь экономикой или прогнозировать вылов рыбы если занимаетесь исследованиями продуктивности рыбных рек). Не обращая внимания на глубину понимания и справедливость теории, вы можете экстраполировать затем ряд на основе найденной модели, т.е. предсказать его будущие значения (прогнозирование). Но при исследовании сложных систем здесь возникает проблема адекватности прогнозной модели. Практика исследования сложных систем говорит нам, что мы не можем построить абсолютно адекватную модель поведения сложных систем, а, следовательно, не можем абсолютно достоверную модель будущего состояния системы. Единственно достоверным методом прогнозирования на настоящий момент остаётся только паттерн-анализ (по мнению В.В. Налимова), который основан на выделении устойчивых повторяющихся сочетаний (паттернов), которые впоследствии можно использовать в качестве индикаторов процесса.

Как и большинство других видов анализа, анализ временных рядов предполагает, что данные содержат систематическую составляющую

(обычно включающую несколько компонент) и случайный шум (ошибку), который затрудняет обнаружение регулярных компонент. Большинство методов исследования временных рядов включает различные способы фильтрации шума, позволяющие увидеть регулярную составляющую более отчетливо.

Большинство регулярных составляющих временных рядов принадлежит к двум классам: они являются либо трендом, либо сезонной составляющей. После исключения из временного ряда этих двух компонент, остаётся стационарный временной ряд или же не остаётся ничего, тогда выяснятся, что ряд целиком состоит из тренда или сезонной составляющей. Для выявления периодичности временного ряда используются автокорреляционные функции, ряд Фурье и другие сложные методы.

Тренд представляет собой общую систематическую линейную или нелинейную компоненту, которая может изменяться во времени. Сезонная составляющая – это периодически повторяющаяся компонента. Оба эти вида регулярных компонент часто присутствуют в ряде одновременно. Например, численность популяции может возрастать из года в год, но она также содержат сезонную составляющую (как правило, существует период особой активности – брачный период). Любой ряд динамики разделён на три компоненты:

$$x(t) = f(t) + g(t) + h,$$

где f(t) – детерминированная (определяемая) компонента, представляющая аналитическую функцию, которая выражает тенденцию в ряду динамики; g(t) – стохастическая (вероятностная) компонента, моделирующая периодический характер вариаций исследуемого явления; h – случайная компонента типа «белый шум», т.е. необъяснённые факторы или, так называемые, флуктуации.

Отметим также некоторые особенности временных рядов. Биометрические данные часто имеют пропуски наблюдений, для восстановления которых используются различные алгоритмы. Как правило, пропущенный участок получают путём осреднения значений соседних интервалов или с помощью более сложных алгоритмов. Другая особенность временных рядов это – выбросы. Под выбросами обычно понимают наблюдения, являющиеся в том или ином смысле аномальными (на графике они выражаются через резкие пики или падения значений, причём зачастую единичные). Такие случаи анализируются и исключаются из общего рассмотрения при создании тренда. Также интересны разрывы. Разрыв

временного ряда – это скачкообразное изменения уровня временного ряда, т.е. выброс в ряду значений. Очевидно, что к идентификации выбросов и разрывов в экологических рядах следует подходить с особой осторожностью, чтобы не потерять значимые данные, т.к. они могут характеризовать некий периодический или системный процесс.

Не существует «автоматического» способа обнаружения тренда во временном ряде. Однако если тренд является монотонным (устойчиво возрастает или устойчиво убывает), то анализировать такой ряд обычно нетрудно. Если временные ряды содержат значительную ошибку, то первым шагом выделения тренда является сглаживание. Как правило, сглаживание подразумевает изменение масштаба для выявления более общей тенденции.

Сглаживание всегда включает некоторый способ локального усреднения данных, при котором несистематические компоненты взаимно погашают друг друга. Самый общий метод сглаживания – скользящее среднее, в котором каждый член ряда заменяется простым или взвешенным средним n соседних членов. Вместо среднего можно использовать медиану значений, попавших в окно значений. Основное преимущество медианного сглаживания, в сравнении со сглаживанием скользящим средним, состоит в том, что результаты становятся более устойчивыми к выбросам (имеющимся внутри окна).

Если в данных имеются выбросы (связанные, например, с ошибками измерений), то сглаживание медианой обычно приводит к более гладким или, по крайней мере, более «надежным» кривым, по сравнению со скользящим средним с тем же самым окном. Основной недостаток медианного сглаживания в том, что при отсутствии явных выбросов, он приводит к более «зубчатым» кривым (чем сглаживание скользящим средним) и не позволяет использовать веса. Также используется взвешенное сглаживание. В данном случа определяются взвешенные средние, взятые с разных точек ряда динамики.

Метод экспоненциального сглаживания (метод Брауна) применяется для нестационарных временных рядов.

Для целей прогнозирования используются сходные методы. Например, частым методом прогнозирования является метод скользящих средних:

$$m_t = \frac{1}{n}\sum_{i=t}^{t-n+1} d_i \ ,$$

т.е. метод основан на составлении нового ряда из простых средних арифметических, которые были вычислены для предыдущих промежутков.

Аналогично применяются и другие методы сглаживания (взвешенное, медианное, экспоненциальное).

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

13. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
14. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
15. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

9. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
10. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.

Методические указания для аспирантов
по дисциплине Б1.В.ДВ.1 Экспериментальная экология
к практическому занятию № 10

Тема: Общая схема статистического анализа.

Цель занятия: Обобщить и систематизировать знания аспирантов в области статистического анализа результатов исследований.

Задачи: Получить представление о порядке статистической обработки результатов исследований.

Аспирант должен знать:

1) до изучения темы:
   - Понятие статистического анализа;

2) после изучения темы:
   - Порядок и методы обработки цифровых данных.

Изучение темы занятия направлено на формирование профессиональных компетенций: готовность к исследованию современных явлений и тенденций в биосфере, к изучению структурных элементов экосистем, закономерностей формирования системы связей на биогеоценотическом, ландшафтном и природно-зональном уровнях (ПК-2).

Методика проведения занятия.

1. Определение темы занятия. Преподаватель поясняет цели и задачи занятия, значение полученных знаний для будущей работы по специальности.

2. **Теоретическая часть.**

Математическая обработка результатов исследований имеет целью выявить направленность наблюдаемых сдвигов, оценить их достоверность и определить общие закономерности их изменения.

При оценке результатов измерений необходима их предварительная подготовка с целью ускорения и упрощения дальнейших расчетов. Для этого необходимо выполнить следующие условия:

- результаты измерений расположить в возрастающем или убывающем порядке, т.е. произвести их ранжирование;
- цифровые данные, если они выражены в виде десятичных дробей, привести к целым числам путем умножения на какую-то избранную постоянную величину (10, 100 и т.п.);
- исключить заведомо сомнительные значения исходных данных (как правило, это крайние варианты).

Статистический анализ материала начинают с вычислений средней арифметической величины, описывающей одним числом результаты ряда измерений. Однако, вычисление одного этого статистического показателя, характеризующего ряд наблюдений через их среднее значение, недостаточно для описания полученных результатов измерений в связи с тем, что не учитывается отклонение отдельных членов ряда от средней арифметической величины и самой средней от истинного значения радиоактивности объекта. В связи с этим возникает необходимость оценить погрешность полученных величин и самой средней арифметической, для чего определяют среднее квадратическое отклонение (ς) и стандартную ошибку средней арифметической величины (m), позволяющие судить о степени рассеивания (разброса) наблюдаемых значений.

Наиболее просто значение среднего квадратического отклонения отдельных измерений и стандартной ошибки средней арифметической величины определяют по размаху варьирования, т.е. по разности значений

крайних вариант в каждом ранжированном ряду (эту разность делят соответственно на коэффициенты Кς или Km).

Одна из задач статистической обработки – оценка достоверности различий средних значений – решается путем вычисления t-критерия по Стьюденту. Разницу между показателями считают достоверной при уровне значимости $P \geq 95\%$.

В тех случаях, когда необходимо установить связь между исследуемыми признаками и оценить ее степень, применяется корреляционный анализ. Количественным выражением степени корреляции является коэффициент корреляции (коэффициент по абсолютной величине не может быть больше единицы).

Применение регрессионного анализа (метод наименьших квадратов) позволяет количественно оценить биологическую закономерность, связывающую признаки между собой и их влияние друг на друга.

Применение критерия $\chi 2$ (хи-квадрат) дает возможность сравнить эмпирические распределения качественных признаков и установить достоверность различий между ними.

Заключительным этапом исследования является обобщение и анализ полученных результатов измерений. От того, насколько всесторонне и глубоко выполнена эта работа с учетом материалов математической обработки цифровых данных, зависит объективность оценок, правомерность и убедительность выводов и рекомендаций.

3) Практическая работа:

Решение предложенных задач.

Рекомендуемая литература:

Основная литература:

16. Ивантер, Э. В., Коросов, А. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
17. Лакин Г. Ф. Биометрия: Учеб. пособие для биол. спец. вузов-4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.
18. Плохинский Н. А. Биометрия. М.: Изд-во МГУ, 1970.

Дополнительная литература:

11. Фишер Р. Статистические методы для исследователей. М.: Госстатиздат, 1958.
12. Коросов А. В. Экологические приложения компонентного анализа. Петрозаводск, 1996.