Министерство образования и науки РФ
ФГАОУ ВПО «Казанский (Приволжский) федеральный университет»
Институт экологии и природопользования
кафедра почвоведения

Составитель
К.Г. Гиниятуллин

Статистическая обработка результатов научных исследований Краткий конспект лекций

Казань-2014

Направление: 021900.62 «Почвоведение» (направление), профиль – не предусмотрен.

Учебный план –не предусмотрен, очное 2013 г.,

Дисциплина: «Статистическая обработка результатов научных исследований» (бакалавриат, 4 курс, 7-8 семестры, очное обучение)

Количество часов: 72 ч. (в том числе: лекции – 16, практические занятия – 16, самостоятельная работа – 40), форма контроля: зачет.

Аннотация: *Дисциплина знакомит студентов с методами статистической обработки результатов почвенных исследований, полевых опытов. Рассматриваются способы статистически обоснованного представления научных результатов в почвоведении. При прохождении курса студенты должны приобрести навыки анализа, статистической обработки собственных научных результатов, получаемых при выполнении курсового и дипломного проектов, с применением компьютерной техники и пакетов статистических программ.*

.Темы курса:

1. Предмет и задачи курса. Закон нормального распределения. Оценка соответствия выборки закону нормального распределения. Критерий кси-квадрат. Критерий Колмогорова-Смирнова. Критерия Шапиро-Уилка.

2. Характеристика выборки. Дисперсия. Стандартное отклонение. Коэффициент вариации. Ошибка опыта.

3. Параметрические методы проверки статистических гипотез. Использование критерия Стьюдента. Парный двухвыборочный t-тест. Анализ независимых выборок. Использование критерия Фишера.

4. Непараметрические методы проверки статистических гипотез. Использование критериев Вилкоксона и Манн-Уитни.

5. Обработка результатов исследований в электронных таблицах MS Excel. Графическое представление результатов исследований..

6. Анализ однофакторных дисперсионных комплексов

7. Анализ многофакторных дисперсионных комплексов

8. Использование корреляционного анализа для исследования зависимостей
9. Использование регрессионного анализа для исследования зависимостей

Составитель курса: Гиниятуллин Камиль Гашикович, доцент кафедры почвоведения, тел.: 89178912874, email: ginijatullin@mail.ru

URL электронного курса в MOODLE:

http://tulpar.kfu.ru/course/view.php?id=2304

Дата начала эксплуатации: 1 сентября 2015 года

Оглавление

Тема 1. Предмет и задачи курса. Закон нормального распределения. Оценка соответствия выборки закону нормального распределения. Критерий кси-квадрат. Критерий Колмогорова-Смирнова. Критерий Шапиро-Уилка.

АННОТАЦИЯ: Тема рассматривает цели и задачи математической статистики. Даются основные понятия математической статистики и теории вероятностей. Рассматриваются проблемы использования различных критериев для оценки нормальности распределения.

Ключевые слова. генеральная совокупность, выборка, критерий кси-квадрат, Критерий Шапиро-Уилка, Критерий Колмогорова-Смирнова.

Методические рекомендации по изучению темы

- К теме составлен конспект лекции. В данной лекции дается характеристика методов, задач и методологические основы изучаемой дисциплины. Также в лекции рассматривается закон нормального распределения и критерии позволяющие оценить соответствие выборок нормальному распределению.
- В качестве самостоятельной работы предлагается написание эссе по следующим темам (на выбор студента):

1. Роль статистики в почвоведении.
2. Закон нормального распределения Гаусса-Лапласа.
3. Условия применения критерия кси-квадрат для проверки нормальности распределения.
4. Условия применения критерия Шапиро-Уилка для проверки нормальности распределения.

- Оформить отчет в виде файла эссе и прикрепить этот файл к заданию
- Для проверки усвоения темы имеются вопросы и темы для самоконтроля знаний.

- Для проверки усвоения темы пройти тест (Тестирование является завершающей стадией освоения темы).

Рекомендуемые информационные ресурсы (при составлении курса лекций использовалась следующая литература и электронные ресурсы):

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб. пособие для студентов вузов / В.Е. Гмурман.—12-е изд., перераб..—Москва: Высш. образование, 2007.—478 с.

Попов, Владимир Александрович. Руководство к решению задач по теории вероятностей и математической статистике / В. А. Попов, М. Х. Бренерман; Казан. гос. ун-т, Физ. фак..?Казань: Изд-во Казанского государственного университета, 2008. 117, [2] с.: ил.; 20.?Библиогр.: с. 118 (9 назв.), 200.

Салимов, Фарид Ибрагимович. Основы статистической обработки: учебное пособие для студентов очной и заочной формы обучения бюджетного и договорного отделения КГУ / Салимов Ф. И..?Казань: Казанский государственный университет, 2010.108 с.: ил.; 21.?Библиогр.: с. 6 (11 назв.), 520.

Решение задач корреляционного и регрессионного анализа в электронных таблицах MS EXCEL: методическое пособие к практическим занятиям дисциплины "Математические модели в почвоведении" / Казан. гос. ун-т, Биол.-почв. фак.; [сост. к.б.н., доц. К. Г. Гиниятуллин]. Казань: Изд-во Казанского государственного университета, 2008. 31,[1] с.: ил

Теория вероятностей и математическая статистика: Учеб. пособие / Под ред. В.И. Ермакова. - М.: ИНФРА-М, 2004. - 287 с. http://znanium.com/bookread.php?book=76845

Теория вероятностей и математическая статистика: Учебное пособие / С.В. Павлов. - М.: ИЦ РИОР: ИНФРА-М, 2010. - 186 с. http://znanium.com/bookread.php?book=217167

Теория вероятностей и математическая статистика /Б.А. Горлач. - М.: "Лань", 2013. - 320 стр. http://e.lanbook.com/view/book/4864/

Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов..—4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351 с.

Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов..—М.: Изд. МГУ, 1970.—367с.

Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.

Дмитриев Е.А. Математическая статистика в почвоведении.М.: МГУ, 1995.291с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww

Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

Выборка – выборочная совокупность.

χ^2 критерий кси-квадрат

As – коэффициент асимметрии

Es – коэффициент эксцесса

Глоссарий

Генеральная совокупность, генеральная выборка (от лат. generis — общий, родовой) (в англ. терминологии — population) — *совокупность всех объектов (единиц), относительно которых учёный намерен делать выводы при изучении конкретной проблемы.*

Выборочная совокупность — *часть генеральной совокупности элементов, которая охватывается наблюдением.*

Коэффицие́нт эксце́сса (коэффициент островершинности) в теории вероятностей — *мера остроты пика распределения случайной величины*

Асимметри́я] (от др.-греч. ασυμμετρία «несоразмерность», от др.-греч. μετρέω — «измеряю») — *отсутствие или нарушение симметрии.*

Гистогра́мма в математической статистике - это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.

Вопросы для изучения по теме:

Цели и задачи математической статистики.

Основы теории вероятностей .

Понятия генеральная совокупность и выборочная совокупность.

Репрезентативность выборки и рандомизация.

Виды изменчивости.

Распределение частот и его графическое изображение.

Количественная непрерывная изменчивость, закон нормального распределения

Асимметри́я и эксце́сс

Проверка гипотезы о нормальности распределения с помощью критерия кси-квадрат

Проверка нормальности распределения с помощью критерия Шапиро-Уилка

Проверка нормальности распределения с помощью критерия Колмогорова-Смирнова

Конспект лекций:

Предмет и задачи курса. Определение понятия «математическая статистика.

Математическая статистика *— это один из разделов математики, который позволяет делать умозаключения о всей (генеральной) совокупности на основе наблюдений над выборочной совокупностью, или выборкой.*

Математическая статистика *помогает в выборе оптимальных условий для проведения опыта, дает объективную, количественную оценку экспериментальным данным, но ее результаты не могут иметь самостоятельной научной ценности.*

Математическая статистика основана ***на теории вероятностей.***

Теория вероятностей как научная основа математической статистики. Понятие вероятность. Вероятность случайного события.

Теория вероятностей *— наука, изучающая общие закономерности в массовых случайных явлениях различной природы. Случайным называется явление, исход которого в настоящий момент нельзя точно предсказать.*

Всякое множественное явление (совокупность) представляет собой совокупность особей, случаев, фактов, предметов, т. е. некоторых условных единиц, каждая из которых в отдельности строго индивидуальна и отличается от других рядом признаков. Свойство условных единиц отличаться друг от друга даже в однородных совокупностях называется изменчивостью, или варьированием. Изменчивость - свойство, присущее всем природным объектам.

Вероятность—одно из основных понятий теории вероятностей.

Вероятностью события А называют отношение числа благоприятствующих этому событию исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу. Вероятность события А определяется формулой

$P(A)=m/n$,

m число элементарных исходов благоприятствующих событию А

n - число всевозможных элементарных событий

Из определения вероятности вытекают следующие ее свойства:

Свойство 1. Вероятность достоверного события равна единице.

Действительно, если событие достоверно, то каждый элементарный исход испытания благоприятствует событию. В этом случае $m=n$, *следовательно,*

$P(A)=m/n=1$.

Свойство 2. Вероятность невозможного события равна нулю.

В этом случае $m=0$, *следовательно,*

$P(A) = m/n = 0/n = 0$.

Свойство 3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

$0 < P(A) < 1$.

В этом диапазоне вероятностей могут использоваться методы математической статистики.

Статистическая совокупность. Генеральная совокупность и выборка.

Изменчивость, варьирование признаков создает известную трудность в тех случаях, когда требуется дать общую характеристику определенной варьирующей группе (совокупности) по отдельным признакам или сравнить две такие группы. Обычно изучение всей совокупности невозможно, поэтому ограничиваются изучением ее части, по которой делают общее заключение. Такой метод называется выборочным. Изучаемую группу объектов, называют *генеральной совокупностью, а ту часть объектов, которая попала на проверку, исследование, измерение выборочной совокупностью или выборкой.*

Главная цель выборочного метода — по статистическим показателям малой выборки дать характеристику всей совокупности объектов, т.е. генеральной совокупности.

Распределение вероятностей для дискретных и непрерывных случайных величин. Виды изменчивости.

Различают два типа изменчивости: количественную, которая может быть измерена, и качественную, которая не поддается измерению.

Качественная изменчивость. Альтернативная и безальтернативная качественная изменчивость.

Качественной изменчивостью называется такое варьирование, когда различия между вариантами выражаются качественными показателями, которые одни варианты имеют, а другие нет. Качественная изменчивость может быть альтернативная и безальтернативная качественная изменчивость. Если признак принимает только два взаимоисключающих друг друга значения, то изменчивость называется альтернативной, т. е. двояковозможной.

Количественная изменчивость.

Под количественной изменчивостью понимают такую изменчивость, в которой различия между вариантами выражаются количеством.

Дискретная и непрерывная количественная изменчивость.

Если различия между вариантами выражаются целыми числами, между которыми нет переходов такая количественная изменчивость называется дискретной. При непрерывной изменчивости значения вариант выражаются мерами объема, длины, массы и т. д., которые могут принимать любые числовые значения, ограничиваемые точностью измерений.

Характеристика непрерывной количественной изменчивости.

Возможные значения варьирующего признака X называют *в а р и а н т а м и* *и обозначают $X_x, X_2, \ldots X_n$. Ряд варьирующих величин можно упорядочить — расположить значения признака в порядке их возрастания (или убывания). Такое упорядочение ряда, т. е. расположение вариант в порядке возрастания (или убывания), называется* *р а н ж и р о в а н и е м* .

Построение вариационных рядов и их графическое представление.

Для первичной систематизации необходимо сгруппировать значения X1, $X_2, \ldots, X_n$ в k групп с интервалом каждой группы i. Ориентировочно число групп равно корню квадратному из объема выборки, которое, однако, не должно быть меньше 5 и больше 20.

Величину интервала групп определяют по соотношению:

$$i = \frac{X\max - X\min}{\textit{число.групп}} = \frac{R}{k}$$

где, R-размах варьирования

Тогда для каждого интервала можно вычислить частоту – количество значений X попадающих в данный интервал

Графическое изображение вариационного ряда называется кривой распределения или *в а р и а ц и о н н о й   к р и в о й* *. Числа, которые характеризуют, сколько раз повторяется каждое значение признака у членов данной совокупности, называются частотами признака. Сумма всех частот (f) равна объему выборки, т. е. числу членов ряда — n. В результате такой обработки первичных наблюдений получаем так называемый* *в а р и а ц и о н н ы й* *ряд.*

Репрезентативность выборки.

Чтобы выборка правильно отражала свойства объекта, она должна быть представительна, репрезентативна. Лучшим способом обеспечения *репрезентативности выборки считается рандомизация т.е. случайность отбора компонентов физической совокупности на испытание.*

Эмпирические и теоретические распределения.

Эмпирическое распределение — распределение результатов измерений, полученных при изучении выборки (экпериментальные данные). На основе теоретических распределений строятся статистические критерии, которые используются для проверки статистических гипотез. Наиболее часто в исследовательской работе опираются на нормальное распределение (распределение аппроксимируемое уравнением Гаусса-Лапласа) или специальные распределения, получаемые из нормального для определенно поставленной задачи и при ограниченном числе степеней свободы (t, F χ^2 - распределение, распределение Пуассона).

Нормальное распределение и его особенности. Закон нормального распределения. *Функция f(x), связывающая значения Xi переменной случайной величины X с их вероятностями P(x), называется законом распределения этой величины, или законно Гаусса-Лапласа. Вероятность P любого значения Xi непрерывно распределяющейся случайной величины X выражается формулой:*

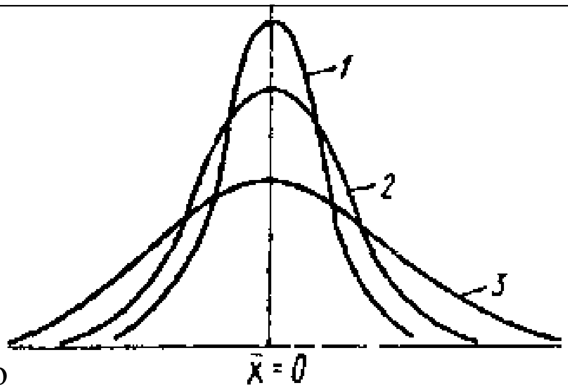
$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2} * \frac{(x-\mu)^2}{\sigma}}$$

P(X - вероятность; μ — генеральная средняя (математическое ожидание); σ — стандартное отклонение генеральной совокупности (при n → ∞); π и e — константы. Положение и форма кривой нормального распределения полностью определяются двумя параметрами: генеральной средней μ, которая находится в центре распределения, и стандартным отклонением σ которое - измеряет вариацию отдельных наблюдений около *средней. Максимум, или центр, нормального распределения лежит в точке $\overline{X} = \mu$; точки перегиба кривой находятся при $X = \mu - \sigma$ и $X_2 = \mu + \sigma$. При $X \to \pm\infty$ кривая достигает нулевого значения.*

Центральная тенденция нормального распределения отражает соответствие максимальной частоты вероятности среднему выборочному значению. Форма кривой нормального распределения соответствует правилу трех сигм. Она сводится к тому, что размах колебаний от μ вправо и влево зависит от величины σ и укладывается в основном в пределах трех стандартных отклонений. В области $\mu \pm \sigma$ лежит 68,26% (почти две трети) всех наблюдений; внутри пределов $\mu \pm 2\sigma$ находится 95,46% всех значений случайной величины; интервал $\mu \pm 3\sigma$ охватывает 99,73%, следовательно, практически все значения.

Проверка нормальности распределения. Асимметрия и эксцессивность и их характеристика.

Кривая зависимости вероятности случайной величины соответствующая закону нормального распределения имеет определенную форму (рис.1). При увеличении σ кривая расширяется (рис.1., $\sigma_1 < \sigma_2 < \sigma_3$)



ко

Рис. 1. Зависимость кривой нормального распределения от величины дисперсии

Для оценки несоответствия вариационной кривой можно использовать показатель характеризующий ее асимметричность (рис.2) или островершинность (выположенность) - коэффициент экцсесса (рис.3)

Рис.2. Положительная и отрицательная асимметрия

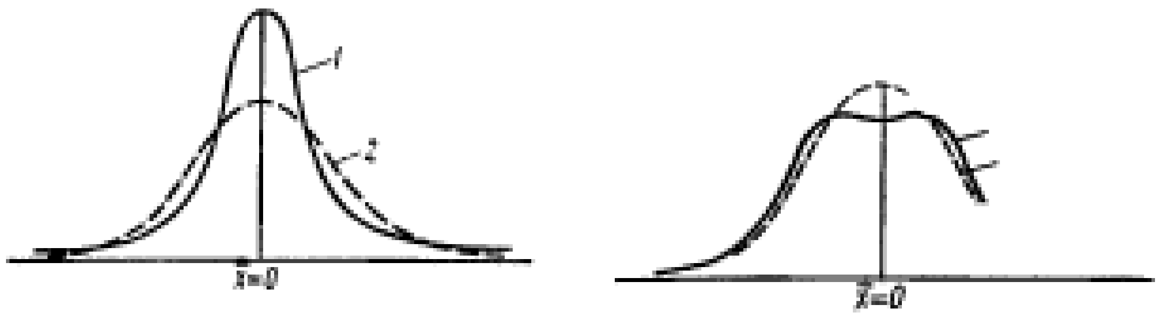


Рис.3. Положительный и отрицательный экцсесс

Применение коэффициентов характеризующих наличие асимметрии и экцсесса требует выборок с большим объемом данных, поэтому проверку гипотез о законах распределения обычно производят с помощью специально выработанных критериев.

Критерий Кси-квадрат (χ^2). Расчет критерия кси-квадрат. Применение критерия кси-квадрат для оценки нормальности распределения. Ограничения применения критерия кси-квадрат.

χ2-критерий представляет собой сумму квадратов отклонений эмпирических частот f от вычисленных или ожидаемых частот отнесенную к теоретическим частотам, т. е.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f - f')^2}{f'}$$

f – эмпирическая частота признака в интервале, f '`- теоретическая частота признака в интервале, k- число степеней свободы

Для того чтобы оценки были более точными, выборка, распределяемая в вариационный ряд, должна содержать не менее 50 вариант. Считается, что применение критерия χ2 требует, чтобы в крайних классах вариационного ряда содержалось не менее пяти вариант.

Число степеней свободы критерия рассчитывают по формуле k=N-3 (с учетом трех ограничений свободы вариации: n, s_x^2 *и* $\overline{X}$ *). Где N – общее количество интервалов.*

Нулевая гипотеза сводится к предположению, что различия, наблюдаемые между эмпирическими и вычисленными или ожидаемыми частотами, носят исключительно случайный характер. Для проверки нулевой гипотезы фактически полученную величину χ2 сравнивают с ее критическим значением $\chi 2_{st}$. *Если* $\chi 2 > \chi 2_{st}$. *то нулевая гипотеза должна быть отвергнута на принятом уровне значимости с числом степеней свободы k. Критические точки* $\chi 2_{st}$ *приведены в специальной таблице.*

Критерий Шапиро-Уилка. Расчет критерия Шапиро-Уилка. Применение критерия Шапиро-Уилка для оценки нормальности распределения. Ограничения применения критерия Шапиро-Уилка.

Для проверки нормальности распределения, когда объем выборки $3 < n < 50$, *можно использовать критерий Уилка-Шапиро. В современных пакетах статистических программ можно использовать данный критерий для оценки нормальности выборок с объемом больше 50. Применимость этого критерия ограничена условием несгруппированности исходных данных. Данный критерий разработан специально для оценки соответствия выборок закону нормального распределения и характеризуется как наиболее объективный критерий.*

Процедуру проверки нулевой гипотезы, сводящейся к предположению, что выборка принадлежит нормально распределенной величине, начинают с построения ранжированного ряда значений $(i = 1, 2, ..., n)$ *от наименьшего* $(i_{min}$

= 1) до наибольшего (i_{max} = n). По выборочным данным вычисляют сумму квадратов центральных отклонений С

$$C2 = \sum (Xi - \bar{X})^2$$

Далее находят величину k согласно следующим правилам: если n - нечетное, то $k = \frac{n-1}{2}$ *если _ n _ четное* $- k = \frac{n}{2}$,

Затем вычисляют вспомогательную величину В по формуле:

$$B = \sum_{i=1}^{i=k} a_{n-i+1}\left(x_{n-i+1} - x_i\right) = a_n\left(x_n - x_1\right) + a_{n-1}\left(x_{n-1} - x_2\right) + \\ + \ldots + a_{n-k+1}\left(x_{n-k+1} - x_k\right), \quad (8.2)$$

где a_{n-i+1} – некоторые коэффициенты, значения которых в зависимости от n для $i = 1, 2, ..., k$ приведены в специальной таблице (см.

В формуле сомножители, стоящие в скобках, представляют собой разности между значениями, расположенными симметрично относительно концов ранжированного ряда.

Затем, вычисляют величину W по формуле,

$$W = \frac{B^2}{C} ,$$

Значение Wслужит статистикой доя проверки гипотезы о нормальности распределения. Если W< Wst то с уровнем значимости а распределение считается отличным от нормального. Если W превышает критические значения Wst, то распределение допустимо рассматривать как нормальное. Критические значения Wst находят в зависимости от α (уровня значимости) и n (объема выборки) из специальных таблиц.

Критерий Колмогорова-Смирнова. Применение критерия Колмогорова-Смирнова. для оценки нормальности распределения. Ограничения применения критерия Колмогорова-Смирнова.

Критерий Колмогорова-Смирнова является непараметрическим критерием позволяющим оценить соответствие эмпирического распределения любому теоретическому распределению. Критерий Колмогорова-Смирнова можно также использовать для оценки выборки на соответствие закону нормального распределения. Критерий Колмогорова-Смирнова и критерий Шапиро-Уилка основы на различных принципах и характеризуются различной мощностью, поэтому выводы о соответствии выборки нормальному закону, по данным критериям могут не совпадать.

Вопросы для самоконтроля знаний

- Что такое центральная тенденция при нормальном законе распределения?
- Может ли кривая нормального рапределения быть асимметричной?
- Чем выражается эксцесс в распределении?
- Задачи курса
- Признаки и их классификация.
- Статистическая совокупность.
- Генеральная совокупность и выборка.
- Получение выборки.
- Принцип построения вариационного ряда.
- Репрезентативность и рендомизация.
- Представление распределений
- Частота и вероятность.
- Вариационная кривая.
- Особенности нормального распределения.
- Формула Гаусса-Лапласа

Тема 2. Характеристика выборки. Дисперсия. Стандартное отклонение. Коэффициент вариации. Ошибка опыта

Аннотация. Данная тема дает представление о параметрических и непараметрических методах характеристики выборки.

Ключевые слова. Дисперсия, Стандартное отклонение, Коэффициент вариации, Мода, Медиана: Квантили.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где дается общее представление об параметрических и непараметрических показателях выборки.
- В качестве самостоятельной работы студентам предлагается ответить на вопросы самоконтроля:
- Для темы разработаны вспомогательные материалы (глоссарий, список сокращений и др.)
- Освоение темы предполагает выполнение задания к практическому занятию. Рекомендации для выполнения самостоятельной работы представлены внутри задания.
- Оформить отчет в виде файла результатов выполнения самостоятельного задания и прикрепить этот файл к заданию
- Для проверки усвоения темы имеются вопросы для самоконтроля знаний.
- Для проверки усвоения темы необходимо пройти тест (Тестирование является завершающей стадией освоения темы).

Рекомендуемые информационные ресурсы (при составлении курса лекций использовалась следующая литература и электронные ресурсы):

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб. пособие для студентов вузов / В.Е. Гмурман.—12-е изд., перераб..—Москва: Высш. образование, 2007.—478, [1] с.

Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов..—4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351[1]с.

Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов..—М.: Изд. МГУ, 1970.—367с.

Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.

Теория вероятностей и математическая статистика: Учеб. пособие / Под ред. В.И. Ермакова. - М.: ИНФРА-М, 2004. - 287 с. http://znanium.com/bookread.php?book=76845

Теория вероятностей и математическая статистика /Б.А. Горлач. - М.: "Лань", 2013. - 320 стр. http://e.lanbook.com/view/book/4864/

Дмитриев Е.А. Математическая статистика в почвоведении.М.: МГУ, 1995.291с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww

Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

$\bar{X}$ или М - средняя арифметическая,

s^2 - дисперсия,

s - стандартное отклонение,

s_x или m - ошибка средней арифметической ,

V% - коэффициент вариации

s_x % - относительная ошибка выборочной средней (

Ме- медиана

Мо - мода

Глоссарий

Сре́днее арифмети́ческое — одна из наиболее распространённых мер центральной тенденции, представляющая собой сумму всех зафиксированных значений, деленную на их количество.

Диспе́рсия случа́йной величины́ *— мера разброса данной случайной величины, то есть её отклонения от математического ожидания*

Мо́да *— значение во множестве наблюдений, которое встречается наиболее часто. (Мода = типичность.)*

Медиа́на (50-й перцентиль, квантиль 0,5) — *возможное значение признака, которое делит ранжированную совокупность (вариационный ряд выборки) на две равные части:.*

Вопросы для изучения по теме:

Параметрические показатели выборки

Средняя арифметическая величина

Дисперсия

Стандартное отклонение

Коэффициент вариации

Ошибка выборочной средней

Непараметрические характеристики выборки

Мода

Медиана

Квантили

Конспект лекций:

Параметрические и непараметрические показатели выборки. Условия применения параметрических показателей выборки.

Параметрические методы могут использоваться только для характеристики выборок имеющих непрерывную количественную изменчивость и подчиняющихся закону нормального распределения. Для таких выборок допустимо использовать параметры – среднее выборочное и выборочная дисперсия, на которых основываются все остальные статистически е показатели.

Основными статистическими характеристиками количественной изменчивости являются средняя арифметическая ($\overline{X}$ *или М), дисперсия (*s^2*), стандартное отклонение (s), ошибка средней арифметической (*s_x*), коэффициент вариации (V) и относительная ошибка выборочной средней (*s_x *%).*

Основные параметрический показатели выборки:

Средняя арифметическая ($\overline{X}$ или М). Расчет средней арифметической. Условия применения среднего арифметического.

Средняя арифметическая $\overline{X}$ *представляет собой обобщенную, абстрактную характеристику всей совокупности в целом. Рассчитывается средняя арифметическая по формуле:*

$$\overline{X} = \Sigma X/n.$$

где: сумма всех вариант (X1+X2+.,.+Xn)= ΣX, число всех вариант - n

Основное свойство средней арифметической заключается в равенстве суммы всех положительных и всех отрицательных отклонений от нее, т.е.

сумма центральных отклонений всех отдельных вариант от $\overline{X}$ *равна нулю* $\Sigma(Xi-\overline{X})=0$.

Дисперсия (s^2),

Дисперсия s_2 *и стандартное отклонение s служат основными .мерами вариации, рассеяния изучаемого признака.*

Дисперсия представляет собой частное от деления суммы квадратов отклонений $\Sigma(Xi-\overline{X})^2$ *на число всех измерений без единицы (n—1):*

$$s^2 = \frac{\sum (Xi + \overline{X})^2}{n-1}$$

Стандартное отклонение (s),

Размерность дисперсии равна квадрату размерности изучаемого признака, что неудобно и заставляет ввести для измерения рассеяния другую характеристику, имеющую размерность варьирующей величины и называемую стандартным или средним квадратическим отклонением. Его получают извлечением квадратного корня из дисперсии:

$$s = \sqrt{\frac{\sum (Xi + \overline{X})^2}{n-1}}$$

При определении любых средних величин сумму всех показателей необходимо делить на число независимых друг от друга величии. В связи с этим в формулах сумму квадратов отклонений делят не на общее число наблюдений, а на число без единицы, так как одно любое отклонение зависимое и может быть найдено из равенства $\Sigma(Xi-\overline{X})=0$. *Остальные отклонения могут свободно варьировать, принимать любые значения. Число свободно варьирующих величин называется числом степеней свободы или числом степеней свободы вариации. Оно обозначается v (k). в простейшем случае v (k) равно n-1, в более сложных формулах может быть n-2 или n-3.*

Ошибка средней арифметической (s_x или m),

Ошибка выборочной средней или ошибка выборки S_X *или m является мерой отклонения выборочной средней от средней всей (генеральной) совокупности μ. Величина ошибок зависит от степени изменчивости изучаемого признака и от объема выборки.*

$$s_x = \frac{s}{\sqrt{n}}$$

Ошибки выборки выражают в тех же единицах измерения, что и варьирующий признак, и приписывают к соответствующим средним со знаками ±, т. е. $\overline{X} \pm s_x$ *или M±m.*

Ошибка средней арифметической тем меньше, чем меньше варьирует опытный материал и чем из большего числа измерений вычислено среднее арифметическое. При n→∞ *Средняя арифметическая* $\overline{X}$ *теоретически становится равным средней всей (генеральной) совокупности μ (генеральной средней))*

Коэффициент вариации (V) и

Коэффициент вариации вычисляется по формуле

$$V = \frac{s}{\overline{X}} * 100\%$$

Коэффициент вариации V, s *— стандартное отклонение, выраженное в процентах к средней арифметической* $\overline{X}$ *данной совокупности:*

Коэффициент вариации является относительным показателем изменчивости. Использование коэффициента вариации имеет смысл при изучении вариации признака, принимающего только положительные значения.• Не имеет смысла, коэффициент вариации, вычисленный для характеристики изменчивости, когда варьирующий признак принимает как положительные, так и отрицательные значения.

Грубая градация коэффициента вариации: изменчивость принято считать незначительной, если коэффициент вариации не превышает 10% средней, если V выше 10%, но менее 20%, и значительной, если коэффициент вариации более 20%.

Более детальная градация:

Меньше 5% - очень слабая

5-10% -слабая

10-20%- средняя

20-30% - сильная

больше 30% -очень сильная

При изучении вариабельности признаков одинаковой размеренности необходима осторожность — коэффициент вариации может дать искаженное представление об изменчивости..

Относительная ошибка выборочной средней (s_x %).

Ошибка выборки, выраженная в процентах, от соответствующей средней, называется относительной ошибкой выборочной средней:

$$S_x\% = \frac{s_x}{\bar{x}} * 100\%$$

Среднее арифметическое взвешенное .

Среднее арифметическое взвешенное учитывает вклад каждого варианта с учетом его доли в среднее значение. Среднее арифметическое взвешенное набора вещественных чисел $x_1, \ldots, x_n$ *с неотрицательными вещественными весами (или долями)* $w_1, \ldots, w_n$ *определяется как*

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i}.$$

При этом не допускается одновременное равенство всех весов нулю (но допускается равенство некоторых из них).

Часто подразумевают, что сумма весов равна 1, тогда формула выглядит:

$$\bar{x} = \sum_{i=1}^{n} w_i \cdot x_i.$$

В том случае, если все веса равны между собой, среднее арифметическое взвешенное будет равно среднему арифметическому.

Среднее геометрическое

Средним геометрическим нескольких положительных вещественных чисел называется такое число, которым можно заменить каждое из этих чисел так, чтобы их произведение не изменилось. Более формально:

$$G(x_1, x_2, \ldots, x_n) = \sqrt[n]{x_1 x_2 \cdots x_n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

Среднее геометрическое двух чисел также называется их ***средним пропорциональным***[

Условия применения непараметрических показателей выборки. Непараметрические показатели выборки и нормальное распределении.

Все параметрические критерии основаны, в конечном счете, на учете основных параметров – выборочной средней арифметической и выборочной дисперсии, применение которых правомерно только для распределений аппроксимируемых нормальным законом, или приводимых к нормальному различными преобразованиями (к примеру логнормальное распределение). Однако это не всегда имеет место, так как не все признаки природных объектов распределяются нормально или могут быть описаны непрерывными *распределениями. В таких случаях необходимо использовать непараметрические критерии.*

Известен целый ряд непараметрических критериев, среди которых видное место занимают так называемые ранговые критерии, применение которых основано на ранжировании членов сравниваемых групп. При этом сравниваются не сами по себе члены ранжированных рядов, а их порядковые номера, или ранги.

Мода (Mo).

Мода - это наиболее часто встречающееся значение случайной величины. Для признаков, измеренных на именной или порядковой шкале, мода представляет собой тот класс, которому соответствует наибольшая вероятность.

Для дискретных количественных признаков модой служит то значение случайной величины, которому соответствует наибольшая вероятность.

Для непрерывных случайных величин мода представляет собой то значение, которому соответствует наибольшая плотность вероятности, т.е. наибольшая ордината на кривой плотности распределения. Иными словами, модальному значению непрерывной случайной величины соответствует вершина кривой распределения.

Рассчитывается мода по формуле:

$$M_o = X_{Mo} + h_{Mo} * (f_{Mo} - f_{Mo-1}) : ((f_{Mo} - f_{Mo1}) + (f_{Mo} - f_{Mo+1})),$$

здесь X_{Mo} *— левая граница модального интервала,* h_{Mo} *— длина модального интервала,* f_{Mo-1} *— частота премодального интервала,* f_{Mo} *— частота модального интервала,* f_{Mo+1} *— частота послемодального интервала.*

Медиана (Me).

Медиа́на (50-й перцентиль, квантиль 0,5) — возможное значение признака, которое делит ранжированную совокупность (вариационный ряд выборки) на две равные части: 50 % «нижних» единиц ряда данных будут иметь значение признака не больше, чем медиана, а «верхние» 50 % — значения признака не меньше, чем медиана. Медиана является важной характеристикой распределения случайной величины и так же, как математическое ожидание, может быть использовано для центрирования распределения. Медиана может быть более предпочтительной для распределений с т.н. тяжёлыми хвостами. Медиана определяется для широкого класса распределений (например, для всех непрерывных).

Квантили. Квартили. Децили. Перцентили. Соотношения медианы, квартилей, децилей и перцентилей. Абсолютное отклонение крайних квартилей от медианы, выраженное через стандартное отклонение. Межквартильный размах оценка стандарта по размаху.

Наряду с медианой и модой к структурным характеристикам вариационного ряда относятся так называемые квантили, отсекающие в пределах ряда определенную часть его членов. К ним относятся квартили, децили и перцентили (процентили). Квартили — это три значения признака (Q1, Q2, Q3), делящие ранжированный вариационный ряд на четыре равные части. Аналогично, девять децилей делят ряд на 10 равных частей, а 99 перцентилей — на 100 равных частей.

В практике используют обычно перцентили P3, P10, P25, P50, P75, P90 и P97. Причем P25 и P75 соответствуют первому и третьему квартилям, между которыми находится 50% всех Членов ряда, а P50 соответствует второму квартилю и равен Медиане, т. е. P50=Me. Абсолютное отклонение крайних квартилей от медианы, выраженное через стандартное отклонение, равно 0,674σ. Поэтому межквартильный размах Rq равен 1,34 σ, и это может быть использовано для оценки стандарта по размаху:

Вопросы для самоконтроля знаний:

- Среднее арифметическое и его свойства.
- Мода и медиана
- Оценки моды, медианы и среднего арифметического.
- Стандартное отклонение и его свойства.
- Оценка дисперсии.
- Число степеней свободы
- Коэффициент вариации и его оценка.
- Ошибка среднего.
- Ошибка среднего и смешанные образцы
- Использование квентилей

Тема 3. Параметрические методы проверки статистических гипотез. Использование критерия Стьюдента. Парный двухвыборочный t-тест. Анализ независимых выборок. Использование критерия Фишера

Аннотация. Данная тема дает представление о параметрических критериях используемых для проверки статистических гипотез, разработанных для анализа зависимых и независимых выборок, соответствующих закону нормального распределения.

Ключевые слова. Парный двухвыборочный t-тест, t-тест независимых выборок,. выбраковка сомнительных дат

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литература: и электронные ресурсы:

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб. пособие для студентов вузов / В.Е. Гмурман.—12-е изд., перераб..—Москва: Высш. образование, 2007.—478, [1] с.

Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов..—4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351[1]с.

Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов..—М.: Изд. МГУ, 1970.—367с.

Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.

Дмитриев Е.А. Математическая статистика в почвоведении.М.: МГУ, 1995.291с.

Теория вероятностей и математическая статистика: Учебное пособие / С.В. Павлов. - М.: ИЦ РИОР: ИНФРА-М, 2010. - 186 с. http://znanium.com/bookread.php?book=217167

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

t – критерий Стьюдента

F – критерий Фишера

τ - тата критерий для выбраковки сомнительных дат

Глоссарий

t-критерий Стьюдента — общее название для класса методов статистической проверки гипотез (статистических критериев), основанных на распределении Стьюдента.

Независимые выборки - выборки в которых единицы наблюдения первой выборки не связаны никаким общим условием с единицами наблюдения второй выборки;

Зависимые (парные, сопряженные) выборки - выборки в которых единицы наблюдения первой выборки связаны (сопряжены) каким-то общим условием с единицами наблюдения второй выборки.

Вопросы для изучения по теме:

Статистические гипотезы и их проверка

Использование критерия Стьюдента.

Парный двухвыборочный t-тест.

Анализ независимых выборок.

F-тест на равенство дисперсий. Две выборки.

Выбраковка сомнительных дат

Конспект лекций

Статистические гипотезы и их проверка. Понятие о статистической гипотезе. Алгоритм проверки гипотез. Нулевая и альтернативные гипотезы.

Статистические методы или критерии проверки гипотез — надежная основа принятия решений при неопределенности, обусловленной случайной вариацией изучаемых явлений. Практически проверка гипотез сводится к сравнению статистических характеристик, оценивающих параметры законов распределения, т. е. к проверке определенных статистических гипотез. Статистической гипотезой называют научное предположение о тех или иных *статистических законах распределения рассматриваемых случайных величин, которое может быть проверено на основе выборки. В большинстве случаев задача сводится к проверке гипотезы об отсутствии реального различия между фактическими и теоретически ожидаемыми наблюдениями. Эту гипотезу называют нулевой гипотезой и обозначают символом* H_o.

Принятие нулевой гипотезы означает, что данные наблюдений не противоречат предположению об отсутствии различий между фактическими и гипотетическими (теоретическими) или между двумя рядами фактических распределений, но не доказывают отсутствия такого различия. Отбрасывание гипотезы означает, что эмпирические данные несовместимы с Но, а верна другая, альтернативная гипотеза.

Справедливость нулевой гипотезы проверяется вычислением статистических критериев проверки для определенного уровня значимости. В почвоведение как и в других науках связанных с изучением природных объектов принято использовать 95% уровень значимости допускающий ошибку 5%.

Параметрические и непараметрические критерии.

Параметрическими критерии, основаны на предположении, что распределение признака в совокупности подчиняется некоторому известному закону, например закону нормального распределения. К таким критериям относятся, в частности, критерии t и F, применение которых требует вычисления оценок параметров распределения.

Непараметрическими критерии, требуют предварительного вычисления оценок неизвестных параметров распределения и даже приближенного значения закона распределения признака. Они могут применяться как тогда, когда распределение сильно отклоняется от нормального так и соответствующему нормальному. Непараметрические критерии менее эффективны по сравнению с параметрическими, и поэтому их применение для выборок с небольшим объемом может быть не оправдано.

t- распределение Стьюдента. Форма кривой t- распределения Стьюдента. Зависимость формы кривой t- распределения Стьюдента от объема выборки.

Нормальная кривая с параметрами μ = 0 и σ=1 называется нормальной *или* стандартизованной кривой. *Она описывается формулой*

$$P(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

где

Любую нормальную кривую можно привести к стандартной

(вычитанием μ из Xi и делением на σ. $t = \frac{X_i - \overline{\mu}}{\sigma}$

Тогда для выборочных значений будет стандартизация $t = \frac{Xi - \overline{X}}{s_x}$

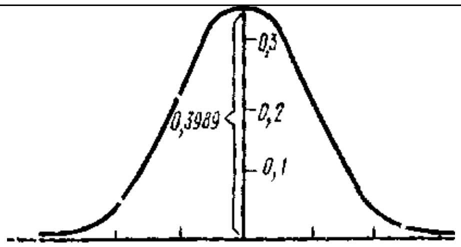


Рис. 4. Вид стандартизованной кривой нормального распределения

Стандартная кривая (рис.4) имеет площадь, равную единице. Максимальная ордината Умах , соответствует началу прямоугольных координат, перенесенному в центр распределения, где Xi—μ=0. Вправо и влево от этого центра случайная величина Х может принимать любые значения, и величина каждого отклонения (Xι—μ) определяется функцией его нормированного отклонения f(t).

При распределении совокупности наблюдений по нормальному закону в интервале от μ—t до μ+t окажется 68,3% от общего числа вариант, составляющих данную совокупность. В интервале от μ—2t до μ

+ 2t будет находиться 95,4% от числа всех вариант совокупности. И в интервале от μ—3t до μ + 3t *окажется 99,7% от общего объема совокупности.*

Вероятность (P) того, что случайно отобранная варианта не отклонится от средней μ более чем на μ±2 t , равна 0,954 . Это означает, что 95,4% от всех вариант нормально распределяющейся совокупности находится в пределах μ±2σ, что соответствует правилу трех сигм.

F- распределение Фишера. Форма кривой F- распределения Фишера. Зависимость формы кривой F- распределение Фишера от объема выборки. Критические (стандартные) значения t- распределения Стьюдента и F- распределения Фишера и их зависимость от объема выборки и заданного уровня значимости.

F-распределение было открыто английским ученым Р. А. Фишером, который открыл закон распределения отношения средних квадратов (дисперсий). Если из одной генеральной совокупности подчиняющейся закону нормального распределения выделить две выборки, посчитать для них дисперсии и поделить большую дисперсию на меньшую, то получается величина F которая представляет собой непрерывную величину кривая вероятности распределения которой от F представляет собой асимметричную кривую форма которой зависит от объемов (степеней свободы) выборок.

Интервальные и точечные оценки среднего и дисперсии.

Одновыборочный t-критерий

Применяется для проверки нулевой гипотезы о равенстве математического ожидания $E(X)$ *некоторому известному значению* m.

Используя несмещенную оценку дисперсии $s_X^2=\sum_{t=1}^n(X_t-\overline{X})^2/(n-1)$

имеем t-статистику:

$$t = \frac{|\overline{X} - m|}{s_X / \sqrt{n}}$$

При превышении значения t критического значения t_{st} нулевая гипотеза отвергается. . Стандартное значение критерия Стьюдента для различных уровней значимости и числа степеней свободы приводятся в специальной таблице (Приложение, табл.1).

Доверительная вероятность и уровень значимости. Доверительный интервал. Использование доверительного интервала для проверки статистических гипотез. Сравнение средних с постоянными величинами.

Оценка может быть представлена одним числом, точкой (точечная оценка) или некоторым интервалом (интервальная оценка), в котором с определенной вероятностью может находиться искомый параметр. Выборочная средняя является несмещенной и наиболее эффективной точечной оценкой генеральной средней μ, а выборочная дисперсия s^2 несмещенной точечной оценкой генеральной дисперсии σ. Обозначая ошибку выборочной средней s_x, точечную оценку генеральное средней можно записать в виде - $X \pm s_x$. Интервальной называют оценку, которая характеризуется двумя числами — концами интервала, покрывающего оцениваемый параметр. Доверительным называют такой интервал, который с заданной вероятностью покрывает оцениваемый параметр.

Доверительный интервал для генеральной средней записывается как:

$$\overline{x} - ts_{\overline{x}} \leqslant \mu \leqslant \overline{x} + ts_{\overline{x}},$$

где, ts_x — предельная ошибка выборочной средней при данном числе степеней свободы и принятом уровне значимости. Стандартное значение критерия Стьюдента для различных уровней значимости и числа степеней свободы приводятся в специальной таблице (Приложение, табл.1). Если доверительные интервалы рассчитанные для двух выборок пересекаются то нулевую гипотезу игнорировать нельзя, если наоборот то нулевая гипотеза *может быть отвергнута и разница средних арифметических может быть оценена как значима.*

Анализ двух и более выборок. Зависимые (парные) и независимые выборки.

При сравнении средних величин может быть 2 случая 1) сравниваются две сопряженные выборки, в которых единицы наблюдения первой выборки связаны (сопряжены) с единицами наблюдения второй выборки; 2) сравниваются средние двух независимых выборок, когда единицы наблюдения первой выборки не связаны с единицами наблюдения второй выборки. В случае сопряженных (парных) выборок по критерию Стьюдента оценивается существенность существенность средней разности ($d=\Sigma d:n$), *а во втором разности средних* ($d=\bar{X}_1-\bar{X}_2$)

Парный двухвыборочный t-тест.

Для вычисления эмпирического значения t-критерия в ситуации проверки гипотезы о различиях между двумя зависимыми выборками применяется следующая формула:

$$t = \frac{|M_d|}{s_d/\sqrt{n}}$$

где M_d *— средняя разность значений,* s_d *— стандартное отклонение разностей, n — количество наблюдений. Значение критерия сравнивается со стандартным значением критерия Стьюдента (Приложение, табл.1), которое определяется уровнем значимости 95% (принятом в почвоведении) и степенью свободы* n-1. *Если* $t_f \geq t_{st}$ *нулевая гипотеза об отсутствии существенных различий между средними (нулевая гипотеза) между средними опровергается*

Независимые выборки. Проверка гипотезы о равенстве средних независимых выборок. Двухвыборочный t-тест независимых выборок.

Двухвыборочный t-критерий для независимых выборок

Tckb имеются две независимые выборки объемами n_1 *и* n_2 *нормально распределенных случайных величин.* $\bar{X}_1, \bar{X}_2$. *Нулевая гипотеза будет означать равенства математических ожиданий этих случайных величин* $\mu_1=\mu_2$.

Разность выборочных средних. $d=\bar{X}_1-\bar{X}2$

Несмещенная оценка дисперсии $s^2 = \frac{\sum_{t=1}^{n}(X_t - \overline{X})^2}{n-1}$

Несмещенная оценку дисперсии разности выборочных средних: $s_\Delta^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}.$

t-статистика для проверки нулевой гипотезы равна

$$t = \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}}$$

Стандартное значение t_{st} определяется при степени свободы k=df и 5% уровне значимости.

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}$$

Случай одинаковой дисперсии

В случае, если дисперсии выборок предполагаются одинаковыми, тогда t-статистика равна:

$$t = \frac{|\overline{X}_1 - \overline{X}_2|}{s_X\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\ ,\quad s_X = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Стандартное значение t_{st} определяется при $k=n_1+n_2-2$ и 5% уровне значения.

Использование критерия Фишера для оценки разности дисперсий.

Если имеются две выборки объемом m и n соответственно случайных величин X и Y, имеющих нормальное распределение. Необходимо проверить равенство их дисперсий. Статистика теста

$$F = \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \ \sim\ F(m-1, n-1)$$

где $\hat{\sigma}^2$ — выборочная дисперсия посчитанная отдельно для случайных величин .

Если статистика больше критического, то дисперсии не одинаковы, в противном X и Y случае дисперсии выборок одинаковы. В зависимости от результата оценки одинаковости дисперсий выбирают метод оценки значимости разницы средних для независимых выборок по критерию Стьюдента.

Выбраковка сомнительных дат. Использование таута-критерия для выбраковки сомнительных дат.

Гипотезу о принадлежности «сомнительных», наиболее отклоняющихся вариант X_{min} и X_{max} к данной совокупности в малых выборках проверяют по критерию τ. Фактическое значение критерия, представляющее собой отношение разности между сомнительной и соседней с ней датой к размаху варьирования, сравнивают с теоретическим на 5%-уровне значимости.

Если $\tau_{факт} > \tau_{теор}$, то варианта отбрасывается, если $\tau_{факт} < \tau_{теор}$, то варианта оставляется и нулевая гипотеза о принадлежности ее к данной совокупности не отвергается.

Критические значения критерия τ теор, которые зависят от принятого уровня значимости и от объема выборки n, даны в таблице приложения.

Для расчета фактического значения критерия τ, варианты располагают в порядке возрастания: X1, X1...Xn-1, Xn.

Критерий τ вычисляют по отношениям:

$$\text{для } X_1 \tau = \frac{X_2 - X_1}{X_{n-1} - X_1} \quad \text{и} \quad \text{для } X_n \tau = \frac{X_n - X_{n-1}}{X_n - X_2}.$$

где, $X_1 = X_{min}$, $X_n = X_{max}$, X_2 – второе значение в ранжированном ряду, X_{n-1}- предпоследнее значение.

Таблица 1.

Критические точки t-критерия Стьюдента при различных уровнях значимости

Число степеней свободы	α, %		Число степеней свободы	α, %	
	5	1		5	1
1	12.71	63,66	I8	2,10	2,88
2	4,30	9,92	19	2,09	2,86
3	3,18	5,84	20	2,09	2,85
4	2,78	4,60	21	2,08	2,83
5	2,57	4,03	22	2,07	2,82
8	2,45	3,71	23	2,07	2,81
7	2,37	3,50	24	2,06	2,80
8	2,31	3,36	25	2,06	2,79
9	2,26	3,25	26	2,06	2,78
10	2,23	3,17	27	2,05	2,77
11	2,20	3,11	28	2,05	2,76
12	2,18	3,05	29	2,05	2,76
13	2,16	3,01	30	2,04	2,75
14	2,14	2,98	40	2,02	2,70
15	2,13	2,95	60	2,00	2,66
16	2,12	2,92	120	1,98	2,62.
17	2,11	2,90	∞	1,96	2,58

Таблица 2.

Стандартные значения критерия F при уровне значимости 1% (верхнее значение) и уровне значимости 5% (нижнее значение)

ν_2	ν_1	1	2	3	4	5	6	7	8	9	10	11	12	24	
2		98,5	99,0	92,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5
		18,5	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5
3		34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,4	27,2	27,1	27,1	26,6	26,1
		10,1	9,6	9,3	9,1	9,0	8,9	8,9	8,8	8,8	8,8	8,8	8,7	8,6	8,5
4		21,2	18,8	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,7	14,5	14,4	13,9	13,5
		7,7	6,9	6,6	6,4	6,3	6,2	6,1	6,0	6,0	6,0	5,9	5,9	5,8	5,6
5		16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	10,0	9,9	9,5	9,0
		6,6	,5,8	5,4	5,2	5,1	5,0	4,9	4,8	4,8	4,7	4,7	4,7	4,5	4,4
6		13,4	10,9	9,8	9,2	8,8	8,5	8,3	8,1	8,0	7,9	7,8	7,7	7,3	6,9
		6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,1	4,1	4,1	4,0	4,0	3,8	3,7
7		12,3	9,6	8,5	7,9	7,5	7,2	7,0	6,8	6,7	6,6	6,5	6,4	6,1	5,7
		5,6	4,7	4,4	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,6	3,6	3,4	3,2
8		11,3	8,7	7,6	7,0	6,6	6,4	6,2	6,0	5,9	5,8	5,7	5,7	5,3	4,9

	5,3	4,6	4,1	3,8	3,7	3,6	3,5	ЗА	3,4	3,3	3,1	3,3	3,1	2,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,6	5,5	5,4	5,3	5,2	5,1	4,7	4,3
	5,1	4,8	3,6	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,1	3,1	2,9	2,7
10	10,0	7,9	6,6	6,0	5,6	5,4	5,2	5,1	5,0	4,9	4,8	4,7	4,3	3,9
	5,0	4,1	3,7	3,5	3,3	3,2	3,1	3,1	3,0	2,9	2,9	2,9	2,7	2,5

ν_2	ν_1	1	2	3	4	5	6	7	8	9	10	11	12	24	
11		9,7	7,2	6,2	5,7	5,3	5,1	4,9	4,7	4,6	4,5	4,5	4,4	4,0	3,6
		4,8	4,0	3,6	3,4	3,2	3,1	3,0	3,0	2,9	2,9	2,8	2,8	2,6	2,4
12		9,3	6,9	6,0	5,4	5,1	4,8	4,7	4,5	4,4	4,3	4,2	4,2	3,8	3,4
		4,8	3,9	3,5	3,3	3,1	3,0	2,9	2,9	2,8	2,8	2,7	2,7	2,5	2,3
13		9,1	6,7	5,7	5,2	4,9	4,6	4,4	4,3	4,2	4,1	4,0	4,0	3,6	3,2
		4,7	3,8	3,4	3,2	3,0	2,9	2,8	2,8	2,7	2,7	2,6	2,6	2,4	2,2
14		8,9	6,5	5,6	5,0	4,7	4,5	4,3	4,1	4,0	3,9	3,9	3,8	3,4	3,0
		4,6	3,7	3,3	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,6	2,5	2,3	2,1
15		8,7	6,4	5,4	4,9	4,6	4,3	4,1	4,0	3,9	3,8	3,7	3,6	3,3	2,9
		4,5	3,7	3,3	3,1	2,9	2,8	2,7	2,6	2,6	2,6	2,5	2,5	2,3	2,1
16		8,5	6,2	5,3	4,8	4,4	4,2	4,0	3,9	3,8	3,7	3,6	3,5	3,2	2,8
		4,5	3,6	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,5	2,5	2,4	2,2	2,0
17		8,4	6,1	5,2	4,7	4,3	4,1	3,9	3,8	3,7	3,6	3,5	3,5	3,1	2,7
		4,5	3,6	3,2	3,0	2,8	2,7	2,6	2,6	2,5	2,5	2,4	2,4	2,2	2,0
18		8,3	6,0	5,1	4,6	4,2	4,0	3,8	3,7	3,6	3,5	3,4	3,4	3,0	2,6
		4,4	3,5	3,2	2,9	2,8	2,7	2,6	2,5	2,7	2,4	2,4	2,3	2,1	1,9
19		8,2	5,9	5,0	4,5	4,2	3,9	3,8	3,6	3,5	3,4	3,4	3,3	2,9	2,5
		4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,1	1,9
20		8,1	5,8	4,9	4,4	4,1	3,9	3,7	3,6	3,4	3,4	3,3	3,2	2,9	2,4
		4,3	3,5	3,1	2,9	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,3	2,1	1,8
21		8,0	5,8	4,9	4,4	4,0	3,8	3,6	3,5	3,4	3,3	3,2	3,2	2,8	2,4
		4,3	3,5	3,1	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,0	1,8
22		7,9	5,7	4,8	4,3	4,0	3,8	3,6	3,4	3,3	3,3	3,2	3,1	2,8	2,3
		4,3	3,4	3,0	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,0	1,8
23		7,9	5,7	4,8	4,8	4,0	3,7	3,5	3,4	3,3	3,2	3,1	3,1	2,7	2,3
		4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,0	1,8
24		7,8	5,6	4,1	4,2	3,9	3,7	3,5	3,4	3,2	3,2	3,1	3,0	2,7	2,2
		4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2	2,0	1,7
25		7,8	5,6	4,7	4,2	3,9	3,6	3,5	3,3	3,2	3,1	3,0	3,0	2,6	2,2
		4,2	3,4	3,0	2,8	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,2	2,0	1,7
26		7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,2	3,1	3,0	3,0	2,6	2,1
		4,2	3,4	3,0	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,0	1,7
27		7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,1	3,1	3,0	2,9	2,5	2,1
		4,2	3,3	3,0	2,7	2,6	2,5	2,4	2,3	2,2	2,2	2,2	2,1	1,9	1,7
28		7,6	5,4	4,6	4,1	3,8	3,5	3,4	3,2	3,1	3,0	2,9	2,9	2,5	2,1
		4,2	3,3	2,9	2,7	2,6	2,4	2,4	2,3	2,2	2,2	2,1	2,1	1,9	1,7
29		7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,9	2,5	2,0
		4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,1	1,9	1,6
30		7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,8	2,5	2,0
		4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,1	1,9	1,6

17. Критические значения критерия τ для 5%-ного и 1%-ного уровня значимости

n	τ		n	τ	
	0,01	0,05		0,01	0,05
4	0,991	0,955	14	0,502	0,395
5	0,916	0,807	16	0,472	0,369
6	0,805	0,689	18	0,449	0,349
7	0,740	0,610	20	0,430	0,334
8	0,683	0,554	22	0,414	0,320
9	0,635	0,512	24	0,400	0,309
10	0,597	0,477	26	0,389	0,299
11	0,566	0,450	28	0,378	0,291
12	0,541	0,428	30	0,369	0,283

Вопросы для самоконтроля знаний:

- Особенности распределения Стъюдента.
- Доверительные границы среднего.
- Гарантированные минимум и максимум среднего.
- Сравнение средних с постоянными величинами.
- Сравнение двух или нескольких дисперсий.
- Сравнение двух средних.
- Сравнение средних, имеющих одинаковые дисперсии.
- Сравнение средних, имеющих разные дисперсии.
- Сравнение выборочных дисперсий

Тема 4. Непараметрические методы проверки статистических гипотез.

Аннотация. Данная тема дает представление о непараметрических критериях используемых для проверки статистических гипотез, разработанных для анализа зависимых и независимых выборок, не соответствующих (или соответствующих) закону нормального распределения.

Ключевые слова. Критерий Уилкоксона, критерий Манн-Уитни

Методические рекомендации по изучению темы

✓ - Тема содержит лекционную часть, где дается общее представление о непараметрических критериях используемых для проверки статистических гипотез

✓ В качестве самостоятельной работы студентам провести оценку значимости разности средних с применением непараметрических критериев.

✓ Для проверки усвоения темы имеются вопросы.

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литературы: м электронные ресурсы:

Самсонова В.П., Мешалкина Ю.Л., Дядькина С.Е. Практикум на компьютере по курсу: "математическая статистика" М. изд-во МГУ, 2005 г. 150 с. (доступно на http://www.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=25#top)

Домашнее задание по некоторым простейшим задачам математической статистики: Для студентов физ.фак. / ; Казан.гос.ун-т;Сост.:Билялов Р.Ф.,Никитин Б.С..?Казань: Б.и., 1996.12с..?500р.

Балахчев, Генрих Николаевич. Основы математической обработки и статистическое обоснование экспериментального материала: [методическое пособие для курса математическая статистика в почвоведении] / Г. Н. Балахчев, Г. Ф. Копосов; Казан. гос. ун-т, Биол.-почв. фак..?Казань: [Казанский государственный университет], 2007.?30, [1] с.: ил

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб. пособие для студентов вузов / В.Е. Гмурман.—12-е изд., перераб..—Москва: Высш. образование, 2007.—478, [1] с.

Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов..—4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351[1]с.

Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов..—М.: Изд. МГУ, 1970.—367с.

Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.

Дмитриев Е.А. Математическая статистика в почвоведении.М.: МГУ, 1995.291с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww

Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

W- непараметрический критерий Ван-дер-Вардена

T- непараметрический критерий Уилкоксона

U – непараметрический критерий Манн-Уитни

Глоссарий

Т-критерий Уилкоксона **—** непараметрический статистический тест (критерий), используемый для проверки различий между двумя выборками парных измерений.

U-критерий Манна — Уитни (англ. Mann — Whitney U-test) — статистический критерий, используемый для оценки различий между двумя независимыми выборками по уровню какого-либо признака, измеренного количественно. Позволяет выявлять различия в значении параметра между малыми выборками. Другие названия: критерий Манна — Уитни — Уилкоксона (англ. *Mann — Whitney — Wilcoxon, MWW*), критерий суммы рангов Уилкоксона (англ. *Wilcoxon rank-sum test*) или критерий Уилкоксона — Манна — Уитни (англ. *Wilcoxon — Mann — Whitney test*). Реже: критерий числа инверсий

Вопросы для изучения по теме:

Непараметрические критерии проверки гипотез

W-критерий Ван-дер-Вардена

T-критерий Уилкоксона

U – критерий Манн-Уитни

Конспект лекций

Непараметрические методы проверки статистических гипотез. Условия применения непараметрических методов проверки статистических гипотез. Непараметрические методы проверки статистических гипотез и нормальное распределение.

В отличие от параметрических критериев оценки статистических гипотез разности средних выборочных значений, построенных на основе параметров совокупности (среднего значения $\bar{x}$ и выборочной дисперсии s^2_x) и представляющие функции этих параметров, непараметрические критерии гипотез разности медиан, представляют собой функции, зависящие от вариант совокупности с их частотами. Первые параметры (как уже говорилось выше) применимы только для совокупностей, распределяемых по нормальному закону, вторые - к любым совокупностям независимо от формы их распределения. Известен целый ряд непараметрических критериев, среди которых видное место занимают так называемые ранговые критерии, применение которых основано на ранжировании членов сравниваемых групп. При этом сравниваются не сами по себе члены ранжированных рядов, а их порядковые номера, или ранги.

Ранжирование рядов. Выделение ранговых величин.

Если значения варьирующего признака X называют обозначить X_1, X_2, ... X_n. Ряд варьирующих величин можно упорядочить — расположить значения признака в порядке их возрастания. Такое упорядочение ряда, т. е. расположение вариант в порядке возрастания, называется ***р а н ж и р о в а н и е м*** *. Для определения рангов признаки располагаются в* *возрастающий ряд. Ранжированные значения признаков обозначаются порядковыми числами 1,2,3,4, …… Если значения признаков совпадают, то им присуждаются одинаковые ранги соответствующие среднему арифметическому из соответствующих чисел натурального ряда (например значений* $x_5 = x_6$ *ранги будут соответствовать* $\frac{5+6}{2} = 5{,}5$*).*

Использование критерия Вилкоксона для оценки зависимых (парных) выборок. Ограничения использования критерия Вилкоксона. Расчет критерия Вилкоксона. Использование критерия Вилкоксона для проверки статистических гипотез о равенстве средних. Т-критерий Уилкоксона.

Для выборок в которых отдельные варианты связаны попарно некоторыми общими условиями (случай зависимых выборок), различия между ними можно оценить с помощью рангового критерия Вилкоксона (Т-критерий). Парный Т-критерий Вилкоксона является более мощным, чем критерий знаков Ван-дер-вардена. Критерий предназначен для сопоставления показателей, измеренных в двух разных условиях на одной и той же выборке испытуемых. Расчет критерия проводится дледующим образом:

1. Рассчитывают разности попарно связанных вариант двух выборок.
2. Ранжируют, как положительные, как и отрицательные, в один общий ряд. При этом нулевые разности в расчет не принимают, а все остальные независимо от знака ранжируют так, чтобы наименьшая абсолютная разность получила первый ранг.
3. Находят отдельно суммы положительных и отрицательных разностей. Меньшую из двух сумм разностей, без учета ее знака, используют в качестве фактически установленной величины Т-критерия.

3. Сравнивают величину с критическим значением $T_{ст}$ *для принятого уровня значимости* $\alpha=0{,}05$ *и числа парных наблюдений n, которое берут без нулевых разностей. Нулевую гипотезу отвергают, если* $T_{факт}<T_{ст}$*,тогда разницу медиан считают существенной. Если* $T_{факт} > T_{ст}$ *то принимают нулевую гипотезу. Критические значения парного критерия Вилкоксона Тст приведены в таблице приложения.*

Ограничения критерия: 1. Объем выборки — от 5 до 50 элементов (парных значений). 2. Нулевые сдвиги исключаются из рассмотрения.

Существует урезанный вариант теста для сравнения одной выборки с известным значением медианы (по аналогии с точечной оценкой принятой в параметрической статистике).

Непараметрический анализ не зависимых выборок. Использование критерия Манн-Уитни для оценки независимых выборок. Ограничения использования критерия Манн-Уитни. Расчет критерия Манн-Уитни. Использование критерия Манн-Уитни для проверки статистических гипотез о равенстве средних.

U **-критерий Манна—Уитни.**

Гипотезу о принадлежности сравниваемых независимых выборок к одной и той же генеральной совокупности можно проверить с помощью популярного рангового критерия Манна—Уитни (U-критерия).

Для расчета U-критерия необходимо:

1. Расположить числовые значения двух сравниваемых выборок в порядке возрастания в один общий ранжированный ряд и пронумеровать члены общего ряда 1 до N=n1+n2 (это будут ранги).
2. Отдельно для каждого ряда найти суммы рангов R1 R2 и рассчитать значения.

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2},$$

которые отображают связь между суммами рангов первой и второй выборки.

3. В качестве U-критерия используют меньшую величину $U_{факт}$, которую сравнтвают с табличным значением U_{cm}. Условием для сохранения принятой Но-гипотезы служит неравенство $U_{факе}>U_{cm}$. Критические точки U-критерия U_{cm} для n1 n2 и принимаемого уровня (значимости α=0,05) в таблице 2 приложенияI

Ограничения применимости критерия

1.В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти.
2.В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало.

П Р И Л О Ж Е Н И Е

Таблица 1. Критические значения Т-критерия Вилкоксона в зависимости от уровня значимости и числа парных значений n.

n	*p<0,05*	*p<0,01*
5	*0*	*-*
6	*2*	*-*
7	*3*	*0*
8	*5*	*1*
9	*8*	*3*
10	*10*	*5*
11	*13*	*7*
12	*17*	*9*
13	*21*	*12*
14	*25*	*15*
15	*30*	*19*
16	*35*	*23*
17	*41*	*27*
18	*47*	*32*
19	*53*	*37*
20	*60*	*43*
21	*67*	*49*
22	*75*	*55*
23	*83*	*62*
24	*91*	*69*
25	*100*	*76*
26	*110*	*84*
27	*119*	*92*
28	*130*	*101*
29	*140*	*110*
30	*151*	*120*
31	*163*	*130*
32	*175*	*140*
33	*187*	*151*
34	*200*	*162*
35	*213*	*173*
36	*227*	*185*

37	*241*	*198*
38	*256*	*211*
39	*271*	*224*
40	*286*	*238*
41	*302*	*252*
42	*319*	*266*
43	*336*	*281*
44	*353*	*296*
45	*371*	*312*
46	*389*	*328*
47	*407*	*345*
48	*426*	*362*
49	*446*	*379*
50	*466*	*397*

Критические значения критерия U Манна-Уитни

(для проверки ненаправленных альтернатив)

P=0,05

n_2	**n1**													
	7	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
3	I	2	2	3	3	4	4	5	5	6	6	7	7	8
4	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	24	29	34	39	44	49	54	59	64	70	75	80	85	90

16	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Вопросы для самоконтроля знаний

Что такое ранговая величина?

Как проводится ранжирование выборки?

Как определятся ранговые величины?

Для каких выборок используется критерий Уилкоксона?

Для каких выборок используется критерий Манн-Уитни?

Тема 5. Обработка результатов исследований в электронных таблицах MS Excel. Графическое представление результатов исследований.

Аннотация. Данная тема дает представление о работе в электронных таблицах MS Excel. Кратко рассматриваются способы графического представления результатов исследований.

Ключевые слова. Электронные таблицы. Пакеты статистических программ.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где дается общее представление о программе **Microsoft Excel**
- Для проверки усвоения темы имеются вопросы.
- В качестве домашнего задания предлагается провести расчеты в электронных таблицах.

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литературы: м электронные ресурсы:

Викиучебник Microsoft Excel http://ru.wikibooks.org/wiki/Microsoft_Excel

Гиниятуллин К.Г. Решение задач корреляционного и регрессионного анализа в электронных таблицах MS EXCEL Методическое пособие Казань: Изд-во Казанский государственный университет им. В.И. Ульянова-Ленина, 2008. 32 с.

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

MS - продукт **Microsoft Office**

MAC - продукт **Macintosh**

VBA - язык макропрограммирования Visual Basic for Application

Глоссарий

Microsoft Excel (также иногда называется **Microsoft Office Excel**) — программа для работы с электронными таблицами, созданная корпорацией Microsoft для Microsoft Windows, Windows NT и Mac OS.

Файл (англ.*file*)— именованная область данных на носителе информации.

Формат — спецификация структуры данных, записанных в компьютерном файле. Формат файла иногда указывается в его имени, как часть, отделённая точкой.

Электронная таблица — компьютерная программа, позволяющая проводить вычисления с данными, представленными в виде двумерных массивов, имитирующих бумажные таблицы. Некоторые программы организуют данные в «листы», предлагая, таким образом, третье измерение.

Вопросы для изучения по теме:

О программе MS Excel

Порядок заполнения электронных таблиц

Конспект лекций

Програмный пакет MS Excel. Совместимость програмного пакета MS Excel. Электронные таблицы MS Excel.

Microsoft Excel *(также иногда называется* ***Microsoft Office Excel)*** *— программа для работы с электронными таблицами, созданная корпорацией Microsoft для Microsoft Windows, Windows NT и Mac OS. Она предоставляет возможности экономико-статистических расчетов, графические инструменты. Microsoft Excel входит в состав Microsoft Office и на сегодняшний день Excel является одним из наиболее популярных приложений в мире*

Язык макропрограммирования VBA.

Программный продукт ***Microsoft Excel,*** *за исключением Excel 2008 под Mac OS X, имеет встроенный язык макропрограммирования VBA (Visual Basic for Application), который позволяет создавать внутренние макропрограммные объекты.*

Анализ данных и обмен информацией. Формат сохранения данных.

Программный продукт ***Microsoft Excel*** *имеет встраиваемую надстройку «Пакет анализа», который позволяет с помощью встроенных функций проводить простые математические расчеты и статистический анализ данных электронных таблиц. Пакет предполагает возможности импорта (экспорта) данных из других (в другие программы), а также создание собственных файлов формата xls и xls.*

Структура рабочей книги

Документ Excel имеет расширение ".xls" ("*.xlsx" в Microsoft Office Excel 2007 и более поздних версиях) и называется рабочая книга. Рабочая книга состоит из листов. По умолчанию их создается три. Переключаться между листами можно, используя закладки (ярлычки) в нижней части окна "Лист 1" и т.д. Каждый лист представляет собой таблицу. Таблица состоит из столбцов и строк. Столбцов в листе 16384 (2 в 14 степени), а строк 1048576 (2 в 20 степени). Количество ячеек - 17179869184 (2 в 34 степени). Столбцы обозначаются буквами латинского алфавита (в обычном режиме) от "А" до "Z", затем идет "AA-AZ", "BA-BZ", "AAA" и т.п. до "XFD" (16384). Строки обозначаются обычными арабскими числами.*

Особенности расчетов в электронных таблицах. Порядок заполнения электронных таблиц

Составление электронных таблиц необходимо начать с определения формата ячеек, для этого выделяем мышью область будущей таблицы, выбираем категорию «формат», подкатегорию «ячейки». В появившемся окне «формат ячейки» выбираем категорию «число», задаем формат ячеек «числовой», задаем нужное значение «числа десятичных знаков». В левые крайние столбцы вводятся числовые значения признаков х и у в формате ***«00,00»****. Формулы вводятся после выделения необходимой ячейки в окошке «строка формул» на панели инструментов. Знак* ***«=»*** *вначале команды означает перевод ячейки в состояние ввода формулы. Значения ячеек используемых в формулах обозначается в формате «буква в латинской транслитерации»«число» (например - А1), где буква означаем индекс столбца, цифра – строки. Команда* ***«=А1»*** *в любой другой ячейки копирует в нее содержимое ячейки, соответствующей столбцу А строки 1. Знак* ***«+»*** *соответствует суммированию заданных значений,* ***«-»*** *- вычитанию,* ***«*»*** *- произведению,* ***«/»*** *- делению. Команда* ***«=А1*23,5»*** *- будет умножать значение ячейки А1 на 23,5. Команда* ***«=А1*А2*В1»*** *- будет умножать значения ячеек А1,А2 и В2.*

*Для разделения части формул от другой используются обычные, обязательно парные скобки **«()»**. Команда **«(A1+A2)/B1»** - будет делить сумму значений ячеек A1 и A2 на значение ячейки B2.*

*Для ввода более сложных математических формул необходимо использовать встроенные математические функции программы MS Excel. Для ввода функций нажимаем клавишу fx на панели инструментов, или выбираем подкатегорию «функция» в категории «вставка». В появившемся окне «мастер функций» выбирают категорию «статистические», нужную функцию. В появившемся новом окне устанавливают диапазоны ввода «массивов» и значения «чисел», для ввода можно использовать «пойтинг» (см.стр. 9). Функция **«=СЧЁТ(В3:В14)»** - выведет число значений ячеек столбца В строк с 2 по 13. Функция **«=СУММ(D2:D13)»** - суммирует значения ячеек столбца D строк с 2 по 13. Функция **«=СТЕПЕНЬ(В15;2)»** - возведет в степень 2 значение ячейки столбца В строки 15. Функция **«=КОРЕНЬ(В2)»** - выведет корень квадратный значения ячейки столбца В строки 2. Функция **«=LOG10(A1)»** - выведет десятичный логарифм значения ячейки столбца А строки 1. Функция **«=СРЗНАЧ(С11:С13**)», в категории «статистические» - выведет среднее из значений ячеек столбца С строк с 11 по 13.*

Формулы могут быть введены копированием из одной ячейки в другую, при этом происходит автоматическая корректировка ссылок на ячейки с учетом переноса. Например, если ячейка D1 содержит формулу «=A1/B1», то простое копирование иго содержимого в ячейку D2 даст новую формулу «=A2/B2».

Более сложные расчеты могут проводиться в надстройке «пакет анализа» в который встроены статистические методы. Статистическая обработка результатов в пакете анализа ориентирована в основном на применении параметрической статистике.

Графическое представление результатов исследований. Построение диаграмм. Построение точечных графиков.

Цифровой материал, введенный в рабочую книгу или рассчитанные в электронных таблицах может быть представлен в виде графиков. Для построения графиков необходимо запустить Мастер диаграмм. Затем выделить тип графика. Программа предлагает следующие виды графического представления материала: гистограммы (2D и 3D), линейные диаграммы, линейные графики, круговые диаграммы, точечные диаграммы, диаграммы с объластями и др. Полученные графики могут в дальнейшем редактироваться.

Экспорт линии тренда. Подбор линий тренда. Расчет эмпирических уравнений регресии по линиям тренда.

При создании точечных диаграмм полезной функцией является экспорт линии тренда. При вводе линии тренда можно выбрать тип аппроксимации эмпирических данных – линейный, полиноминальный, степенной, логарифмической, экспоненциальны и др. Программа позволяет вывести аппроксимирующую формулу.

Вопросы для самоконтроля знаний:

Какие расширения имеют файлы Microsoft Excel?
Какие файлы могут быть открыты в программе Microsoft Excel?
Что понимают под рабочим листом?
Что понимают под ячейкой в электронных таблица?
Какие форматы могут быть заданы в ячейке?
Какой символ обозначает переход в режим ввода формулы?

Тема 6. Анализ однофакторных дисперсионных комплексов.

Аннотация. В данной теме дается представление о принципах дисперсионного анализа, алгоритмах проведения дисперсионного анализа. Рассматриваются методы проведения однофакторного дисперсионного анализа

Ключевые слова. Дисперсионный анализ, Критерий Фишера. Однофакторные дисперсионные комплексы.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где дается общее представление о принципах дисперсионного анализа, видах дисперсионного анализа. Рассматриваются методы проведения однофакторного дисперсионного анализа
- Для темы разработаны вспомогательные материалы (глоссарий, список сокращений и др.)
- Освоение темы предполагает выполнение задания к практическому занятию. Задание предполагает проведение самостоятельной работы – дисперсионный анализ однофазного дисперсионного комплекса без повторений. Рекомендации для выполнения самостоятельной работы представлены внутри задания.
- Обучающийся должен оформить отчет о самостоятельной работе в виде файла конечных результатов и прикрепить этот файл к заданию
- Для проверки усвоения темы имеются вопросы для самоконтроля знаний.
- Для проверки усвоения темы необходимо пройти тест (Тестирование является завершающей стадией освоения темы).

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литература: м электронные ресурсы:

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб. пособие для студентов вузов / В.Е. Гмурман.—12-е изд., перераб..—Москва: Высш. образование, 2007.—478, [1] с.

Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов..—4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351[1]с.

Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов..—М.: Изд. МГУ, 1970.—367с.

Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.

Дмитриев Е.А. Математическая статистика в почвоведении.М.: МГУ, 1995.291с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

ANOVA –аббревиатура принятая в западной литературе для обозначения дисперсионного анализа.

НСР – наименьшая статистически значимая разница

НЗР – наименьшая значимая разница, современное обозначение НСР принятая в некоторых программах.

Сy - сумма квадратов центральных отклонений общая

Cv - сумма квадратов центральных отклонений вариантов

Cp - сумма квадратов центральных отклонений повторений

Cz - сумма квадратов центральных отклонений общая ошибки

Остаток - - сумма квадратов центральных отклонений общая ошибки

F – фактическое значение отношений дисперсий (критерия Фишера)

Fst- стандартное (критическое, табличное) значение отношений дисперсий (критерия Фишера)

N – общее число наблюдений

l – градации фактора

n – количество повторностей

Глоссарий

Дисперсионный анализ — метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях

F-тест критерий Фишера (F-критерий, φ*-критерий) —статистический критерий, тестовая статистика которого при выполнении нулевой гипотезы имеет распределение Фишера (F-распределение).

Критерий Фишера - F-тест (F-критерий, φ*-критерий) —статистический критерий, тестовая статистика которого при выполнении нулевой гипотезы имеет распределение Фишера (F-распределение).

φ*-критерий – обозначение F-критерия в неукоторых источниках

Вопросы для изучения по теме:

Дисперсионный анализ.

Общие принципы дисперсионного анализа.

Однофакторные дисперсионные комплексы.

Конспект лекций

Дисперсионный анализ. Общие принципы.

Дисперсионный анализ (метод ANOVA в англоязычной литературе) является одним из самых разработанных и мощных методов статистического анализа данных. Дисперсионный анализ основан на сравнении дисперсий рассчитываемых специальным способом в зависимости от особенностей группировки данных в дисперсионные комплексы.

F- распределение Фишера. Форма кривой F- распределения Фишера. Зависимость формы кривой F- распределение Фишера от объема выборки. F-распределение Фишера и дисперсионный анализ.

Дисперсионный анализ разработан и введен в практику сельскохозяйственных и биологических исследований английским ученым Р. А. Фишером, который открыл закон распределения: отношения средних квадратов (дисперсий):

$$F = s_1^2/s_2^2 \quad \text{при} \quad s_1^2 \geqslant s_2^2.$$

Закон распределения выводит, что если из одной генеральной совокупности подчиняющейся закону нормального распределения выделить две случайные выборки, посчитать для них дисперсии и поделить большую дисперсию на меньшую, то получается величина F, которая представляет собой непрерывную изменяющуюся величину, кривая зависимости вероятности распределения P(F) от F представляет собой асимметричную кривую (рисунок), форма которой зависит от объемов (степеней свободы) выборок.

Другими словами величина F имеет непрерывную функцию распределения, которая зависит только от чисел степеней свободы $v_1=n_1—1$ *и* $v_2= n_2— 1$*. F полностью определяется выборочными дисперсиями и не зависит от генеральных параметров, так как предполагают, что сравниваемые выборки, характеризуемые дисперсиями* $s^2{}_1$ *и* $s^2{}_2$*, взяты из генеральных совокупностей с* $\sigma^2{}_1 = \sigma^2{}_2$ *. Функция распределения возможных значений величины F по мере увеличения числа испытаний* ($n\to\infty$) *приближается к кривой нормального распределения.*

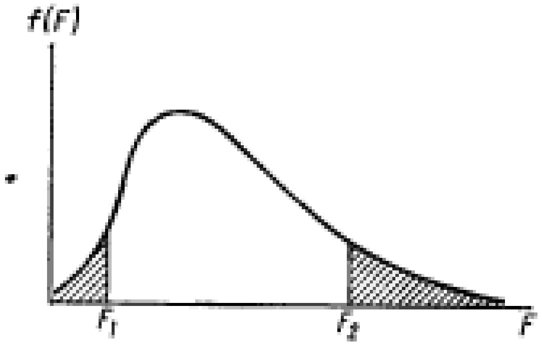


Рисунок. Вид кривой распределения вероятности величины F (f(F)) от его значений

Функция F-распределения может выть табулирована для принятых уровней значимости и чисел степеней свободы для большей дисперсии v_1 и v_2 для меньшей (на рисунке критический точки обозначены F1 и F2) . Критические точки для F-критерия приведены в приложении к лекции 3.

Достоинства дисперсионного анализа перед другими параметрическими методами проверки статистических гипотез.

Дисперсионный анализ имеет ряд преимуществ перед методом попарных сравнений по критерию Стьюдента:

1) вместо индивидуальных ошибок, средних по каждому варианту, в дисперсионном анализе используется обобщенная ошибка средних, которая опирается на большее число наблюдений и, следовательно, является более надежной базой для оценок;

2) методом дисперсионного анализа можно обрабатывать данные простых и сложных, однофакторных и многофакторных опытов или лабораторных экспериментов;

3) дисперсионный анализ позволяет избежать громоздких вычислений при большом числе вариантов и компактно в виде существенных разностей представить итоги •статистической обработки.

Одно и многофакторные дисперсионные комплексы. Равномерные и неравномерные комплексы.

В зависимости от типа и количества переменных, различают

однофакторный и многофакторный дисперсионный анализ (одна или несколько независимых переменных);

одномерный и многомерный дисперсионный анализ (одна или несколько зависимых переменных);

дисперсионный анализ с повторными измерениями (для зависимых выборок);

дисперсионный анализ с постоянными факторами, случайными факторами, и смешанные модели с факторами обоих типов;

Разложение дисперсий при дисперсионном анализе. Однофакторный дисперсионный анализ. Разложение дисперсий при простом однофакторном дисперсионном анализе. Разложение степеней свободы дисперсий при проведении простого однофакторного дисперсионного анализа. Однофакторный дисперсионный анализ с повторениями. Разложение дисперсий при однофакторном дисперсионном анализе с повторениями. Разложение степеней свободы дисперсий при проведении однофакторного дисперсионного анализа с повторениями.

При обработке однофакторных статистических комплексов, состоящих из нескольких независимых выборок, общая изменчивость результативного признака, измеряемая общей суммой квадратов Cy, расчленяется на два компонента: варьирование между выборками (вариантами) Cv и внутри выборок Cz.

$$Cy=Cv+Cz.$$

Общее число степеней свободы (N— 1) также расчленяется на две части — степени свободы для вариантов (l—1) и для случайного варьирования (N—l):

$N\text{-}l = (l\text{-}1)+(N\text{-}l).$

Если обрабатывают однофакторные сопряженные статистические комплексы, когда выборки (варианты) связаны каким-то общим контролируемым условием, например наличием n организованных повторений в полевом опыте, общая сумма квадратов разлагается на три части: варьирование повторений Ср вариантов Сv и случайное Cz. Соответственно разлагаются и степени свободы:

$$C_Y = C_P + C_V + C_Z;$$
$$(N-1) = (n-1) + (l-1) + (n-1)(l-1).$$

Проведение однофакторного дисперсионного анализа.

Этапы однофакторного дисперсионного анализа с повторениями.

1) *Определение общего числа наблюдений $N=l*n$;*
2) *Определение корректирующего фактора (поправки) $C = (\Sigma X)^2:N$;*
3) *Расчет общей сумму квадратов $C_Y = \Sigma X^2 — C$;*
4) *Расчет суммы квадратов для повторений $C_P = \Sigma P^2:l — C$;*
5) *сумма квадратов для вариантов $Cv = \Sigma V^2:n — C$;*
6) *сумма квадратов для ошибки (остаток) $CZ = C_Y — Cp — Cv$*

Расчет критерия Фишера как отношения дисперсий. Определение критических значений критерия Фишера. Определение значимости критерия.

Две последние суммы квадратов Cv и Cz делят на соответствующие им степени свободы. В результате получают два средних квадрата (дисперсии):

вариантов $s_V^2 = \frac{C_V}{l-1}$ и ошибки $s^2 = \frac{C_Z}{(n-1)(l-1)}$.

Эти средние квадраты и используют в дисперсионном анализе для оценки значимости действия изучаемых факторов. Оценка проводится путем сравнения дисперсии вариантов s^2_v *с дисперсией ошибки* s^2 *по критерию* $F = s^2_v/s^2$. *Если* $F > F_{st.}$, *то можно сделать вывод, что разница между вариантами проявляется на фоне случайной ошибки внутри вариантов.*

Вопросы для самоконтроля знаний:

- Общие принципы дисперсионного анализа.
- Как проводят разложение суммы квадратов при дисперсионном анализе?
- Как проводится оценка степени влияния изучаемого фактора?
- Назовите условия применимости дисперсионного анализа.
- Что представляет собой однофакторный равномерный дисперсионный комплекс?
- Что представляет собой однофакторный неравномерный дисперсионный комплекс?

Тема 7. Анализ многофакторных дисперсионных комплексов

Аннотация. В данной теме дается представление о принципах проведения многофакторного дисперсионного анализа. Рассматриваются виды многофакторных дисперсионных комплексов. Дается представление о методах проведения двухфакторного дисперсионного анализа

Ключевые слова: дисперсионный анализ, критерий Фишера, многофакторные дисперсионные комплексы.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где дается общее представление о принципах многофакторного дисперсионного анализа, видах многофакторного дисперсионного анализа. Рассматриваются методы проведения дисперсионного анализа и виды многофакторных дисперсионных комплексов.

- Для темы разработаны вспомогательные материалы (глоссарий, список сокращений и др.)
- Освоение темы предполагает выполнение задания к практическому занятию. Задание предполагает проведение самостоятельной работы – анализ многофакторного дисперсионного комплекса. Рекомендации для выполнения самостоятельной работы представлены внутри задания.
- Обучающийся должен оформить отчет о самостоятельной работе в виде файла конечных результатов и прикрепить этот файл к заданию
- Для проверки усвоения темы имеются вопросы для самоконтроля знаний.
- Для проверки усвоения темы необходимо пройти тест (**Тестирование является завершающей стадией освоения темы).**

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литературы: м электронные ресурсы:

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб. пособие для студентов вузов / В.Е. Гмурман.—12-е изд., перераб..—Москва: Высш. образование, 2007.—478, [1] с.

Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов..—4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351[1]с.

Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов..—М.: Изд. МГУ, 1970.—367с.

Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.

Дмитриев Е.А. Математическая статистика в почвоведении.М.: МГУ, 1995.291с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

Но нулевая гипотеза

Dy - сумма квадратов центральных отклонений (девиата) общая

Da - сумма квадратов центральных отклонений (девиата) по фактору a

Db - сумма квадратов центральных отклонений (девиата) по фактору b

Dab - сумма квадратов центральных отклонений (девиата) по двум факторам a и b

Dz - сумма квадратов центральных отклонений (девиата) общая ошибки

a — число градаций фактора A;

b — число градаций фактора B;

n — количество вариант в отдельных градациях комплекса;

na=na — общая численность вариант в каждой градации фактора A;

nb=nb — общая численность вариант в каждой градации фактора B;

ΣX_{α} — сумма вариант в градациях фактора A;

ΣX_{b}— сумма вариант в градациях фактора B.

Xi — варианты, входящие в состав дисперсионного комплекса;

$N=\Sigma n=n_{ab}$— общая численность вариант, или объем комплекса

k_y – степень свободы для общей дисперсии;

k_x - степень свободы для общей дисперсии;

k_x - степень свободы для межгрупповой дисперсии, характеризующей влияние обоих факторов А и В на результативный признак Х;

k_e- степень свободы для внутригрупповой, или остаточной, дисперсии;

k_a - степень свободы для факториальной дисперсии А;

k_b - степень свободы для факториальной дисперсии В;

k_{ab} - степень свободы для дисперсии совместного действия факторов А и В. для межгрупповой дисперсии, характеризующей влияние обоих факторов А и В на результативный признак Х;

Глоссарий

Дисперсионный анализ — метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях

Синерги́я (греч. συνεργία — сотрудничество, содействие, помощь, соучастие, сообщничество; от греч. σύν — вместе, греч. ἔργον — дело, труд, работа, (воз)действие) — суммирующий эффект взаимодействия двух или более факторов,

Антагони́зм (от др.-греч. ανταγωνισης — «противник», от др.-греч. αγων — «спор, борьба») — соперничество, конкуренция, борьба, противостояние, противоречия. Термин употребляется в различных областях

Аддитивность (лат. *additivus* — прибавляемый) — свойство величин, состоящее в том, что значение величины, соответствующее целому объекту, равно сумме значений величин, соответствующих его частям, в некотором классе возможных разбиений объекта на части. Например, аддитивность объёма означает, что объём целого тела равен сумме объёмов составляющих его частей.

Вопросы по теме для изучения:

Многофакторный дисперсионный анализ

Виды дисперсионных комплексов

Двухфакторный дисперсионный анализ

Краткий курс лекций:

Принципы многофакторного дисперсионного анализа. Многофакторный дисперсионный анализ

Ситуация, когда некоторое явление может полностью описываться одной единственной переменной, в природных объектах практически не встречаются. При проведении типичного эксперимента приходится иметь дело с большим количеством факторов. Основная причина, по которой использование дисперсионного анализа предпочтительнее повторного сравнения двух выборок при разных уровнях факторов с помощью серий t-критерия, заключается в том, что дисперсионный анализ существенно более эффективен и, для малых выборок, более информативен. Главное, что обеспечивает объективный результат дисперсионного анализа правильное составление дисперсионных комплексов.

Условия применения дисперсионного анализа

При применение дисперсионного анализа требуется выполнение трех важных условий 1) нормальность распределений по градациям факторов и 2) однородность (или гомогенность) дисперсий. 3) регулируемые (влияющие) факторы должны быть независимы друг от друга.

Многофакторные дисперсионные комплексы. Многофакторные дисперсионные комплексы. Равномерные, пропорциональные и неравномерные комплексы. Ортогональные и неортогональные дисперсионные комплексы.

Разложение степеней свободы при проведении многофакторного дисперсионного анализа.

Если одновременно исследуют действие на признак двух трех или большего числа регулируемых факторов, такой дисперсионный комплекс называют двух-, трех- и многофакторным. Числовые значения результативного признака, т. е. варианты или даты, могут распределяться по градациям комплекса равномерно, пропорционально и неравномерно. Равномерные и пропорциональные комплексы носят общее название ортогональные, а неравномерные комплексы называют неортогональными. В ортогональных комплексах осуществляется равенство квадратов центральных отклонений (девиат) $C_y=C_v+C_z$; *в неортогональных комплексах это равенство нарушается или* $C_y \neq C_v+C_z$; *Второй случай требует более сложных расчетов поэтому при планировании опытов предусматривающих применение дисперсионного анализа необходимо заранее планировать, что бы в градациях многофакторного комплекса были одинаковые или пропорциональные числа вариант.*

Разложение дисперсий при проведении многофакторного дисперсионного анализа. Условие равенства дисперсий для равномерных и пропорциональных (ортогональных) дисперсионных комплексов. Разложение степеней свободы при проведении многофакторного дисперсионного анализа.

Анализ многофакторных комплексов (по сравнению с однофакторными) не меняет, а лишь несколько усложняет общие схемы, поскольку наряду с действием каждого фактора в отдельности приходится учитывать и их совместное действие на результативный признак. В экспериментах с природными объектами часто эффект от совместного применения изучаемых факторов больше (синергизм) или меньше {антагонизм) суммы эффектов от раздельного применения каждого из них, определяется взаимодействием факторов: в первом случае положительным, а во втором — отрицательным. Если факторы не взаимодействуют, то прибавка от совместного применения *должна быть равна сумме прибавок от раздельного воздействия (аддитивность).*

Тогда сумма квадратов центральных отклонений Су содержит четыре компонента варьирования:

Су=Са + Сb+ +Саb+Сz,

Где, Са - варьирование по фактору а

Сb - варьирование по фактору b

Саb – взаимовлияние факторов

Сz – внутригрупповое варьирование.

Также можно разложить степени свободы:

$k_y=k_a+k_b+k_{ab}+k_z$

Проведение дисперсионного анализа. Расчет девиат. Отнесение девиат к соответствующим числам степени свободы. Определение значений дисперсий. Расчет отношений дисперсий. Устанавливление дисперсионных отношений F. Сравнение дисперсионных отношений с критическими точкам Fst. Оценка Но нулевой гипотезы.

1. Рассчитывают девиаты (как и при обработке однофакторных комплексов): общую для всего комплекса Су, межгрупповую Сх и остаточную Сz. Для этого служат можно использовать формулы:

$$Cy = \sum^{N} X_i^2 - H$$

$$Cx = \frac{\sum (X_i)^2}{n} - H$$

$$Cz = Cy - Cx$$

2. Затем определяют факториальные девиаты:

$$Ca = \frac{\sum^{a} (\sum X_a)^2}{n_a} - H$$

$$Cb = \frac{\sum^{b}(\sum X_b)^2}{n_b} - H$$

Девиату совместного действия факторов в обоих случаях определяют по формуле

$$Cab = Cx - (Ca + Cb)$$

В формулах используется вспомогательная величина, которая рассчитывается по формуле

$$H = \frac{(\sum X_i)^2}{N}$$

где Xi — варианты, входящие в состав дисперсионного комплекса; $N=\Sigma n=n_{ab}$— общая численность вариант, или объем комплекса; а — число градаций фактора А; b — число градаций фактора В; n — количество вариант в отдельных градациях комплекса; $n_a=na$ — общая численность вариант в каждой градации фактора А; $n_b=nb$ — общая численность вариант в каждой градации фактора В; ΣХа — сумма вариант в градациях фактора А; ΣXb— сумма вариант в градациях фактора В.

Слtдующий этап - установление чисел степеней свободы, которые равны: ky=N—1 для общей дисперсии; kx=ab—1 для межгрупповой дисперсии, характеризующей влияние обоих факторов А и В на результативный признак X; kz=N—ab для внутригрупповой, или остаточной, дисперсии; $k._a$ =a—1 для факториальной дисперсии А; k_b=Ь—1 для факториальной дисперсии В; kab=(a—1) (b—1)=ka.kb для дисперсии совместного действия факторов А и В.

Должны выполняться равенства, соответственно, ky==ka + kb + kab+kz, kx= ka + kb + kab и ky=kx+kz.

Отнесением девиант к соответствующим числам степени свободы определяют значения дисперсий, а по их отношениял. к величине остаточной дисперсии устанавливают дисперсионньи отношения F, которые сравнивают *с критическими точкам Fst для соответствующих чисел степеней свободы и принимаемого уровня значимости – Но гипотезу отвергают, если F>Fst. Соответственно можно рассчитать три критерия Fa Fb Fab и оценить их независимо.*

Вопросы для самоконтроля знаний

- Чем отличается двухфакторный бесповторностный дисперсионный анализ?
- В чем особенность двухфакторного дисперсионного анализа с повторностями?
- Что такое синергизм факторов?
- Что такое антогонизм факторов?
- Что такое аддитивность фактров?

Тема 8. Использование корреляционного анализа для исследования зависимостей

Аннотация. В данной теме дается представление о принципах проведения корреляционного анализа. Дается оценка различных методов корреляционного анализа, условия их применения. Рассматривается параметрический показатель – коэффициент корреляции Пирсона и непараметрический показатель – коэффициент корреляции Спирмена.

Ключевые слова. Корреляционный анализ, коэффициент корреляции Пирсона, коэффициент корреляции Спирмена

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где дается общее представление о принципах корреляционного анализа. Рассматриваются методы оценки корреляции

- Для темы разработаны вспомогательные материалы (глоссарий, список сокращений и др.)
- Освоение темы предполагает выполнение задания к практическому занятию. Задание предполагает проведение самостоятельной работы – корреляционный анализ однофакторных статистических зависимостей. Рекомендации для выполнения самостоятельной работы представлены внутри задания.
- Обучающийся должен оформить отчет о самостоятельной работе в виде файла конечных результатов и прикрепить этот файл к заданию
- Для проверки усвоения темы имеются вопросы для самоконтроля знаний.
- Для проверки усвоения темы необходимо пройти тест

(Тестирование является завершающей стадией освоения темы).

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литература: и электронные ресурсы:

Гиниятуллин К.Г. Решение задач корреляционного и регрессионного анализа в электронных таблицах MS EXCEL Методическое пособие Казань: Изд-во Казанский государственный университет им. В.И. Ульянова-Ленина, 2008. 32 с.

Акберова Н.И. Описательная статистика. Интервальные оценки. Учебно-методическое руководство и сборник задач к практическим занятиям по курсу «Математические методы в биохимии». – Казань: Типография издательского центра Казанского гос. ун-та, 2004. - 40 с.

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб.пособие для студентов вузов - 6-е изд., стер.. - М.: Высш. шк., 1998. - 479 с.

Дмитриев Е.А. Математическая статистика в почвоведении. - М.: МГУ, 1972. - 292с.

Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). - 5-е изд., доп. и перераб. - М.: Агропромиздат, 1985. – 351 с.

Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). - 5-е изд., доп. и перераб. - М.: Агропромиздат, 1985. - 351 с.

Лакин Г.Ф. Биометрия: учеб. пособие для биол. спец. Вузов - 4-е изд., перераб. и доп. - М.: Высш. шк., 1990. - 352 с.

Плохинский Н.А. Математические методы в биологиии. учебно-методическое пособие. - М.: Изд-во Моск. ун-та, 1978. - 265 с.

Словарь-справочник почвенно-экологических терминов/ Под.ред. Б.Ф.Апарина и А.И.Попова: учебное пособие. – СПб.: Изд-во С.-Петерб. ун-та, 2006. – 287 с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Список сокращений и обозначений:

r – линейный коэффициент корреляции (коэффициент корреляции Пирсона);

r_s- коэффициент корреляции рангов (коэффициент корреляции Спирмена);

r_{st} – расчетное стандартное значение коэффициента корреляции Спирмена

h – корреляционное отношение;

F – критерий нелинейности;

t_f – фактический критерий значимости;

t_{st} – стандартный критерий значимости

Глоссарий

Корреля́ция (от лат. *correlatio* — соотношение, взаимосвязь), **корреляционная зависимость** — статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми).

Корреляционный анализ — метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя или более переменными

Коэффициент детерминации (R^2 — *R-квадрат*) — это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости, то есть объясняющими переменными.

Вопросы для изучения по теме:

Основные понятия корреляционного и регрессионного анализа

Параметрические методы оценки связи между изучаемыми признаками

Однофакторная линейная корреляция и регрессия.

Непараметрические показатели связи.

Коэффициент корреляции рангов Спирмена

Конспект лекций:

Основные понятия корреляционного и регрессионного анализа

В исследованиях часто изучают изменчивость одновременно двух признаков x и y. Массив выборочных данных в этом случае образует n пар чисел $(x_1, y_1), \ldots, (x_n, y_n)$*. Методы статистического анализа позволяют с помощью выборки делать выводы о степени связи (мере связи) между данными изучаемыми признаками.*

Зависимость (связь) между признаками может быть функциональной или статистической. При функциональной зависимости каждому значению одного признака строго соответствует единственное значение другого признака. При наличии такой зависимости между переменными величинами, можно применить математическое понятие функции $y=f(x)$*. При изучении большинства природных объектов приходится иметь дело со статистической зависимостью, при которой изменение величины одного признака влечет изменение распределения второго признака. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин (x) изменяется среднее значение другой* $\bar{y} = f(x)$*; в этом случае статистическую зависимость между признаками называют корреляционной. При наличии корреляционной связи между признаками изменение результативного (зависящего) признака y (функции) при изменении одного или нескольких факториальных признаков x (аргументов) называется регрессией. Для оценки связи между изучаемыми признаками используются специальные статистические методы, называемые также корреляцией (корреляционным анализом) и регрессией (регрессионным анализом).*

По форме корреляция и регрессия может быть линейной (прямолинейной) и нелинейной (криволинейной), по направлению прямой и обратной. Корреляцию и регрессию называют простой (однофакторной), если исследуется связь между двумя факторами, и множественной (многофакторной), если изучается зависимость между тремя и более признаками. Зависимости могут оцениваться как, на количественном уровне, так и на качественном. При изучении почвенных объектов, как правило, имеют *дело с простыми зависимостями, которые оцениваются на количественном уровне, что закладывается обычно идеологией самих почвенных исследований, поэтому в данном пособии мы ограничимся рассмотрением только однофакторной корреляции и регрессии количественных признаков.*

Существуют два вида статистических показателей: параметрические, построенные на основе параметров совокупности (обычно, среднего значения $\bar{x}$ *и выборочной дисперсии* s^2_x*) и представляющие функции этих параметров, а также непараметрические, представляющие собой функции, зависящие от вариант совокупности с их частотами. Первые параметры применимы для совокупностей, распределяемых по нормальному закону, вторые - к любым совокупностям независимо от формы их распределения.*

Параметрические методы оценки связи между изучаемыми признаками. Однофакторная линейная корреляция и регрессия.

Под линейной корреляционной зависимостью между двумя признаками понимают такую зависимость, при которой одинаковые приращения аргумента x при любых его значениях должны вызывать одинаковые приращения величины функции y. Подобной зависимости соответствует уравнение регрессии $y=a+bx$*, на графике, построенном в прямоугольных координатах - прямая линия.*

Таблица, отражающая возрастающие значения аргумента и соответствующие им экспериментальные значения функции называется эмпирическим рядом регрессии (пример - табл. 1.), графическое изображение

Таблица 1.

Эмпирический ряд регрессии

Значения аргумента (фактора) x	*0*	*2*	*4*	*6*	*8*	*10*	*12*
Эмпирические значения	*4,6*	*16,6*	*28,3*	*33,9*	*41,2*	*56,5*	*69,2*

функции y							

n=7

течения функции линией регрессии (рис. 1., линия а).

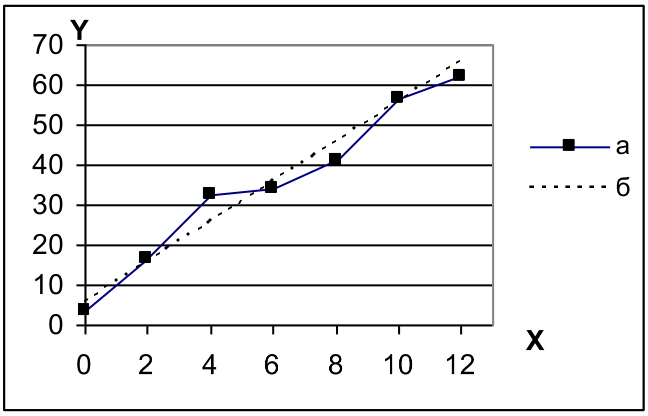


Рис.1. Ряды регрессии
а - эмпирическая линия регрессии по значениям табл. 1.
б - линия соответствующая уравнению регрессии Y=6,87+4,72 X

Эмпирическая линия регрессии представляет собой ломаную линию, что отражает статистический характер зависимости функции от аргумента в экспериментах. Процесс получения усредненного (теоретического) течения функции при равномерном увеличении аргумента

В качестве числового показателя простой линейной корреляции, показывающего тесноту (силу) и направления связи x с y, используют линейный коэффициент корреляции r (коэффициент корреляции Пирсона). Исходная формула расчета коэффициента линейной корреляции выглядит следующим образом:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$

где: xi и yi - значения пар признаков, и - средние значения.

Деление числителя формулы на знаменатель обеспечивает получение результата, как безразмерной величины варьирующей в пределах от -1,0 до1,0.

При наличии, прямой корреляционной связи между признаками, положительные отклонения (x) будут преимущественно сочетаться с положительными отклонениями (y), а отрицательные отклонения с отрицательными. Значение r будет больше 0. При обратной соответственно *меньше 0. Если изучаемые признаки независимы друг от друга, то каждое из значений (x) может сочетаться с любым значением (y) и сумма произведений будет стремиться к нулю.*

При расчете линейного коэффициента корреляции признаки можно менять местами, при этом соблюдаться условие $r_{yx}=r_{xy}$.

Величина $0<r<0,3$ – соответствует слабой прямой корреляции между признаками, $r = 0,3$-$0,7$ средней, $0,7>r>1,0$ сильной и $r =1,0$ полной. Соответственно, $-0,3<r<0$ – слабой обратной корреляции между признаками, $r = -0,3$ - $-0,7$ средней, $-0,7>r>-1,0$ сильной и $r=-1,0$ полной.

Коэффициент корреляции, как и любой другой выборочный показатель, имеет статистическую ошибку, которую можно выразить через показатель стандартная ошибка коэффициента корреляции (s_r)

$$s_r = \sqrt{\frac{1-r^2}{n-2}}$$

и критерий существенности (критерий проверки нулевой гипотезы)

$$t_f = r/s_r \geq t_{st}$$

Теоретическое (критическое) значение критерия t_{st} находят по таблице Стьюдента (Приложение 1), принимая 5% уровень значимости (B_1=0,95), обычный для почвенных исследований, или 1% уровень значимости при более строгом подходе (B_2=0,99). Допустимо для отдельных почвенных исследований использование 10% уровня значимости. Если фактическая величина t_f больше или равна теоретического значения t_{st} то корреляционная зависимость существенна (нулевая гипотеза отвергается), когда меньше – несущественна.

Правильное применение коэффициента корреляции предполагает нормальное распределение двумерной совокупности сопряженных величин x и y.

При наличии значительной корреляционной связи между переменными величинами r>0,5, выборочное распределение коэффициента корреляции для большого числа малых выборок, взятых из нормально распределяющейся генеральной совокупности, значительно отклоняется от нормальной кривой. Следовательно, выборочный коэффициент корреляции не будет точной оценкой генерального параметра, если он вычислялся из малой выборки и r>0,5. Для оценки значимости r, для малых выборок, можно использовать более точный способ оценки – z-преобразование Фишера. Способ сводится к замене r преобразованной величиной z.

$$z = 1{,}15129 \lg \frac{1+r}{1-r}$$

Критерий достоверности показателя z определяется по формуле $t_z = z\sqrt{n-3}$, *нулевая гипотеза отвергается при* $t_z \geq t_{st}$ *при заданном уровне значимости и числе степеней свободы k=n-2*

Непараметрические показатели связи. Коэффициент корреляции рангов Спирмена

Непараметрические показатели связи между изучаемыми признаками позволяют оценить ее направленность и характер при любом распределении признаков. Наиболее используемый непараметрический показатель - коэффициент корреляции Спирмена (называемый также коэффициентом корреляции рангов). Коэффициент корреляции Спирмена вычисляется по следующей формуле:

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

где: $d = R_x - R_y$ *- разность рангов между сопряженными значениями x и y;* n *– число сопряженных пар исследований (объем выборки).*

Как и другие выборочные показатели, коэффициент корреляции Спирмена является величиной случайной. Значимость определяется сравнением критерия r_s *с критическим значением* r_{st}*. Критерий* r_s *можно определить по формуле:*

$$r_{st} = \frac{t}{\sqrt{n-1}}\left(1 - \frac{m}{n-1}\right),$$

где: n *– объем выборки; при 1% уровне значимости вспомогательные величины* t *и* m *составляют 1,96 и 0,16; при 5% уровне значимости 2,58, и 0,69. Если абсолютное значение вычисленного коэффициента корреляции рангов* r_s *больше или равно критического значения* r_{st}*, то нулевая гипотеза отвергается.*

Вопросы для самоконтроля знаний

- Какие связи называют функциональные и корреляционные?
- Как рассчитывают коэффициент корреляции Пирсона?
- Как оценивается коэффициент корреляции Пирсона?
- Как рассчитывают коэффициент корреляции Спирмена?
- Как оценивается коэффициент корреляции Спирмена?

Тема 9. Использование регрессионного анализа для исследования зависимостей

Аннотация. В данной теме дается представление о принципах проведения регрессионного анализа. Дается оценка различных методов регрессионного анализа. Отдельно рассматривается однофакторная линейная и нелинейная регрессия. Рассматриваются области применения метода наименьших квадратов.

Ключевые слова. регрессионный анализ, линейная регрессия, нелинейная регрессия, метод наименьших квадратов.

Методические рекомендации по изучению темы

- Тема содержит лекционную часть, где дается общее представление о принципах регрессионного анализа. Рассматриваются методы оценки регрессии
- Для темы разработаны вспомогательные материалы (глоссарий, список сокращений и др.)
- Освоение темы предполагает выполнение задания к практическому занятию. Задание предполагает проведение самостоятельной работы – регрессионный анализ однофакторных статистических зависимостей. Рекомендации для выполнения самостоятельной работы представлены внутри задания.
- Обучающийся должен оформить отчет о самостоятельной работе в виде файла конечных результатов и прикрепить этот файл к заданию
- Для проверки усвоения темы имеются вопросы для самоконтроля знаний.
- Для проверки усвоения темы необходимо пройти тест (Тестирование является завершающей стадией освоения темы).

Рекомендуемые информационные ресурсы:

При составлении курса лекций использовалась следующая литература и электронные ресурсы:

Гиниятуллин К.Г. Решение задач корреляционного и регрессионного анализа в электронных таблицах MS EXCEL Методическое пособие Казань: Изд-во Казанский государственный университет им. В.И. Ульянова-Ленина, 2008. 32 с.

Акберова Н.И. Описательная статистика. Интервальные оценки. Учебно-методическое руководство и сборник задач к практическим занятиям по курсу «Математические методы в биохимии». – Казань: Типография издательского центра Казанского гос. ун-та, 2004. - 40 с.

Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб.пособие для студентов вузов - 6-е изд., стер.. - М.: Высш. шк., 1998. - 479 с.

Дмитриев Е.А. Математическая статистика в почвоведении. - М.: МГУ, 1972. - 292с.

Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). - 5-е изд., доп. и перераб. - М.: Агропромиздат, 1985. – 351 с.

Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). - 5-е изд., доп. и перераб. - М.: Агропромиздат, 1985. - 351 с.

Лакин Г.Ф. Биометрия: учеб. пособие для биол. спец. Вузов - 4-е изд., перераб. и доп. - М.: Высш. шк., 1990. - 352 с.

Плохинский Н.А. Математические методы в биологиии. учебно-методическое пособие. - М.: Изд-во Моск. ун-та, 1978. - 265 с.

Словарь-справочник почвенно-экологических терминов/ Под.ред. Б.Ф.Апарина и А.И.Попова: учебное пособие. – СПб.: Изд-во С.-Петерб. ун-та, 2006. – 287 с.

Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww Электронная библиотека МГУ -

http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

Сайт теория вероятностей и математическая статистика http://www.teorver.ru/

Сайт "Теория вероятностей" http://tever.ru/

Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/

w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)

Принятые сокращения и обозначения:

b – коэффициент регрессии

s_b – ошибка коэффициента регрессии

t_f – фактический критерий значимости;

t_{st} – стандартный критерий значимости

Глоссарий

Регрессио́нный анализ — статистический метод исследования влияния одной или нескольких независимых переменных $X_1, X_2, ..., X_p$ на зависимую переменную Y.

Мультиколлинеарность (*multicollinearity*) — в эконометрике (регрессионный анализ) — наличие линейной зависимости между независимыми переменными (факторами) регрессионной модели. При этом различают *полную коллинеарность*, которая означает наличие функциональной (тождественной) линейной зависимости и *частичную* или просто *мультиколлинеарность* — наличие сильной корреляции между факторами.

Независимая переменная — в эксперименте переменная, которая намеренно манипулируется или выбирается экспериментатором с целью выяснить ее влияние на зависимую переменную.

Зависимая переменная — в научном эксперименте измеряемая переменная, изменения которой связывают с изменениями независимой переменной.

Вопросы для изучения по теме:

Основные понятия корреляционного и регрессионного анализа

Параметрические методы оценки связи между изучаемыми признаками

Однофакторная линейная корреляция и регрессия.

Непараметрические показатели связи.

Коэффициент корреляции рангов Спирмена

Краткое изложение теоретического материала:

Понятие о регрессионном анализе.

Линейный коэффициент корреляции указывает на направление и степень сопряженности в изменчивости признаков, но не показывает, как количественно меняется результативный признак при изменении факториального на единицу измерения. Таким показателем является коэффициент линейной регрессии (b_{yx} и b_{xy}).

Однофакторная линейная регрессия

Уравнение однофакторной линейной регрессии может быть записано как:

$$y = \bar{y} + |b_{yx}(x - \bar{x}),\ \textit{тогда}$$

$$b_{yx} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}.$$

Ошибка коэффициента регрессии (s_b) вычисляется по формуле:

$$s_b = s_r\sqrt{\frac{\sum(y_i-\bar{y})^2}{\sum(x_i-\bar{x})^2}}$$

Критерий существенности (t_b):

$$t_b = {b_{yx}}/{s_b}$$

Стандартное значение критерия t_{st} определяют по таблице Стьюдента (Приложение к лекции 3) при степени свободы n-2, и заданном уровне значимости. Если $t_b \geq t_{st}$, то нулевая гипотеза отвергается.

Подставляем найденные значения в уравнение $y = \bar{y} + b_{yx}(x-\bar{x})$, *получаем y=35,2+4,72(x-6,0). Уравнение линейной регрессии примет вид y=6,87+4,72x. Данное уравнение соответствует теоретической линии регрессии (Рис.1, лекции 8). Линейное уравнение регрессии может быть получено графически, после ручного выравнивания эмпирического ряда регрессии (рис.1) в теоретическую прямую линию. В данном случае, значение a соответствует y при x=0, значение* $\mathrm{b} = \frac{y_i - a}{x_i}$.

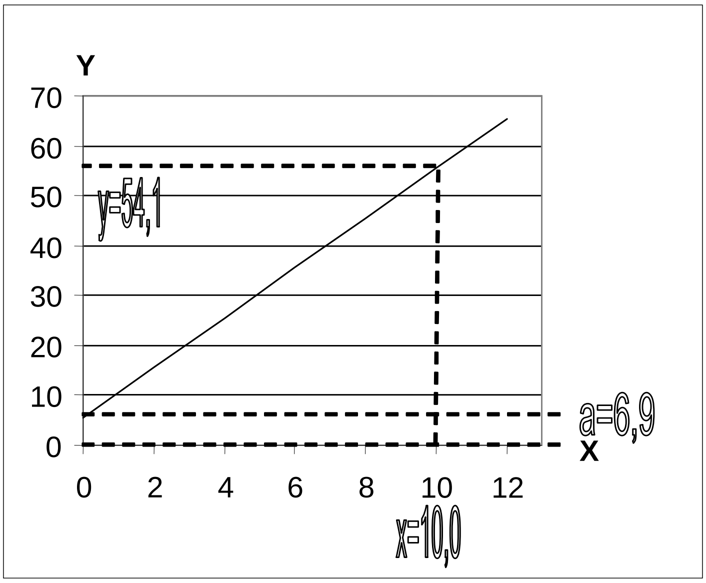


Рис. 1. Графическое определение линейного уравнения регрессии (a=6,9 при y=0, $b = \frac{54{,}1 - 6{,}9}{10{,}0} = 4{,}72$.

Поскольку показатели регрессии выражают корреляционную связь между признаками двухсторонне, уравнение регрессии может быть записано так же, как $x = \bar{x} - b_{xy}(y - \bar{y})$, для данного уравнения коэффициент регрессии будет вычисляться по формуле:

$$b_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum(y_i - \bar{y})^2},$$ при этом должно выполняться условие:

$$\sqrt{b_{yx} b_{xy}} = r.$$

Однофакторная нелинейная регрессия

Криволинейными называют зависимости, для которых равные приращения аргумента x вызывают неодинаковые приращения функции y,

Таблица 6.

Эмпирический ряд регрессии

Значения аргумента (фактора) x	*0*	*2*	*4*	*6*	*8*	*10*	*12*
Эмпирические значения функции y	*6,9*	*4,9*	*42,7*	*108,5*	*108,4*	*152,3*	*235,2*

n=7

например, эмпирический ряд регрессии, приведенный в таблице 6 и рисунке 2. Наличие корреляционной связи между признаками соответствующим криволинейной регрессионной зависимости может быть

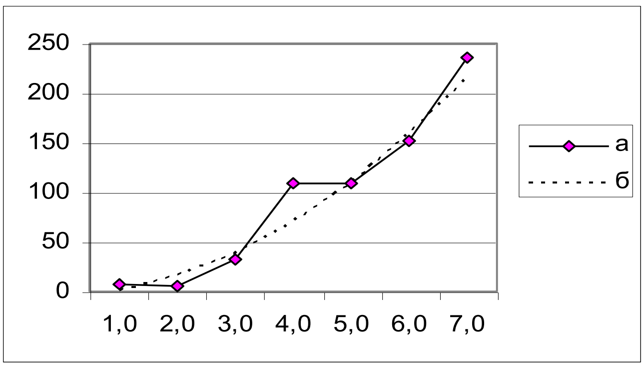


Рис.3. Ряды криволинейной регрессии
а - эмпирическая линия регрессии по значениям табл. 6.
б - линия соответствующая уравнению регрессии
$y=0{,}6452+6{,}2321x+1{,}0511x^2$

охарактеризовано через показатели линейного корреляционного анализа. Однако показатели линейной корреляции могут быть уверенно использованы для оценки связи между изучаемыми признаками, только в том случае если эта связь имеет не слишком значительную кривизну и достаточно близка к линейной. Критерии оценки линейности корреляционной связи рассматриваются в главе 2.4. Для эмпирического ряда приведенного в таблице 6 показатели линейной корреляции, рассчитанные в электронных таблицах 2б и 4б, составляют: $r=0{,}965$ $z=2{,}01$ $t_z=3{,}94$ ($t_{st}=4{,}03$). *Получаем тесную корреляционную связь между изучаемыми признаками, коэффициент корреляции существенен при 5% уровне значимости.*

Нелинейная регрессия может быть выражена с помощью ряда уравнений, наиболее широко используются:

уравнение параболы второго порядка – $y=a+вx+cx^2$;

уравнение параболы третьего порядка – $y=a+вx+cx^2+dx^3$;

уравнение гиперболы первого порядка – $y=a+\frac{b}{x}$;

уравнение параболы второго порядка – $y=a+\frac{b}{x^2}$;

уравнение параболы третьего порядка – $y=a+\frac{b}{x^3}$;

уравнение параболы первого порядка с тремя неизвестными –

$y=a+bx+\frac{c}{x}$;

уравнения экспоненциального или показательного типа – $y=a+b^x$ *или*

$y=a+e^{bx}$;

уравнение степенного типа – $y=a+x^b$.

Для эмпирического ряда таблицы 6 наиболее подходящим уравнением, максимально отвечающим течению ряда регрессии является уравнение параболы второй степени $y=a+вx+cx^2$. *Задача нелинейного регрессионного* *анализа, в данном случае, будет определяться нахождением коэффициентов регрессии (а, в и с). Для выполнения данной задачи наиболее удобным и точным методом является выравнивание эмпирических рядов методом наименьших квадратов. Метод основан на предположении, что сумма квадратов отклонений вариант* x_i *от среднего значения есть величина минимальная* $\sum_{i=1}^{n}(x_i - \bar{x})^2 = min$. *Тогда теоретические точки, соответствующие линии регрессии* $\bar{y}_x$, *должны быть получены таким способом, чтобы сумма квадратов отклонений была также минимальна* $\sum(y_i - \bar{y}_x)^2 = Q_{min}$.

Создается система нормальных уравнений:

$$\begin{cases} an + b\sum x + c\sum x^2 = \sum y \\ a\sum x + b\sum x^2 + c\sum x^3 = \sum xy \\ a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum x^2 y \end{cases}$$

Для эмпирических значений таблицы 6 система уравнений примет вид:

$$\begin{cases} a \cdot 7 + b \cdot 42{,}0 + c \cdot 364{,}0 = 648{,}9 \\ a \cdot 42{,}0 + b \cdot 364{,}0 + c \cdot 3528{,}0 = 6004{,}2 \\ a \cdot 364{,}0 + b \cdot 3528{,}0 + c \cdot 36400{,}0 = 60485{,}2 \end{cases} \quad (1)$$

Делим уравнения на числовое значение при коэффициенте уравнения регрессии а (7,0 42,0 364,0), получаем систему уравнений:

$$\begin{cases} a + b \cdot 6{,}000 + c \cdot 52{,}0000 = 92{,}7000 \\ a + b \cdot 8{,}6667 + c \cdot 84{,}0000 = 142{,}9571 \\ a + b \cdot 9{,}6923 + c \cdot 100{,}0000 = 166{,}1681 \end{cases} \quad (2)$$

Вычитаем первое уравнение из второго, второе из третьего, получаем:

$$\begin{cases} b\cdot 2{,}6667 + c\cdot 32{,}0000 = 50{,}2671 \\ b\cdot 1{,}0256 + c\cdot 16{,}0000 = 23{,}2110 \end{cases} \quad (3)$$

Делим каждое уравнение на число при коэффициенте регрессии b *(2,6667 и 1,0256), получаем:*

$$\begin{cases} b + c\cdot 12{,}0000 = 18{,}8464 \\ b + c\cdot 15{,}6000 = 22{,}6307 \end{cases} \quad (4)$$

Вычитаем первое уравнение из второго, получаем:

$\{c\cdot 3{,}6000 = 3{,}7843\}$ *(5), коэффициент с=1,0511.*

Подставляя значение коэффициента с в любое из уравнений, системы уравнений (4) получаем:

$\{b + 1{,}0511\cdot 12{,}0000 = 18{,}8464\}$*(6), коэффициент* b*=.6,2321.*

Подставляя в первое (исходное) уравнение системы уравнений (1) значение коэффициентов b *и с получаем:*

$\{a\cdot 7 + 6{,}2321\cdot 42{,}0 + 1{,}0511\cdot 364{,}0 = 648{,}9\}$ *(7), коэффициент а=0,6452.*

Уравнение регрессии будет выглядеть как: $\boldsymbol{y=0{,}6452+6{,}2321x+1{,}0511x^2}$ *(рис. 3.)*

Вопросы для самоконтроля знаний

- Понятие о регрессионном анализе.
- Прямолинейная регрессия
- Нелинейная регрессия

- Уравнения применяемые для аппроксимации нелинейных эмпирических рядов
- Значимость параметров уравнения регрессии и доверительная зона регрессии.
- Нахождение параметров уравнений связи.

Общий список сокращений и обозначений к ЭОР «Статистическая обработка результатов научных исследований»

- Выборка – выборочная совокупность.
- НЗР – наименьшая значимая разница, современное обозначение НСР принятая в некоторых программах.
- НСР – наименьшая статистически значимая разница
- Остаток - - сумма квадратов центральных отклонений общая ошибки

- a — число градаций фактора А;
- ANOVA –аббревиатура принятая в западной литературе для обозначения дисперсионного анализа.
- As – коэффициент асимметрии
- b – коэффициент регрессии
- b — число градаций фактора В;
- Cp - сумма квадратов центральных отклонений повторений
- Cv - сумма квадратов центральных отклонений вариантов
- Cy - сумма квадратов центральных отклонений общая
- Cz - сумма квадратов центральных отклонений общая ошибки
- Da - сумма квадратов центральных отклонений (девиата) по фактору a
- Db - сумма квадратов центральных отклонений (девиата) по фактору b Dab - сумма квадратов центральных отклонений (девиата) по двум факторам a и b
- Dz - сумма квадратов центральных отклонений (девиата) общая ошибки
- Dy - сумма квадратов центральных отклонений (девиата) общая
- Es – коэффициент эксцесса
- F – критерий нелинейности;
- F – критерий Фишера
- F – фактическое значение отношений дисперсий (критерия Фишера)

- Fst- стандартное (критическое, табличное) значение отношений дисперсий (критерия Фишера)
- h – корреляционное отношение;
- Ho нулевая гипотеза
- k_{ab} - степень свободы для дисперсии совместного действия факторов А и В. для межгрупповой дисперсии, характеризующей влияние обоих факторов А и В на результативный признак X;
- k_b - степень свободы для факториальной дисперсии В;
- k_e- степень свободы для внутригрупповой, или остаточной, дисперсии;
- k_x - степень свободы для межгрупповой дисперсии, характеризующей влияние обоих факторов А и В на результативный признак X;
- k_x - степень свободы для общей дисперсии;
- k_y – степень свободы для общей дисперсии;
- k_a - степень свободы для факториальной дисперсии А;
- l – градации фактора
- MAC - продукт Macintosh
- Me- медиана
- Mo - мода
- MS - продукт Microsoft Office
- n — количество вариант в отдельных градациях комплекса;
- n – количество повторностей
- N – общее число наблюдений
- na=na — общая численность вариант в каждой градации фактора А; nb=nb — общая численность вариант в каждой градации фактора В;
- r – линейный коэффициент корреляции (коэффициент корреляции Пирсона);
- r_s- коэффициент корреляции рангов (коэффициент корреляции Спирмена);
- r_{st} – расчетное стандартное значение коэффициента корреляции Спирмена

- s - стандартное отклонение,
- s^2 - дисперсия,
- s_b – ошибка коэффициента регрессии
- s_x % - относительная ошибка выборочной средней (
- s_x или m - ошибка средней арифметической ,
- t – критерий Стьюдента
- T- непараметрический критерий Уилкоксона
- t-тест – оценка значимости средних с помощью t-критерия Стьюдента
- t_f – фактический критерий значимости;
- t_f – фактический критерий значимости;
- t_{st} – стандартный критерий значимости
- t_{st} – стандартный критерий значимости
- U – непараметрический критерий Манн-Уитни
- V% - коэффициент вариации
- VBA - язык макропрограммирования Visual Basic for Application
- W- непараметрический критерий Ван-дер-Вардена
- Xi — варианты, входящие в состав дисперсионного комплекса;

  $N=\Sigma n=n_{ab}$— общая численность вариант, или объем комплекса
- ΣX_b— сумма вариант в градациях фактора B.
- ΣX_α — сумма вариант в градациях фактора A;
- τ - таута-критерий для выбраковки сомнительных дат
- χ^2 критерий кси-квадрат
- $\overline{X}$ или M - средняя арифметическая,

Общий глоссарий к ЭОР «Статистическая обработка результатов научных исследований»

- **Аддитивность** (лат. *additivus* — прибавляемый) — свойство величин, состоящее в том, что значение величины, соответствующее целому объекту, равно сумме значений величин, соответствующих его частям, в некотором классе возможных разбиений объекта на части. Например, аддитивность объёма означает, что объём целого тела равен сумме объёмов составляющих его частей.
- **Антагони́зм** (от др.-греч. ανταγωνισης — «противник», от др.-греч. αγων — «спор, борьба») — соперничество, конкуренция, борьба, противостояние, противоречия. Термин употребляется в различных областях
- **Асимметри́я** ***(от др.-греч. ασυμμετρία «несоразмерность», от др.-греч. μετρέω — «измеряю»)*** *— отсутствие или нарушение симметрии*.
- **Выборочная совокупность —** *часть генеральной совокупности элементов, которая охватывается наблюдением.*
- **Генеральная совокупность, генеральная выборка** ***(от лат. generis — общий, родовой) (в англ. терминологии — population) —*** *совокупность всех объектов (единиц), относительно которых учёный намерен делать выводы при изучении конкретной проблемы.*
- **Гистогра́мма** в математической статистике - это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.
- **Дисперсионный анализ (ANOVA)** — метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях
- **Дисперсионный анализ (многофакторный)** — метод направленный на поиск зависимостей в экспериментальных данных путём исследования

значимости различий в средних значениях с учетом двух или более независимых факторов.

- **Диспе́рсия случа́йной величины́** — *мера разброса данной случайной величины, то есть её отклонения от математического ожидания*
- **Зависимая переменная** — в научном эксперименте измеряемая переменная, изменения которой связывают с изменениями независимой переменной.
- **Зависимые (парные, сопряженные) выборки** - выборки в которых единицы наблюдения первой выборки связаны (сопряжены) каким-то общим условием с единицами наблюдения второй выборки.
- **Корреляционный анализ** — метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя или более переменными
- **Корреля́ция** (от лат. *correlatio* — соотношение, взаимосвязь), **корреляционная зависимость** — статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми).
- **Коэффициент детерминации** (R^2 — *R-квадрат*) — это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости, то есть объясняющими переменными.
- **Коэффицие́нт эксце́сса (коэффициент островершинности)** ***в теории вероятностей*** **—** *мера остроты пика распределения случайной величины*
- **Медиа́на (50-й перцентиль, квантиль 0,5) —** *возможное значение признака, которое делит ранжированную совокупность (вариационный ряд выборки) на две равные части:.*
- **Мо́да** — *значение во множестве наблюдений, которое встречается наиболее часто. (Мода = типичность.)*
- **Мультиколлинеарность** (*multicollinearity*) — в эконометрике (регрессионный анализ) — наличие линейной зависимости между независимыми переменными (факторами) регрессионной модели. При

этом различают *полную коллинеарность*, которая означает наличие функциональной (тождественной) линейной зависимости и *частичную* или просто *мультиколлинеарность* — наличие сильной корреляции между факторами.

- **Независимая переменная** — в эксперименте переменная, которая намеренно манипулируется или выбирается экспериментатором с целью выяснить ее влияние на зависимую переменную.
- **Независимые выборки** - выборки в которых единицы наблюдения первой выборки не связаны никаким общим условием с единицами наблюдения второй выборки;
- **Регрессио́нный анализ** — статистический метод исследования влияния одной или нескольких независимых переменных $X_1, X_2, ..., X_p$ на зависимую переменную Y.
- **Синерги́я** (греч. συνεργία — сотрудничество, содействие, помощь, соучастие, сообщничество; от греч. σύν — вместе, греч. ἔργον — дело, труд, работа, (воз)действие) — суммирующий эффект взаимодействия двух или более факторов,
- Сре́днее арифмети́ческое — одна из наиболее распространённых мер центральной тенденции, представляющая собой сумму всех зафиксированных значений, деленную на их количество.
- **Т-критерий Уилкоксона —** непараметрический статистический тест (критерий), используемый для проверки различий между двумя выборками парных измерений.
- **Файл** (англ.*file*)— именованная область данных на носителе информации.
- **Формат** — спецификация структуры данных, записанных в компьютерном файле. Формат файла иногда указывается в его имени, как часть, отделённая точкой.
- **Электронная таблица —** компьютерная программа, позволяющая проводить вычисления с данными, представленными в виде двумерных массивов, имитирующих бумажные таблицы. Некоторые программы

организуют данные в «листы», предлагая, таким образом, третье измерение.

- **F-тестом** или **критерием Фишера** (F-критерием, φ*-критерием) — называют любой статистический критерий, тестовая статистика которого при выполнении нулевой гипотезы имеет распределение Фишера (F-распределение).
- **Microsoft Excel** (также иногда называется Microsoft Office Excel) — программа для работы с электронными таблицами, созданная корпорацией Microsoft для Microsoft Windows, Windows NT и Mac OS.
- **t-тест –**оценка значимости средних с помощью t-критерия Стьюдента
- **t-критерий Стьюдента** — общее название для класса методов статистической проверки гипотез (статистических критериев), основанных на распределении Стьюдента.
- **U-критерий Манна — Уитни (англ. Mann — Whitney U-test) —** статистический критерий, используемый для оценки различий между двумя независимыми выборками по уровню какого-либо признака, измеренного количественно. Позволяет выявлять различия в значении параметра между малыми выборками. Другие названия: критерий Манна — Уитни — Уилкоксона (англ. *Mann — Whitney — Wilcoxon, MWW*), критерий суммы рангов Уилкоксона (англ. *Wilcoxon rank-sum test*) или критерий Уилкоксона — Манна — Уитни (англ. *Wilcoxon — Mann — Whitney test*). Реже: критерий числа инверсий

Общие информационные ресурсы ЭОР «Статистическая обработка результатов научных исследований»:

- Акберова Н.И. Описательная статистика. Интервальные оценки. Учебно-методическое руководство и сборник задач к практическим занятиям по курсу «Математические методы в биохимии»/ Н.И. Акберова; Казань: Типография издательского центра Казанского гос. ун-та, 2004. 40 с.
- Балахчев, Генрих Николаевич. Основы математической обработки и статистическое обоснование экспериментального материала: [методическое пособие для курса математическая статистика в почвоведении] / Г. Н. Балахчев, Г. Ф. Копосов; Казан. гос. ун-т, Биол.-почв. фак..Казань , 2007. 30 с.
- Википедия - свободная энциклопедия http://ru.wikipedia.org/wiki/ w.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=8#top)
- Гмурман, Владимир Ефимович. Теория вероятностей и математическая статистика: учеб.пособие для студентов вузов/ В.Е. Гмурман - 6-е изд., стер.. - М.: Высш. шк., 1998. 479 с.
- Дмитриев Е.А. Математическая статистика в почвоведении / Е. А. Дмитриев - М.: МГУ, 1972. 292с.
- Дмитриев Е.А. Математическая статистика в почвоведении. Е. А. Дмитриев, - М.: МГУ, 1995.291с.
- Дмитриев Евгений Анатольевич Математическая статистика в почвоведении. Изд. 3, испр. и доп. 2009. 328 с. (доступно на Интернет ресурсе http://ww Электронная библиотека МГУ - http://www.pochva.com/studentu/study/books/index.php?query=&by=author&format_search=d;

- Домашнее задание по некоторым простейшим задачам математической статистики: Для студентов физ.фак. / ; Казан.гос.ун-т;Сост.: Билялов Р.Ф.,Никитин Б.С..- Казань: Б.и., 1996. - 12с..
- Доспехов Борис Александрович. Методика полевого опыта: (С основами стат. обраб. результатов исслед.) [По агр. спец.] / Б. А. Доспехов; [Предисл. Д. В. Васильевой и др.].—5-е изд., доп. и перераб..—М.: Агропромиздат, 1985.—351 с.
- Лакин Г.Ф. Биометрия.: Учеб. пособие для студ. биолог. спец. вузов./ Г.Ф. Лакин —4-е изд., перераб. и доп..—М.: Высш. шк., 1990.—351[1]с.
- Плохинский Н.А. Биометрия.: Учеб. пособие для студ. биол. спец. ун-тов/ Н.А. Плохинский.—М.: Изд. МГУ, 1970. 367с.
- Попов, Владимир Александрович. Руководство к решению задач по теории вероятностей и математической статистике / В. А. Попов, М. Х. Бренерман; Казан. гос. ун-т, Физ. фак..?Казань: Изд-во Казанского государственного университета, 2008. 117, [2] с.: ил.; 20.?Библиогр.: с. 118 (9 назв.), 200.
- Решение задач корреляционного и регрессионного анализа в электронных таблицах MS EXCEL: методическое пособие к практическим занятиям дисциплины "Математические модели в почвоведении" / Казан. гос. ун-т, Биол.-почв. фак.; [сост. к.б.н., доц. К. Г. Гиниятуллин]. Казань: Изд-во Казанского государственного университета, 2008. 31 с.:
- Сайт теория вероятностей и математическая статистика http://www.teorver.ru/
- Салимов, Фарид Ибрагимович. Основы статистической обработки: учебное пособие для студентов очной и заочной формы обучения бюджетного и договорного отделения КГУ / Салимов Ф. И..?Казань: Казанский государственный университет, 2010.108 с.: ил.; 21.?Библиогр.: с. 6 (11 назв.), 520.
- Самсонова В.П., Мешалкина Ю.Л., Дядькина С.Е. Практикум на компьютере по курсу: "математическая статистика" М. изд-во МГУ, 2005

г. 150 с. (доступно на http://www.pochva.com/studentu/study/books/index.php?query=%20&by=author&format_search=d&n=25#top)

- Словарь-справочник почвенно-экологических терминов/ Под.ред. Б.Ф.Апарина и А.И.Попова: учебное пособие. – СПб.: Изд-во С.-Петерб. ун-та, 2006. 287 с.
- Теория вероятностей и математическая статистика /Б.А. Горлач. - М.: "Лань", 2013. - 320 стр. http://e.lanbook.com/view/book/4864/
- Теория вероятностей и математическая статистика: Учеб. пособие / Под ред. В.И. Ермакова. - М.: ИНФРА-М, 2004. - 287 с. http://znanium.com/bookread.php?book=76845
- Теория вероятностей и математическая статистика: Учебное пособие / С.В. Павлов. - М.: ИЦ РИОР: ИНФРА-М, 2010. - 186 с. http://znanium.com/bookread.php?book=217167

Вопросы к итоговой оценке знаний (зачету) по дисциплине «Статистическая обработка результатов научных исследований»:

1. Нормальный и логарифмически нормальный законы распределений варьирования.
2. t преобразование нормального закона. Распределение t величины.
3. Непараметрические методы проверки статистических гипотез. Использование критерия Манн-Уитни.
4. Проверка нормальности распределения. Критерий кси-квадрат.
5. Непараметрические методы проверки статистических гипотез. Использование критерия Вилкоксона.
6. Проверка нормальности распределения. Критерий Колмогорова-Смирнова.
7. Точечные оценки. Проверка гипотезы о равенстве среднего определенной величине.
8. Проверка нормальности распределения. Критерий Шапиро-Уилка.
9. Способы выбраковки сомнительных данных.
10. F-преобразование Фишера. Использование F критерия для оценки гипотезы о равенстве средних величин.
11. Парный двухвыборочный t-тест.
12. Основы дисперсионного анализа. Отношение дисперсий и F величина.
13. Распределение F величины.
14. Двухвыборочный t-тест независимых данных.
15. Особенности распределение дискретных величин.
16. Дисперсионный анализ. Многофакторные дисперсионные комплексы. Виды дисперсионных комплексов. Взаимовлияние факторов.
17. Регрессионный анализ. Уравнение регрессии. Эмпирическая и теоретическая линии регрессии. Коэффициенты регрессии и их смысл. Линейная регрессия. Статистическая значимость параметров регрессии и ее оценка.

18. Параметрические характеристики выборки. Среднее арифметическое (выборочное). Дисперсия. Стандартное отклонение. Ошибка среднего. Коэффициент вариации. Взвешенное среднее арифметическое.
19. Корреляционный анализ. Непараметрические методы корреляционного анализа.
20. Непараметрические характеристики выборки. Мода. Медиана. Квантили.
21. Дисперсионный анализ. Однофакторные дисперсионные комплексы. Равномерные и неравномерные комплексы. Анализ однофакторных дисперсионных комплексов
22. Асимметрия и эксцесс.
23. Корреляционный анализ. Параметрические методы корреляционного анализа.