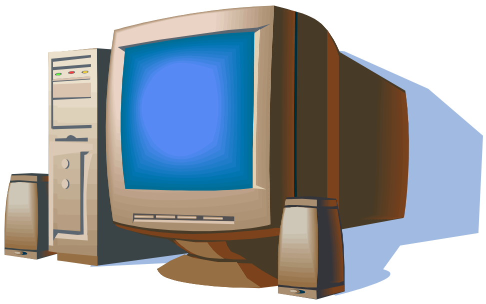
Булавко Г.И., Мусатова О.В.

БИОМЕТРИЯ

Методические указания к выполнению лабораторных работ по биометрии



УДК 573:001.8
ББК 28 в631 я73
Б 90

Авторы: доцент кафедры экологии и охраны природы УО «ВГУ им. П.М.Машерова», кандидат биологических наук ***Г.И.Булавко***
старший преподаватель кафедры экологии и охраны природы УО «ВГУ им. П.М.Машерова» ***О.В.Мусатова***

Рецензенты: профессор кафедры информатики УО «Витебский государственный технологический университет», кандидат технических наук ***В.Л.Шарстнев***
профессор кафедры зоологии УО «Витебский государственный университет им. П.М.Машерова», кандидат биологических наук ***С.И.Денисова***

Научный редактор: заведующий кафедрой экологии и охраны природы УО «ВГУ им. П.М.Машерова», кандидат биологических наук ***А.М.Дорофеев***

Учебно-методическое пособие подготовлено в соответствии с типовой учебной программой по курсу «Биометрия» для студентов, обучающихся по биологическим специальностям вузов. Рассматриваются наиболее часто используемые биометрические методы для характеристики свойств эмпирических совокупностей, обработки массовых экспериментальных материалов. Описаны возможности пакета анализа MS Excel для решения задач прикладной статистики.

Предназначено для студентов, обучающихся по биологическим специальностям вузов, учителям биологии и экологии, а также лицам, ведущим исследования в различных областях биологии.

УДК 573:001.8
ББК 28 в631 я73

Содержание

Введение

Системный подход при моделировании сложных явлений природы – одна из ведущих идей современного естествознания. В реализации этого подхода важное место занимают экспериментальные методы и методы многомерной статистики, которые позволяют исследователям выявить закономерности происходящих природных процессов, раскрыть причинно-следственные связи между элементами живой природы, сделать их доступными описанию точными математическими моделями.

При научных исследованиях биологических явлений наиболее эффективным является метод массовых наблюдений, использование которого предполагает проведение большого числа наблюдений. Собранный материал обрабатывают, анализируют, делают соответствующие выводы и устанавливают те или иные закономерности. Рассмотренный путь называется прямым индуктивным методом, когда от отдельных фактов переходят к общим положениям. Однако точность и достоверность результатов биологических экспериментов, равно как и корректность формулируемых выводов, зависят не только от качеств экспериментальных методик. Свойства самих биологических объектов и явлений сильно варьируют в пределах сообществ. Применение статистического анализа экспериментально полученных данных дополняет и углубляет познания биологических явлений природы, позволяет объективно оценить полученные результаты, нивелировав субъективизм исследователя и методические ошибки при постановке эксперимента, страхует экспериментатора от неточных и необоснованных выводов и заключений в отношении изучаемого явления.

Цель настоящих методических рекомендаций, описывающих стандартные и наиболее часто применяемые биометрические методы, – закрепить теоретические знания и приобрести опыт математической обработки данных наблюдения. Пособие включает практические расчетные работы, в которых отрабатываются основные математические методы обработки массовых экспериментальных материалов, повторяются ключевые теоретические моменты отдельных разделов курса «Биометрия», а также задания для самостоятельной индивидуальной работы. В пособии дается оценка методам математической статистики с точки зрения их возможностей и границ применения. При выполнении конкретных видов работ студенты могут использовать в качестве матриц экспериментальных данных результаты собственных исследований.

Отдельный раздел пособия посвящен возможностям решения задач прикладной статистики средствами Microsoft Excel. В практических работах приводятся примеры применения стандартных функций и пакета анализа данных MS Excel для решения конкретных задач с подробным описанием алгоритма решения. Приложение включает необходимые справочные данные для выполнения лабораторных работ.

При подготовке методических рекомендаций использован опыт других вузов, научная и методическая литература, основной список которой приводится.

Учебное пособие подготовлено для студентов биологических специальностей вузов широкого профиля, может быть с успехом использовано студентами научной специальности 1-33 01 01 «Биоэкология» при подготовке курсовых и дипломных работ, а также учителями – предметниками профильных классов и классов с углубленным изучением дисциплин естественно-научного цикла гимназий, лицеев при планировании учебной исследовательской работы школьников.

Лабораторная работа №1. БИОЛОГИЧЕСКИЕ ПРИЗНАКИ

Цель работы: освоить навыки выбора и измерения мерных и счетных признаков на различных биологических объектах.

Контрольные вопросы:

1. Биологические признаки, их классификация.
2. Ошибки наблюдений.
3. Понятие выборочной совокупности и генеральной совокупности.
4. Статистическая совокупность и ее свойства.
5. Требования к выборке.
6. Статистические таблицы.

Изучение биологических явлений проводится не по отдельным наблюдениям, которые могут оказаться случайными, нетипичными, неполно выражающими сущность данного явления, а на множестве однородных наблюдений, что дает более полную информацию об изучаемом объекте. Некоторое множество относительно однородных предметов или объектов, объединяемых по тому или иному признаку для совместного изучения, называют статистической совокупностью. Элементы, входящие в состав совокупности, называются ее членами, или в а р и а н т а м и (от лат. varians — изменяющийся). Варианты — это отдельные наблюдения или числовые значения признака. Так, если обозначить признак через X (икс большое), то его значения или варианты будут обозначаться через x (икс малое), т. е. как x_1, x_2, x_3, x_n и т.д. Общее число вариант, входящих в состав данной совокупности, называется ее объемом и обозначается буквой n.

Вместо сплошного обследования генеральной совокупности изучению подвергается обычно какая-то ее часть, получившая название выборочной совокупности, или в ы б о р к и .

Одной из наиболее распространенных форм группировок выборочных данных служат статистические таблицы. Статистические таблицы имеют иллюстративное значение, показывая какие-то общие итоги, положение отдельных элементов в общей серии наблюдений и т. д. Все это способствует пониманию описываемых явлений, выяснению того существенного, чем они отличаются друг от друга.

Задание:

1. На предложенных объектах провести выделение мерных и счетных биологических признаков и произвести их измерение и подсчет.
2. Произвести систематизацию и группировку полученных данных
3. Оформить статистические таблицы по образцу (табл.1)

Таблица 1. Результаты измерений признаков биологических объектов

№ п.п.	Признак			
	Высота растения	Масса растения	Количество листьев	Количество соцветий
1.				
2.				
3.				
….				
n				
Среднее значение				

Лабораторная работа №2. ВАРИАЦИОННЫЙ РЯД

Цель работы: получить навыки построения и графического изображения вариационного ряда статистической совокупности.

Контрольные вопросы:

1. Вариационный ряд значений.
2. Ранжирование (способы и значение).
3. Графическое выражение распределений
   - гистограмма
   - полигон
   - кумулята
   - огива.

Помимо статистических таблиц формой первичной группировки выборочных данных служит способ р а н ж и р о в а н и я, т. е. расположение вариант в определенном порядке — по возрастающим или убывающим значениям признака. При большом числе наблюдений ранжировать выборочную совокупность принято в виде двойного ряда, т. е. с указанием частоты или повторяемости отдельных вариант ранжированного ряда. Такой двойной ряд ранжированных значений признака называется в а р и а ц и о н н ы м рядом, или р я д о м р а с п р е д е л е н и я .

Для составления вариационного ряда необходимо установить число интервалов (классов), которое зависит от числа выборки (таблица 1).

Таблица 1. Число интервалов вариационного ряда по П.Ф.Рокицкому

Число наблюдений	25-40	41-60	61-100	101-200	≥201
Число интервалов	5-6	6-8	7-10	8-12	9-15

Число интервалов можно определить через логарифм численности выборки по формуле

$$k = 1 + 3{,}322\lg n, \text{ где}$$

k- число интервалов вариационного ряда;

n-численность выборки.

Величина классового интервала обозначается через i; она определяется по разности между максимальной и минимальной вариантами, отнесенной к избранному числу классов, т. е. по следующей приближенной формуле: $i = \dfrac{x_{\max} - x_{\min}}{k}$, где

i - классовый интервал, который берется целым числом;

x_{max} - максимальная и x_{min} — минимальная варианты выборки;

k - число классов, на которые разбивается выборочная совокупность.

Вычисленные величины интервалов округляют до целых чисел. При установлении пределов интервалов к минимальному значению изучаемого признака последовательно прибавляют принятую величину интервала до тех пор, пока не достигнута максимальная величина признака. Среднее значение интервала определяется как полусумма предельных значений данного интервала.

После того, как установлен классовый интервал, и выборочная совокупность разбита на классы, производится разноска вариант по классам, определяются частоты каждого класса.

Чтобы выразить эту закономерность более наглядно, принято изображать вариационные ряды графически в виде гистограммы, полигона, кумуляты или огивы. Гистограмма получается, если по оси абсцисс отложить границы классов, а по оси ординат - частоты классов вариационного ряда. Таким образом, гистограмма изображает распределение вариант при непрерывном варьировании признака. Прямоугольники соответствуют классам, а их высота — количеству вариант, заключенных в каждом классе.

Если из срединных точек вершин прямоугольников гистограммы опустить перпендикуляры на ось абсцисс, а затем эти точки соединить между собой, получится график прерывистого варьирования, называемый полигоном распределения. Если на оси абсцисс нанести классовые варианты, т. е. срединные значения классов, а на оси ординат — накопленные частоты классов, соединив затем соответствующие точки в системе координат, то получится график, называемый к у м у л я т о й.

Накопленные частоты классов получаются последовательным суммированием (кумуляцией) всех частот вариационного ряда в направлении от минимальной варианты до конца ряда. Если ряд накопленных частот нанести на ось абсцисс, а срединные значения классов — на ось ординат и построить соответствующий график, то он будет называться о г и в о й. Нетрудно понять, что огива есть не что иное, как кумулята, повернутая на 180^0.

Ниже приводятся примеры построения гистограммы, полигона распределения, кумуляты и огивы на примере измерений длины листа выборки одуванчика лекарственного (рисунок 1). Результаты измерений приводятся в таблице 2.

Таблица 2. Результаты измерений длины листа одуванчика лекарственного

Значения вариант	8,8	9,3	9,8	10,3	10,8	11,3	11,8	12,3	12,8	13,2	13,8
Частота	1	4	6	9	12	14	16	11	7	5	3
Накопленные частоты	1	5	11	20	32	46	62	73	80	85	88

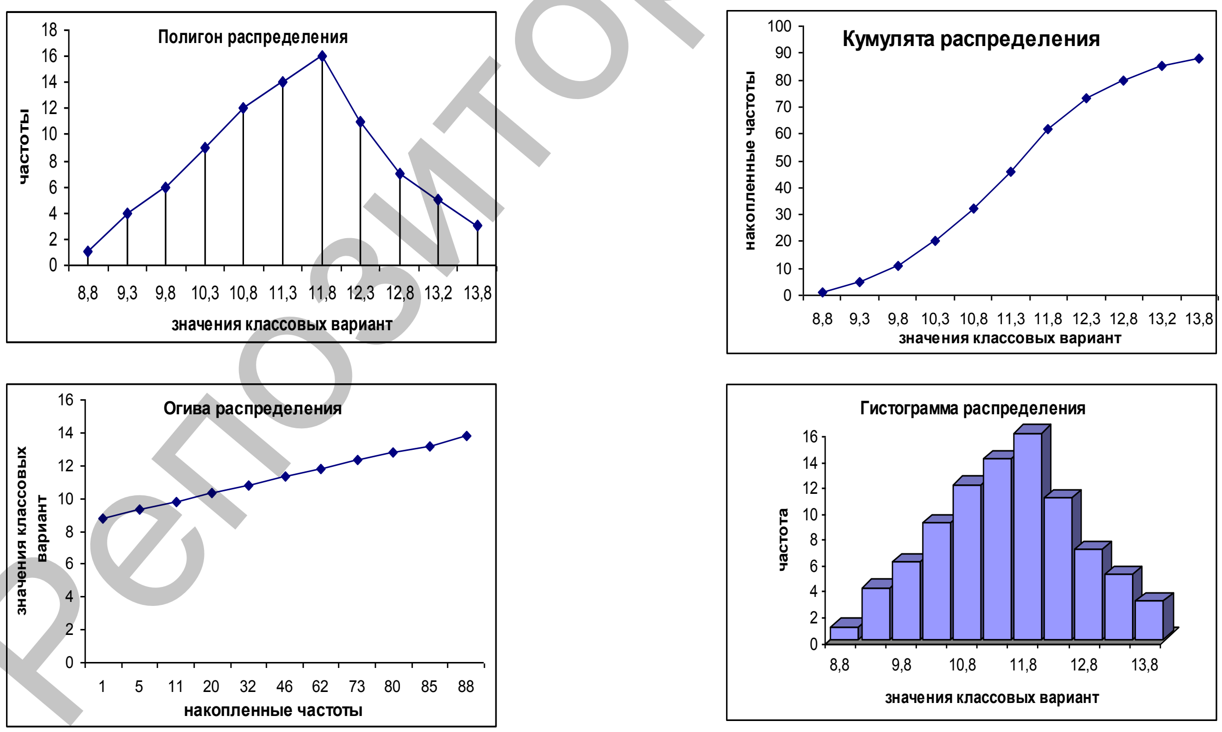


Рис. 2. Графическое изображение закономерностей распределения признаков

Задание:

1. Ранжировать данные измерений биологических признаков, полученные на предыдущем занятии.
2. Построить вариационный ряд значений.

3. Изобразить вариационный ряд значений в виде диаграммы, полигона распределения, кумуляты, огивы.

Лабораторная работа №3. СРЕДНИЕ ВЕЛИЧИНЫ И СПОСОБЫ ИХ ВЫЧИСЛЕНИЯ

Цель работы: получить навыки выбора вычисления средних величин на больших и малых выборках.

Контрольные вопросы:

1. Средняя арифметическая и способы ее вычисления.
2. Основные свойства средней арифметической.
3. Непараметрические средние
   - медиана
   - мода.
4. Средняя арифметическая в оценке качественных признаков.

Наиболее часто и широко как в практической деятельности человека, так и в научных исследованиях используется средняя величина. Она дает суммарную характеристику любого признака, указывая на то типичное и устойчивое в явлении, что наиболее полно выражает его содержание. Средняя арифметическая, которую принято обозначать через $\bar{X}$, или через М, есть не что иное, как частное от деления суммы всех вариант совокупности на их число, т. е. $\bar{x} = \frac{x_1 + x_2 + x_3 + \dots x_k}{n} = \frac{\Sigma x}{n} = \frac{1}{n}\Sigma x$ (1).

Это и есть общая формула средней арифметической, где x_1 x_2, x_3, ..., x_k обозначают варианты, входящие в состав данной совокупности; Σ — знак суммирования; n — общее число вариант, или объем выборочной совокупности. Средняя арифметическая - число именованное, она выражается теми же единицами меры или счета, что и характеризуемый ею признак.

При повторяемости отдельных вариант среднюю арифметическую можно представить как сумму произведений отдельных вариант на их частоты, отнесенную к общему числу всех вариант данной совокупности, т. е. как $\bar{x} = \frac{\Sigma xp}{n} = \frac{1}{n}\Sigma xp$. (2)

Существует упрощенный способ, позволяющий быстро и точно определять среднюю величину. Сущность этого способа очень проста: одну из вариант, все равно какую, условно принимают за среднюю арифметическую. Обычно в качестве условной средней берется варианта с большей частотой, хотя это совершенно не обязательно. Условную среднюю обозначим *А*. После выбора условной средней остается найти величину той поправки, которую нужно прибавить или отнять от условной средней, чтобы получить истинное значение средней арифметической данной совокупности. Эта поправка, называемая центральным моментом первого порядка, равна сумме произведений частот вариационного ряда на отклонения вариант от условной средней (А), отнесенной к числу всех вариант данного ряда.

Формула средней арифметической, вычисляемой по этому способу, принимает следующий вид: $\bar{x} = A + \frac{\Sigma pa}{n}$ (3), где

$\bar{x}$ - средняя арифметическая; А - условная средняя; *Σpa* — сумма произведений частот *(p)* на отклонения вариант от условной средней, *a* = *x*-А; n - объем выборки.

Преимущество этого способа более заметно на больших выборках, особенно в тех случаях, когда при вычислении средней арифметической по формуле (2) приходится перемножать многозначные числа.

Медиана - показатель описательного характера - не зависит от параметрических характеристик ряда. Она служит серединой вариационного ряда, который делит на две равные части: в обе стороны от медианы располагается одинаковое число вариант. Срединное значение вычисляется по формуле $Me = x_{\min} + i\frac{0{,}5N - \sum n}{n_e}$, где

x_{min} – минимальное значение предела интервала, где находится срединное значение;

i-величина интервала;

N-численность выборочной совокупности;

Σn-суммарная численность до интервала, в котором находится срединное значение;

N_e-численность интервала, где находится срединное значение.

Медиану можно определить и графически по кумуляте. Для этого из точки, лежащей на оси ординат и соответствующей 5-% всей численности выборки, проводится прямая, параллельная оси абсцисс, до пересечения с кумулятой. Перпендикуляр, опущенный из точки пересечения на ось абсцисс, укажет срединное значение признака.

Модой называется наиболее часто встречающаяся величина. В непрерывных вариационных рядах мода находится обычно в том классе, который имеет наибольшее число вариант. Этот класс называется модальным классом. Мода, как и медиана, служит вспомогательной характеристикой вариационного ряда. Правильное вычисление моды возможно в том случае, когда известен закон распределения значений статистической величины. Приближенно моду вычисляют, используя формулу Пирсона:

$$Mo = 3Me - 2M, \text{ где}$$

Me – медиана ряда распределения;

M-среднее значение признака.

В рядах симметричных (нормальных) среднее значение, медиана и мода совпадают. В действительных рядах такое совпадение маловероятно, так как фактическое распределение численностей биологических явлений природы отличается от нормального, имеет ту или иную кососТь.

Задание:

Для данных, полученных при измерении натуральных биологических объектов (занятие 1), определить:

- среднюю арифметическую с использованием всех вариантов вычисления4
- медиану;
- моду

Лабораторная работа № 4. ПОКАЗАТЕЛИ ВАРИАЦИИ И СПОСОБЫ ИХ ВЫЧИСЛЕНИЯ

Цель работы: получить навыки оценки степени вариации признаков на больших и малых выборках.

Контрольные вопросы:

1. Среднее линейное отклонение.
2. Дисперсия или варианса.
3. Среднее квадратическое отклонение.
4. Коэффициент вариации.
5. Ошибка средней арифметической.

Средняя арифметическая служит одной из важнейших характеристик вариационного ряда. Но она ничего не говорит о величине вариации характеризуемого признака. Величина вариации может быть оценена и по разности между максимальной и минимальной вариантами совокупности. Этот показатель получил название р а з м а х а вариации.

Средним линейным отклонением называют сумму отклонений вариант от средней арифметической, отнесенную к общему числу вариант данной совокупности: $\Delta = \frac{\sum(\bar{x} - x_i)}{n}$.

Среднее квадратичное отклонение, или основное отклонение - важнейшая характеристика вариационного ряда, являющаяся мерой рассеяния ряда распределения, показывает отклонение (для 68 случаев из 100) статистических величин от среднего значения. Определяется эта величина по формуле $S_x(\sigma_x) = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$, где

Σ -знак суммирования произведений отклонений вариант x_i от их средней x на веса или частоты p этих отклонений в пределах от первого до i-го класса;

n - общее число наблюдений ,или объем выборки.

(n -1) - число степеней свободы.

Дисперсия, или *варианса*, генеральной совокупности обозначается символом σ^2, а дисперсия выборки - S^2. В случае нормального распределения их величины совпадают.

Ошибка средней (m_x) является величиной, на которую отличается среднее значение выборочной (опытной) совокупности от среднего значения генеральной совокупности при условии, что распределение изучаемого признака приближается к нормальному. Основная ошибка среднего рассчитывается по формуле $m_x = \frac{\sigma}{\sqrt{n-1}}$. Среднее значение необходимо записать с основной ошибкой ($X \pm m_x$), только в этом случае можно судить о точности опыта.

Значение σ не всегда достаточно полно характеризует вариабельность рассматриваемой величины, особенно если приходится сравнивать изменчивость разных в количественном и качественно отношении признаков. Более информативным и удобным при сравнении различных статистических совокупностей является *коэффициент изменчивости*, или *вариации*, так как его величина не зависит от единиц, используемых при измерениях.

Коэффициент изменчивости – это основное отклонение, выраженное в процентах от среднего значения, которое рассчитывается по формуле $C = \frac{\sigma}{X} \times 100\%$.

На основании величины коэффициента изменчивости можно судить о характере и степени варьирования признака (таблица 1).

Таблица 1. Характер изменчивости признаков (по М.Л.Дворецкому, 1971)

Коэффициент изменчивости, С	до 5%	6-10%	11-20%	21-50%	более 50%
Характер изменчивости	слабая	умеренная	значительная	большая	очень большая

После вычисления того или иного статистического показателя необходимо проверить степень его надежности (достоверности) путем деления величины этого показателя на его ошибку. Достоверность среднего значения определяется по формуле $t = \frac{X}{m_x}$. Если значение t больше четырех, то среднее значение показателя является достоверным. Таким показателем можно пользоваться для сопоставления и формулировки корректных выводов. Часто о достоверности показателя судят, если $t \geq 3$. Если же t меньше трех, то по таким показателям нельзя делать категорические заключения или проводить сопоставления.

Важным показателем, характеризующим процент расхождения между выборочной и генеральной средними является *точность опыта (p,%),* или *ошибка наблюдений.* Эта величина характеризует субъективную ошибку исследователя. Ошибка выборки выражается в процентах от соответствующей средней: $p\% = \frac{m_x}{\bar{x}}100$. Точность опыта показывает, на сколько процентов можно ошибиться, если утверждать, что генеральная средняя равна полученной выборочной средней. В 68 случаях из 100 расхождение между выборочной и генеральной средними не будет превышать однократного значения точности опыта (в ту или иную сторону).

В некоторых экспериментах требуется очень высокая точность опыта (например, в медико-биологических, токсикометрических и др.), когда ошибка не должна превышать 1%. Рассчитанный по формуле процент ошибки необходимо сопоставить с заданным. Если он не выше заданного, то точность достаточная, а если выше, то точность результата является неудовлетворительной, необходимо увеличить число наблюдений.

Для определения объема выборки с заданной точностью опыта используют формулы: $n = C^2/p^2$ (если точность указана в процентах) и $n = \sigma^2/m_x^2$ (если точность дается в абсолютных величинах).

Задание:

Для данных, полученных при измерении признаков на натуральных объектах, вычислить:

- линейное отклонение
- дисперсию
- среднеквадратическое отклонение
- ошибку средней
- коэффициент вариации.

Проверить достоверность полученной средней, оценить точность опыта. Рассчитать оптимальный объем выборки для вычисления среднего значения с точностью 1%, 0,2 см (г, экз. и т.п.).

Лабораторная работа № 5. АСИММЕТРИЯ И ЭКСЦЕСС

Цель работы: получить навыки вычисления асимметрии и эксцесса

Контрольные вопросы:

1. Нормальное распределение.
2. Основные свойства нормального распределения.
3. Асимметрия.
4. Причины асимметрии.
5. Измерение асимметрии
6. Эксцесс и его измерение.

Наряду с практически симметричными распределениями встречаются и скошенные, асимметричные ряды. Аналитически они характеризуются нарушением равенства между модой, медианой и средней арифметической распределения. Графически они выражаются асимметричными кривыми распределения. Принято различать правостороннюю, или отрицательную, асимметрию и положительную или левостороннюю. В случаях правосторонней, или отрицательной, асимметрии варианты накапливаются преимущественно в правой части ряда; вершина такого ряда сдвинута вправо. В случае левосторонней асимметрии правая ветвь кривой, начиная от вершины, больше левой.

Пирсон предложил оценивать степень асимметрии по разности между средней арифметической и модой, отнесенной к величине среднего квадратического отклонения: $A_s = \dfrac{\bar{x} - M_o}{\sigma}$, где

A_s — мера скошенности рядов распределения, или коэффициент асимметрии.

В качестве показателя асимметрии также может служить утроенная разность между средней арифметической и медианой, отнесенная к величине среднего квадратического отклонения, т. е. $A_s = \dfrac{3(\bar{x} - M_e)}{\sigma}$.

Величина этого показателя обычно не выходит за пределы —3 и +3, что указывает на отрицательную или положительную асимметрию. При симметричном распределении коэффициент асимметрии равен нулю. Мера косости меньше 0,5 считается малой, от 0,5 до 1 – средней, выше 1 – большой.

Наиболее совершенным показателем асимметрии служит *центральный* момент третьего порядка, отнесенный к кубу среднего квадратического отклонения: $A_s = \sum pa^3 / n\sigma^3$, или $A_s = \dfrac{\Sigma p(\bar{x} - x_i)^3}{n\sigma^3}$.

Наряду с симметричными и скошенными распределениями вариационные ряды могут быть высоко- и плосковершинными, многовершинными, или эксцессивными. Это свойство распределения называется *эксцесс*. Величина эксцесса, обозначаемая знаком E_x, измеряется центральным моментом четвертого порядка, отнесенным к среднему квадратическому отклонению в четвертой степени $E_x = \dfrac{\Sigma pa^4}{n\sigma^4}$.

Для строго симметричных распределений эксцесс равен нулю. При положительном эксцессе показатель E_x - число положительное, а при отрицательном эксцессе - отрицательное. В обоих случаях коэффициент эксцесса - величина отвлеченная, не именованная.

При $E_x \le 0{,}2$ эксцесс практически отсутствует. Если же $0{,}5 \le E_x \le 1$, эксцесс считается заметным, но небольшим. Крайняя степень положительного эксцесса теоретически безгранична. Предельное значение отрицательного эксцесса равно -2, что указывает на наличие двух вариационных рядов, т. е. рядов с самостоятельными центрами распределения, объединенных в одной общей совокупности.

Задание:

- Для вариационных рядов значений исследуемых биологических признаков определить величину асимметрии и оценить степень скошенности распределения.
- Определить величину эксцесса, объяснить его причину.

Лабораторная работа № 6. НОРМИРОВАННОЕ ОТКЛОНЕНИЕ И ПОНЯТИЕ НОРМЫ

Цель работы: изучить особенности нормального распределения

Контрольные вопросы:

1. Нормальное распределение.
2. Основные свойства нормального распределения.
3. Статистические границы нормы.
4. Нормированное отклонение.

Наблюдения показывают, что большинство учитываемых признаков у человека, животных и растительных организмов распределяется по нормальному закону. Следовательно, общие статистические границы нормы дает критерий $x \pm 3\sigma$, поскольку все варианты практически симметричного распределения укладываются в эти пределы.

В теоретической и прикладной статистике большое значение имеет нормирование, позволяющее использовать среднее квадратическое отклонение для оценки отдельных вариант по отношению их к средней величине данной совокупности. Такого рода оценка производится по разности между вариантой и средней арифметической, отнесенной к величине среднего квадратического отклонения, т.е. $t = \frac{\bar{x} - x}{\sigma}$. Здесь t называется нормированным отклонением. Этот показатель удобен и прост как при оценке единичных вариант, так и при относительной характеристике сравниваемых друг с другом индивидов.

Разумеется, статистические границы нормы не могут быть очень жесткими: в зависимости от задачи исследования и различных обстоятельств их можно сузить до x: ±0,5σ или, наоборот, расширить до x±1σ. Варианты, которые распределяются между пределами от 0,67σ до 2σ, должны считаться субнормальными**.** Все же остальные члены, совокупности, выходящие за пределы x ± 2σ, следует отнести к категории аномалий

Задание:

Для исследуемых качественных и количественных признаков установить степень отклонения от нормы отдельных вариант.

Лабораторная работа №7. ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ

Цель работы: получить навыки определения достоверности эмпирических показателей.

Контрольные вопросы:

1. Ошибка отдельно взятой варианты.
2. Ошибка средней арифметической.
3. Ошибка среднего квадратического отклонения.
4. Ошибка коэффициента вариации.
5. Ошибки показателей асимметрии и эксцесса.
6. Оценка достоверности различий между дисперсиями

Расхождение между величиной средней арифметической (x) выборки и величиной средней арифметической генеральной совокупности (M) принято называть *ошибкой репрезентативности*, т. е. ошибкой, допускаемой не в самом процессе измерительной и вычислительной работы, а в результате случайного отбора вариант из генеральной совокупности при образовании выборки.

Если судить о величине статистической ошибки отдельно взятой варианты, то она равна среднему квадратическому отклонению, так как любое эмпирическое распределение, следующее нормальному закону, практически укладывается в пределах плюс - минус трех сигм, т. е. x±3σ. Ошибку репрезентативности называют, поэтому, средней квадратической ошибкой, или просто средней ошибкой. Будем ее обозначать через m, указывая при этом и характеристику, которую она сопровождает. Таким образом, средняя квадратическая ошибка отдельно взятой варианты выразится в виде $m_x=\pm\sigma$.

Выборочная средняя (x) отклоняется от своего математического ожидания или средней арифметической (M) генеральной (теоретически рассчитанной) совокупности меньше в √n раз по сравнению с отдельными вариантами данного распределения. Отсюда

$$m_x = \frac{\sigma}{\sqrt{n}}.$$

Поскольку весь вариационный ряд нормально распределяющейся случайной величины *X* практически укладывается в пределах между *x*-3σ и *x*-σ, то можно сказать, что генеральная средняя *(M)* таких распределений не выходит за пределы утроенного значения средней ошибки средней арифметической любой выборки, взятой из данной генеральной совокупности, т. е. она всегда заключена между пределами от x-3m_x до x+3m_x или в пределах $x \pm 3m_x$.Поэтому утроенное значение средней квадратической ошибки называется предельной ошибкой средней арифметической выборочной совокупности. А выражение $x \pm .3m_x$ заключает в себе содержание так называемого «правила утроенной ошибки».

При вычислении ошибки средней арифметической на малых выборках число наблюдений *(n)* берется «числом степеней свободы», и формула принимает следующий вид: $m_x = \frac{\sigma}{\sqrt{n-1}}$. Средняя ошибка среднего квадратического отклонения вычисляется по формуле $m_\sigma = \frac{\sigma}{\sqrt{2n}}$.Средняя ошибка коэффициента вариации (С) определяется по следующей приближенной формуле: $m_C = \frac{C}{\sqrt{2n}} \times \sqrt{1 + 2\left(\frac{C}{100}\right)^2}$

Средняя ошибка показателя асимметрии определяется по следующей приближенной формуле: $m_{A_s} = \sqrt{\frac{6}{n}}$, или более точно по формуле $m_{A_s} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$.

Ошибку коэффициента эксцесса можно вычислить по следующим аналогичным формулам: $m_{E_x} = 2\sqrt{\frac{6}{n}}$, или $m_{Ex} = \sqrt{\frac{24}{n}}$.

Как и в случаях сравнения выборочных средних, разность между двумя средними квадратическими отклонениями - σ_1 и σ_2 - оценивается путем нормирования, т. е. по критерию достоверности: $t = \frac{\sigma_1 - \sigma_2}{m_\sigma} = \frac{D_\sigma}{m_\sigma}$ где m_a - средняя квадратическая ошибка указанной разности. Она вычисляется по следующей формуле: $m_\sigma = \sqrt{\frac{\sigma_2^1}{2n_1} + \frac{\sigma_2^2}{2n_2}}$.

Задание: рассчитать для исследуемых признаков

- Ошибку отдельно взятой варианты.
- Ошибку средней арифметической
- Ошибку среднего квадратического отклонения
- Ошибку коэффициента вариации
- Ошибки показателей асимметрии и эксцесса.
- Достоверности различий между дисперсиями

Лабораторная работа №8. КРИТЕРИИ ДОСТОВЕРНОСТИ

Цель работы: получить навыки оценки достоверности полученных результатов.

Контрольные вопросы:

1. Достоверность различий между выборочными средними.
2. Достоверность различий между двумя дисперсиями.
3. Критерий соответствия между ожидаемыми и наблюдаемыми частотами.

Сравнительный анализ биометрических показателей сводится обычно к оценке степени достоверности наблюдаемых между ними различий. Оценка существенности или достоверности различий, наблюдаемых между двумя выборочными средними x_1 и x_2, производится на основе нормирования, т. е. отнесения разности между средними, которую обозначим через D, к средней квадратической ошибке этой разности:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{m_1^2 + m_2^2}} = \frac{D}{m_d}.$$

Здесь t называется критерием достоверности. Значение этого критерия оценивается по таблицам вероятности Стьюдента (приложение 1) на основании числа степеней свободы для заданного уровня вероятности: Р=0,95; Р=0,99; Р=0,999.число степеней свободы равно $u = n_1 + n_2 - 2$. Если фактическое t больше стандартного (табличного) t_{st} для данного уровня вероятности, различие существенное, достоверное и его нельзя объяснить случайными причинами.

Разница между показателями вариации двух независимых распределений оценивается с помощью критерия Фишера, обозначаемого через F и представляющего отношение дисперсий (варианс): $F = \sigma_1^2 / \sigma_2^2$. Числителем всегда берется большая варианса. Поэтому критерий F может быть равен 1 или больше ее. Если F=1, это указывает на равенство дисперсий. Когда же такого равенства нет, возникает необходимость оценить, случайно расхождение между дисперсиями или нет. Чем больше величина F, тем значительнее расхождение между дисперсиями, и наоборот, чем ближе значение F к 1, тем меньше расхождение между сравниваемыми показателями вариации. Р.А.Фишер получил значения F-критерия для различных уровней значимости и различного числа степеней свободы (приложение 2). Число степеней свободы равно численности выборки без единицы (n-1). Если фактическое значение критерия Фишера F будет больше стандартного (табличного), то различие дисперсий двух выборок доказано.

При сравнении наблюдаемых и ожидаемых результатов применяются особые критерии оценки, в частности критерий хи-квадрат (χ^2). Критерий предложен Карлом Пирсоном и представляет собой сумму отношений между квадратами разностей эмпирических и вычисленных или ожидаемых частот к ожидаемым частотам: $\chi^2 = \sum \frac{(p - p')^2}{p'}$, где Σ - знак суммирования, p – эмпирическая частота, p' – ожидаемая или теоретически вычисленная частота.

Использование χ^2-теста необходимо для того, чтобы узнать, подтверждается ли гипотеза экспериментом, т.е. насколько верны условия эксперимента, позволяют ли они с высокой степенью достоверности подтвердить или опровергнуть исходное предположение. Если бы фактические данные полностью совпадали с теоретическими, значение критерия было бы равно нулю. По мере увеличения разницы между этими показателями значение критерия будет возрастать. Каждому значению χ^2 соответствует определенная вероятность его появления (табличные данные, приложение 3). Значение χ^2 в таблице указывают те границы, до которых полученные значения критерия не дают оснований сомневаться в высказанном предположении с определенной степенью вероятности. Значений χ^2, превышающие табличные, будут указывать на несостоятельность гипотезы, т.е. признание того, что различие между фактическими и теоретически ожидаемыми результатами является достоверным, значимым.

Задание: для предложенных примеров произвести расчет критериев

- достоверности
- Фишера
- хи-квадрат

и оценить их величину.

Лабораторная работа №9. ДИСПЕРСИОННЫЙ АНАЛИЗ

Цель работы: получить навыки изучения статистического влияния одного или нескольких факторов на результативный признак.

В практике нередко возникает необходимость в оценке целых комплексов количественных показателей -необходимость сравнивать между собой одновременно не две, а несколько выборок, объединенных в единый комплекс. Статистический анализ целого комплекса требует особого метода, который был разработан Р. А. Фишером (1925) и получил название дисперсионного анализа.

Ход анализа

Методика дисперсионного анализа сводится к некоторой общей схеме, которую можно свести к следующим 6 пунктам:

1. Собранные данные упорядочиваются, сводятся в таблицу в соответствии с условиями опыта. Затем определяется общая вариация для всего комплекса, равная сумме квадратов отклонений отдельных вариант от общей средней для всего комплекса $(\bar{x}_s)$. Будем обозначать общую вариацию через a_0, ее удобнее определять по следующим аналогичным формулам: $a_0 = \sum x^2 - n\bar{x}_s^2$, или $a_0 = \sum x^2 - \frac{1}{x}(\sum x)_s^2$, где

Σx^2- сумма квадратов вариант, входящих в состав всего комплекса;

$\bar{x}$ - средняя арифметическая комплекса, т. е. общая средняя выборочной совокупности;

n — общее число наблюдений.

2.Определяется межгрупповая вариация (a_1): $a_1 = N\sum\bar{x}^2 - \frac{1}{n}(\sum x)^2$ или $a_1 = N\sum\bar{x}^2 - n\bar{x}^2$, где

где N — число вариант в группах;

$\bar{x}$ - средняя арифметическая групп;

x - значение отдельной варианты (остальные обозначения см. выше).

3.Определяется внутригрупповая, или остаточная, вариация по разности $a_2 = a_0 - a_1$

4.Находится значение дисперсий - межгрупповой и внутригрупповой, для чего соответствующие вариации относятся к числу степеней свободы в группах $(n$-1) и к числу степеней свободы внутригрупповой вариации $(n — k)$, т.е $\sigma_1^2 = \frac{a_1}{N-1}$ и $\sigma_2^2 = \frac{a_2}{n-k}$.

5.Берется отношение дисперсий $F = \frac{\sigma_1^2}{\sigma_2^2}$.

6.Критерий F оценивается по таблицам Фишера для соответствующих степеней свободы и взятого уровня значимости. При этом, когда $F \leq 1$, следует брать отношение большей варианcы к меньшей, т. е. оценивать $F = \frac{\sigma_2^2}{\sigma_1^2}$ и предельное значение F для большего числа степеней свободы находить по столбцам, а для меньшего - по строкам таблицы. Если найденная величина F при заданном уровне значимости и данных числах степеней свободы превышает значение этого показателя, указанное в таблице, то различия, наблюдаемые между групповыми средними, достоверны. В противном случае, т. е. когда критерий F меньше своего предельного значения, эти различия носят случайный характер и не могут быть признаны существенными, достоверными.

Задание: провести дисперсионный анализ для предложенных вариантов.

Пример задания для расчета

Варианты опыта	Урожай в кг по повторностям			Средний урожай, кг
	1	2	3	
Контроль	21,2	28,0	31,2	26,80
1. Удобрения помещались ниже семян на 3-5 см	23,6	22,5	28,0	24,73
2. удобрения помещались в стороне от семян на 3-5 см	24,0	30,0	29,2	27,73

3. Удобрения помещались выше заделки семян на 3-5 см	29,2	28,0	27,0	28,07

Лабораторная работа №10. КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Цель работы: получить навыки оценки степени связи между признаками.

Контрольные вопросы:

1. Суммарный показатель связи.
2. Функциональная зависимость и корреляция.
3. Коэффициент корреляции.
4. Понятие о регрессии.
5. Построение эмпирических рядов регрессии.
6. Уравнение регрессии.
7. Коэффициенты регрессии.

Отличительной чертой биологических объектов является многообразие признаков, характеризующих каждый из них. Например, организмы можно характеризовать возрастом, весом, размерами и т.д. При этом описываемые признаки часто бывают взаимообусловлены. Например, чем старше организм, тем большими размерами он характеризуется. В простейшем случае связь между переменными строго однозначна. Например, вес древесины одного вида полностью определяется ее объемом. Такого рода зависимость называют *функциональной*, когда каждому значению независимой переменной соответствует только одно значение зависимой. Для биологических объектов редко связь между их характеристиками бывает менее «жесткой»: объекты с одинаковыми значениями одного признака имеют, как правило, разные значения по другим признакам. Такую связь между вариациями разных признаков называют *корреляцией*. По взаимонаправленности связь может быть прямой – когда с увеличением значений одного признака в общем увеличиваются значения другого, и обратной – когда с увеличением значений одного признака значения другого признака уменьшаются.

По форме связь может быть прямолинейной (линейной) и криволинейной (нелинейной). Основным мерилом связи, существующей между биологическими, признаками, служит коэффициент *корреляции*. Он показывает степень приближения корреляционной связи к функциональной (для которой всегда равен единице) и колеблется в пределах от минус (для обратной связи) до плюс (для прямой связи) единицы. Значение коэффициента корреляции, равное нулю или близкое к нулю, говорит лишь об отсутствии прямолинейной связи, но не указывает на наличие или отсутствие криволинейной связи, которая при этом может быть тесной.

Рабочая формула, по которой обычно вычисляется коэффициент корреляции во всех случаях, когда варианты не группируются по классам: $r = \dfrac{\sum a_x a_y}{n\sigma_x\sigma_y}$.

В числителе этой формулы стоит сумма произведений отклонений вариант от средней арифметической по одному ряду *(X)* на соответствующие отклонения вариант от средней арифметической по другому ряду (У) т.е. $a_x = x_x - \bar{x}_x$ и $a_y = x_y - \bar{x}_y$; $\sum a_x a_y = \sum (x_x - \bar{x}_x)(x_y - \bar{x}_y)$.

По величине коэффициента корреляции можно установить характер связи (таблица).

Коэффициент корреляции	Теснота связи
до 0,30	слабая
0,31-0,50	умеренная
0,51-0,70	значительная
0,71-0,90	высокая
0,91 и более	очень высокая

Квадратическая ошибка коэффициента корреляции $m_r = \pm\frac{1-r^2}{\sqrt{n}}$

Значение коэффициента корреляции оценивается с помощью критерия достоверности Стьюдента: $t = \frac{r\sqrt{(N-2)}}{\sqrt{1-r^2}}$ или с помощью формулы $t = r/m_r$.

При оценке степени взаимосвязи статистических величин важно провести математическое моделирование, т.е. подобрать аналитическое уравнение, которое соответствовало бы природе изучаемого явления с целью предсказания поведения независимой характеристики объекта при изменении зависимого параметра. Динамика взаимной зависимости между переменными величинами получила название *регрессии*, а методика исследования регрессии носит название *регрессионного анализа*.

Ряды регрессии выражаются не только графически, но и аналитически при помощи следующих уравнений: $Y' = \bar{y} + Ry/x(x-\bar{x})$ -уравнение регрессии *Y* по *X;* $X' = \bar{x} + Rx/y(y-\bar{y})$ - уравнение регрессии *X* по *Y*.

Здесь *Y'* и *X'* — теоретические, т. е. вычисленные по эмпирическим данным, значения регрессии *Y/X* и *X/Y;* $\bar{y}$ и $\bar{x}$ - средние арифметические рядов распределения Y и X; R — коэффициент регрессии, который определяется по следующим аналогичным формулам:

$Ry/x = \frac{\sum a_x a_y}{\sum a_x^2}$ -коэффициент регрессии *Y.X;*

$Rx/y = \frac{\sum a_x a_y}{\sum a_y^2}$ коэффициент регрессии *X.Y,*

$a_x = x - \bar{x}$ и $a_y = y - \bar{y}$.

Когда известны средние квадратические отклонения варьирующих признаков, т. е. σ_x и σ_y, а также вычислен коэффициент корреляции (r), коэффициенты регрессии определяются по формулам: $Ry/x = r \cdot \frac{\sigma_y}{\sigma_x}$ и $Rx/y = r \cdot \frac{\sigma_x}{\sigma_y}$.

Коэффициент регрессии сопровождается средней квадратической ошибкой, которая вычисляется то формулам: $m_{R_{y/x}} = \frac{\sigma_{y/x}}{\sqrt{\sum a_x^2}}$ и $m_{R_{x/y}} = \frac{\sigma_{x/y}}{\sqrt{\sum a_y^2}}$

Задание:

- Рассчитать коэффициенты корреляции и регрессии для биологических признаков, измеренных на 1 занятии (длина листа и высота растения, длина и ширина листа, высота и вес растения).
- Оценить степень связи между признаками

Лабораторная работа № 11. РЕШЕНИЕ ЗАДАЧ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ СРЕДСТВАМИ MS EXCEL

Цель работы: освоить применение стандартных функций MS Excel для решения задач описательной статистики.

Контрольные вопросы:

1. Ввод исходных данных.
2. Вычисление размаха (вариации),
3. Оценка среднего, среднеквадратичного отклонения и дисперсии, асимметрии и эксцесса.

4. Построение таблицы частот и гистограммы.

Пример 1. Проведите анализ данных в рамках описательной статистики с использованием средств **Вставка функций** и **Мастер диаграмм** MS Excel.

1. Запустите MS Excel: **Пуск/Программы/ MS Excel** и переименуйте ярлычок рабочего листа **Лист 1**: двойной щелчок по ярлычку и напечатайте поверх выделения **Статистика 1**. введите исходные данные и заголовки статистической таблицы по **Образцу 1:** выделите ячейку А1 щелчком мыши/введите текст заголовка и зафиксируйте щелчком по инструменту Enter √/расположите заголовок по центру столбцов А-Е – выделите ячейки А1:Е1 и щелкните инструмент **Объединить и поместить в центре**/ аналогично выделяя последовательно ячейки А2 – Е11 введите числа исходных данных таблицы:

х, прирост населения в 50 городах				
27	36	34	46	34
28	29	37	41	43
40	33	50	37	41
32	29	43	34	32
30	43	54	42	47
35	49	49	54	36
36	51	36	24	35
25	33	38	38	36
29	51	32	36	53
30	55	44	46	38

В ячейках G2 – G14 заголовки строк статистической таблицы и число выборок:

Среднее
Среднеквадратичное отклонение (СТАНДОТКЛОН)
Дисперсия (ДИСПА)
Медиана (МЕДИАНА)
Мода (МОДА)
Асимметрия (СКОС)
Эксцесс (ЭКСЦЕСС)
Наименьшее (МИН)
Наибольшее (МАКС)
Кол-во выборок 50

Выполните расчеты указанных в заголовках строк статистической таблицы параметров, вставляя при помощи средства **Вставка функций** расчетные формулы. Например, для расчета среднего выделите щелчком мыши ячейку Н2, щелкните инструмент **Вставка функций**/в окне **Мастер функций** в поле **Категории** щелкните **Статистические,** в поле **Функция** при помощи полосы прокрутки пролистайте список названия функций, найдите и щелкните **СРЗНАЧ** и **ОК**/ в окне вставки функции справа от поля **Число1** щелкните кнопку сворачивания/ выделите мышью диапазон ячеек А2:Е11 (удерживая левую кнопку мыши)/ в свернутом окне вставки функции щелкните кнопку разворачивания/ОК; аналогично вставьте остальные формулы.

2. Сформируйте таблицу частот исследуемой величины, выполнив группировку данных и расчеты непосредственным вводом формул и при помощи средства **Вставка функций**:

– вставьте формулу для вычисления минимального числа интервалов группирования при помощи средства **Вставка функций:** выделите ячейку А14 и введите «мин. кол-во интервалов», выделите ячейку В14/инструмент **Вставка функций**/ в поле **Категории** щелкните **Математические**/ в поле **Функция** найдите и выберите **ОКРУГЛ** и **ОК**/ в окне вставки функции установите курсор в поле **Число разрядов** и введите 0 (округление до целого числа), установите курсор в поле **Число** и введите 5* (множитель)/ в инструменте выбора функции (левый верхний угол рабочей книги) щелкните кнопку списка и выберите позицию **Другие функции…**/в окне **Мастер функций** выберите категорию **LOG10** из категории **Математические** и **ОК**/ в окне вставки функции в поле **Число 1** введите ссылку с числом выборок **Н14** и **ОК**.

– вставьте формулу для расчета ширины интервала при помощи ввода с клавиатуры: выделите ячейку А15 и введите «ширина интервала», выделите ячейку В15/введите знак =(равно) и знак **(**(скобка)/щелкните ячейку с максимальным значением Н10 и нажмите клавишу **F4** для перехода к абсолютной ссылке/ введите знак - /щелкните ячейку с минимальным значением Н9 и нажмите клавишу F4/введите знак **)** (скобка) и знак **/** (наклонная черта) и щелкните ячейку В14 с числом интервалов/ **Enter.**

– аналогично в ячейки А20-А27 вставьте формулы для вычисления правых границ интервалов: щелкните ячейку А20, введите знак =(равно)/щелкните ячейку с минимальным значением Н9 и нажмите клавишу **F4** для перехода к абсолютной ссылке/ введите знак + (плюс) и щелкните ячейку с значением ширины интервала В15/ **Enter;** в ячейку А21 введите формулу =А20+ $В$15; в ячейки А22 и ниже растяните формулу из ячейки А20 при помощи автозаполнения: после ввода формулы в А21 укажите на нижний правый угол ячейки А21 до появления маркера автозаполнения в форме +, нажмите левую кнопку мыши и, удерживая ее, протяните выделение ячейки до А27 и отпустите кнопку мыши.

– Вставьте формулу для расчета частот с применением функции массивов: выделите диапазон ячеек В20-В27/ инструмент **Вставка функций**/найдите и выберите функцию **ЧАСТОТА** из категории **Статистические** и **ОК**/ в окне вставки функции справа от поля **Массив данных** щелкните кнопку сворачивания/ выделите мышью диапазон ячеек А2:Е11/щелкните кнопку разворачивания/справа от поля **Массив интервалов** щелкните кнопку сворачивания/ выделите мышью диапазон ячеек А20:А27/щелкните кнопку разворачивания/одновременно нажмите клавиши **Ctrl, Shift, Enter** для фиксации функции массива.

3. Постройте гистограмму для исследуемой величины с применением мастера диаграмм: выделите диапазон ячеек с таблицей частот А20:В27/инструмент **Мастер диаграмм**/на вкладке **Нестандартные** в поле **Тип** выберите **График/Гистограмма 2** и кнопка **Далее**/в окне **Исходные данные** на вкладке **Диапазон данных** включите переключатель **в столбцах**/на вкладке **Ряд** щелкните кнопку сворачивания справа от поля **Подписи по оси X**/ выделите диапазон ячеек А20:А27 и щелкните кнопку разворачивания/ в поле **Подписи второй оси X** внесите диапазон ячеек В20:В27 и кнопка **Далее**/в окне размещение диаграммы включите переключатель **на имеющемся листе** и **ОК**.

Пример 2. Выполните процедуру генерации случайных чисел и проанализируйте их с помощью средств **Анализ данных** и **Мастер диаграмм** MS Excel.

1. Перейдите на свободный рабочий лист книги и переименуйте его в **Генерация данных**.
2. Подключите надстройку **Пакет анализа** MS Excel: **Сервис/Надстройки**/в окне Надстройки установите флажок **Пакет анализа** и **ОК**.
3. Выполните генерацию 30 случайных чисел, распределенных в соответствии с нормальным законом с нулевым средним и дисперсией 1: щелкните ячейку А1 и **Сервис/Анализ данных**/в поле со списком **Инструмент анализа** щелкните позицию **Генерация случайных чисел** и **ОК**/в поле **Число переменных** введите 1, в поле **Число случайных чисел** введите 30, раскройте список поля **Распределение** и выберите позицию **Нормальное**, введите в полях **Среднее** – 0, **Стандартное отклонение** – 1, в разделе **Параметры вывода** включите переключатель **выходной интервал**, щелкните кнопку сворачивания/щелкните ячейку А1 и кнопку разворачивания, **ОК**.
4. Измените разрядность данных, уменьшите число знаков после запятой до двух: выделите диапазон ячеек А1:А30/щелкните инструмент **Уменьшить разрядность** четыре раза.
5. Выполните процедуру описательной статистики по сгенерированным данным: **Сервис/Анализ данных/Описательная статистика** и **ОК**/в окне Описательная статистика в поле Входной интервал введите ссылку на диапазон ячеек А1:А30/в разделе

Группирование включите переключатель **по столбцам** и уберите флажок **Метки в первой строке**/в разделе **Параметры вывода** включите переключатель **Выходной интервал** и щелкните ячейку С1/установите флажок **Итоговая статистика** и **ОК.**

6. Постройте гистограмму по данным столбца А: **Сервис/Анализ данных/Гистограмма** и **ОК**/ в окне Гистограмма в разделе **Входные данные** в поле **Входной интервал** введите ссылку на диапазон ячеек А1:А30 и установите флажок Метки/ в разделе параметры вывода включите переключатель **Выходной интервал** и укажите любую свободную ячейку рабочего листа/установите флажок **Интегральный процент** и **Вывод графика** и **ОК**. Добавьте на построенную гистограмму 2 линии тренда: полиноминального со степенью 4 и скользящего среднего на 2 точки: щелкните правой клавишей мыши по рядам значений, на появившемся контекстном меню выбрать позицию **Добавить линию тренда/**на вкладке тип выбрать **Полиноминальная**, в поле **степень** ввести 4 и т.д.

Лабораторная работа №12. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ, АНАЛИЗ ФАКТОРОВ В MS EXCEL

Цель работы: освоить применение стандартных функций MS Excel для решения задач анализа связей, применение пакета анализа MS Excel для решения задач анализа связей.

Контрольные вопросы:

1. Построение диаграммы рассеяния.
2. Расчет корреляции с помощью стандартных функций MS Excel.
3. Однофакторный дисперсионный анализ средствами пакета анализа MS Excel.

Пример 1. Проведите визуальный анализ данных и расчет коэффициента корреляции в рамках задачи проверки наличия связи между двумя переменными с использованием средств **Вставка функций** и **Мастер диаграмм** MS Excel.

1. Запустите MS Excel и переименуйте ярлычок рабочего листа в **Диаграмма рассеяния.**

2. Сформируйте массив исходных данных результатов измерений длины первого молярного x и второго молярного y зубов у ископаемого млекопитающего по образцу:

x	y
10,7	11,2
10,8	10,9
10,6	10,5
10,7	9,6
10,1	11,2
11,2	11,3
11,4	12,2
12,1	12,1
12,3	11,7
12,0	11,0
12,3	13,2
12,7	13,0
12,9	12,2
12,8	13,4
13,1	12,6
13,3	12,2
13,3	12,0
13,4	11,2
12,7	11,4
12,5	11,4

3. Используя средство **Мастер диаграмм**, постройте диаграмму рассеяния: выделите диапазон ячеек **А1:В21**/инструмент **Мастер диаграмм/**тип диаграммы **Точечная** вида 1 и **Далее**/на вкладке **Диапазон данных** установите флажок **Ряды в столбцах** и **Далее**/введите заголовки диаграммы **Диаграмма рассеяния**, оси Х – **первый молярный**, оси У – **второй молярный**/снимите отображение легенды/установите отображение основных и промежуточных линий сетки по обеим осям/расположите диаграмму на имеющемся листе/добавьте линейный тренд с включением отображения уравнения (вкладка **Параметры** окна Л**иния тренда**).

4. Проанализируйте полученные данные и сделайте предварительный вывод о наличии линейной связи между рассматриваемыми признаками.

5. Рассчитайте коэффициент корреляции для исследуемых выборок. Выделите ячейку **А22** и введите текст **Коэффициент корреляции**/выделите ячейку **В22**/щелкните инструмент **Вставка функций**/в окне **Мастер функций** в поле **Категории** щелкните **Статистические,** в поле **Функция** при помощи полосы прокрутки пролистайте список

названия функций, найдите и щелкните **КОРЕЛЛ** и **ОК**/в полях **Массив1** и **Массив2** введите последовательно ссылки на диапазоны **А1:А21** и **В1:В21** соответственно и **ОК**.

6. Проверьте значимость полученного значения коэффициента корреляции по критерию Стъюдента. Выделите ячейку **А23** и введите текст **t-статистика**/выделите ячейку **В23** и введите формулу по образцу =**В22*КОРЕНЬ(СЧЕТ(А2:А21)-2)/КОРЕНЬ(1-В22*В22)**. Выделите ячейку **А24** и введите текст **Критическое значение**. Выделите ячейку **В24** и введите формулу по образцу **=СТЪЮДРАСПОБР (0,05;СЧЕТ(А2:А21)-2).** Если значение **t** больше табличного (критического), то принимается наличие значимой линейной связи (отвергается предположение об отсутствии связи).

<u>Пример 2.</u> Проведите визуальный и корреляционный анализ данных в рамках задачи проверки наличия связи между двумя переменными с использованием средств **Анализ данных** и **Мастер диаграмм** MS Excel.

1. Перейдите на **Лист 2** книги и переименуйте его ярлычок в **Корелл анализ**. Скопируйте предыдущий пример на новый лист: выделите диапазон ячеек **А1:В21** листа **Диаграмма рассеяния**/щелкните правой кнопкой мыши для вызова контекстного меню/**Копировать**/перейдите на новый рабочий лист/выделите ячейку **А1**/щелкните правой кнопкой мыши и **Вставить**.

2. Рассчитайте корреляцию для исследуемых данных, используя группу **Корреляция** средства **Анализ данных**: **Сервис/Анализ данных**/в поле со списком **Инструмент анализа** щелкните позицию **Корреляция** и **ОК/**в поле **Входной интервал** введите ссылку на диапазон **А1:В21**/установите флажки в **Группирование по столбцам** и **Метки в первой строке**/ в поле выходной интервал укажите ссылку на ячейку D1 и **ОК**.

<u>Пример 3.</u> Проведите однофакторный дисперсионный анализ влияния одного фактора на характеристики нескольких экспериментальных групп с использованием средств **Анализ данных** MS Excel.

1. Перейдите на **Лист 3** книги и переименуйте его ярлычок в **Однофакт анализ**.

2. Сформируйте массив исходных данных лабораторных испытаний влияния фактора погоды на изменение продолжительности систолической остановки сердца при введении хлорида бария для четырех групп экспериментальных животных по образцу:

	A	B	C	D	E
		Показатели экспериментальных групп			
	Погода	1	2	3	4
	Тихая погода	13,8	11	13,7	12,1
	Ветер и вьюга	16	12,2	15,8	14,3

3. Запустите процедуру однофакторного дисперсионного анализа: **Сервис/Анализ данных/Однофакторный дисперсионный анализ**/укажите диапазон входных значений **В3:Е4**, группирование **по столбцам**, флажок **Метки** снимите, укажите ячейку выходного диапазона ячеек **В6**.

4. Проанализируйте полученные результаты: сравните дисперсию внутри групп (характеризует влияние случайной составляющей) и между группами (характеризует влияние изучаемого фактора – погоды). Если они значимо отличаются (уровень значимости Р=0,05), то фактор считается оказывающим статистически значимое влияние на исследуемую переменную. Сравните расчетное F и критическое значения статистики Фишера. Отличие считается значимым, если расчетное значение больше критического.

5. Сформируйте дополнительную расчетную формулу для принятия решения **Влияет ли фактор?** по образцу. Выделите ячейку Е23 и введите текст **Влияет ли фактор?**/щелкните ячейку F23 и введите формулу =ЕСЛИ(F18>H18;«да»;«нет»).

Лабораторная работа №13. ПРОВЕРКА ГИПОТЕЗ В MS EXCEL. ПАРАМЕТРИЧЕСКИЕ И НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ

Цель работы: освоить применение стандартных функций MS Excel для решения задач проверки гипотез, применение пакета анализа для решения задач проверки гипотез.

Контрольные вопросы:

1. Нулевая гипотеза.
2. Эмпирический тест на нормальность.
3. Проверка гипотезы о равенстве среднего заданному значению.
4. Проверка гипотезы о распределении по критерию хи-квадрат.

Пример 1. Проведите анализ данных в рамках задачи проверки гипотезы о распределении при помощи эмпирического теста на нормальность с использованием средства **Вставка функций** MS Excel.

1. Запустите MS Excel и сохраните созданную при запуске книгу под именем **Примеры гипотезы.**
2. Переименуйте ярлычок рабочего листа в **Тест норм**.
3. сформируйте массив исходных данных результатов 100 замеров отклонений от номинального размера объекта по образцу 1 в диапазоне ячеек **A1:J11**:

Образец 1

x (норм)									
48	39	43	44	34	34	32	43	40	46
25	31	34	49	39	37	45	48	41	49
43	46	34	35	42	32	41	34	42	42
38	40	46	47	34	42	38	40	38	36
30	43	41	40	40	35	35	41	38	45
37	42	38	36	44	39	332	48	43	39
43	30	44	36	42	34	49	49	49	51
37	30	50	48	44	35	45	34	33	41
43	45	44	34	33	39	41	39	46	31
40	52	45	39	35	45	33	42	42	36

4. Рассчитайте среднее и среднеквадратичное отклонение, разместив расчетные формулы в ячейках **L12, L13** соответственно: =СРЗНАЧ(A2:J11) и =СТАНДОТКЛОН (B2:J11).
5. Рассчитайте массив отклонений выборочных значений от среднего: активизируйте ячейку **A15**/инструмент **Вставка функций**/в поле **Категории** выберите **Математические**/ в поле **Функция** найдите и выберите **ABS** и **ОК**/в окне вставки функции установите курсор в поле **Число,** введите ссылку **A2**/введите знак разности «-» /введите ссылку **L12**, перейдите к абсолютной ссылке клавишей **F4** и **ОК**/используя маркер автозаполнения ячейки **A15**, растяните формулу в ячейки **B15:J15**/не снимая выделение с диапазона **A15:J15**, используя маркер автозаполнения выделенного диапазона, растяните формулу в ячейки **A16:J24**.
6. Сформируйте таблицу проверки условий эмпирического теста на нормальность и вставьте расчетные формулы согласно Образцу 2 в ячейки **L15:L22:**

Образец 2

	L	M
15	Значение 3S	=3*L13
16	Значение 0,625 S	=L13*0,625
17	Число выборок	100
18		
19	Условие	Выполняется ли условие
20	<3S	=ЕСЛИ(СЧЕТЕСЛИ(A15:J24;"<16,52")>0,997*100; "да";"нет")
21	<S	=ЕСЛИ(СЧЕТЕСЛИ(A15:J24;"<5,508")>0,683*100; "да";"нет")
22	<0,625S	=ЕСЛИ(СЧЕТЕСЛИ(A15:J24;"<3,44")>0,5*100; "да";"нет")

7. Проинтерпретируйте полученные результаты: в случае невыполнения хотя бы одного из условий эмпирического теста необходима дополнительная проверка исходной гипотезы о нормальности при помощи, например, критерия хи-квадрат. При выполнении всех трех условий гипотеза о нормальном законе распределения принимается.

Пример 2. Проведите анализ данных в рамках задачи проверки гипотезы о распределении при помощи критерия согласия хи-квадрат с использованием средств **Вставка функций** и **Анализ данных** MS Excel.

1. Перейдите на рабочий лист 2 книги **Примеры гипотезы** и переименуйте его в **Тест хи-квадрат**. В ячейке **А1** введите заголовок столбца данных **х, норм**.

2. Скопируйте исходные данные из диапазона ячеек листа 1 **A2:J11** в позицию начиная с ячейки **A2**.

3. Реорганизуйте скопированный массив данных на листе при помощи приема перемещения диапазонов ячеек так, чтобы данные располагались в одном столбце **А**: результирующий массив должен занимать диапазон ячеек **A2:A101**.

4. При помощи средства **Сервис/Анализ данных** рассчитайте по исходным данным описательную статистику и постройте таблицу частот и гистограмму (см. лабораторную работу №. 11) в диапазоне ячеек **C1:G12**.

5. Используя построенную таблицу частот и рассчитанные среднее и среднеквадратичное, а также стандартные и встроенные функции, сформируйте таблицу для расчета статистики хи-квадрат по образцу 3. Обратите внимание на ввод максимального значения вместо текста **Еще** в исходной таблице частот (ячейка **C29**) и расчетной формулы для дополнительного значения интервала группирования в ячейке **C30**. Эти действия необходимы для корректного вычисления теоретических частот. В столбце скорректированных частот выполнено объединение тех карманов, где значение частот менее 3. это условие правильного применения критерия. Расчетная формула для числа степеней свободы распределения хи-квадрат определяется разностью числа карманов (с учетом их объединения) и числа 3 (количеству налагаемых связей плюс 1).

Образец 3

	C	D	E	F	G
18	карман	частота	скоррект. частота	теоретич. частота	хи-квадрат
19	25	1			
20	27,7	0			
21	30,4	3	=D19+D20+D21	=(НОРМРАСП(C22;F3;F7;ИСТИНА)-НОРМРАСП(C19;F3;F7;ИСТИНА))*$	=(E21-F21)^2/F21
22	33,1	8	8	=(НОРМРАСП(C23;F3;F7;ИСТИНА)-НОРМРАСП(C22;F3;F7;ИСТИНА))*$	=(E22-F22)^2/F22
23	35,8	14	14	=(НОРМРАСП(C24;F3;F7;ИСТИНА)-НОРМРАСП(C23;F3;F7;ИСТИНА))*$	=(E23-F23)^2/F23
24	38,5	12	12	=(НОРМРАСП(C25;F3;F7;ИСТИНА)-НОРМРАСП(C24;F3;F7;ИСТИНА))*$	=(E24-F24)^2/F24
25	41,2	19	19	=(НОРМРАСП(C26;F3;F7;ИСТИНА)-НОРМРАСП(C25;F3;F7;ИСТИНА))*$	=(E25-F25)^2/F25
26	43,9	15	15	=(НОРМРАСП(C27;F3;F7;ИСТИНА)-НОРМРАСП(C26;F3;F7;ИСТИНА))*$	=(E26-F26)^2/F26
27	46,6	15	15	=(НОРМРАСП(C28;F3;F7;ИСТИНА)-НОРМРАСП(C27;F3;F7;ИСТИНА))*$	=(E27-F27)^2/F27
28	49,3	10	10	=(НОРМРАСП(C29;F3;F7;ИСТИНА)-НОРМРАСП(C28;F3;F7;ИСТИНА))*$	=(E28-F28)^2/F28
29	52	3	3	=(НОРМРАСП(C30;F3;F7;ИСТИНА)-НОРМРАСП(C29;F3;F7;ИСТИНА))*$	=(E29-F29)^2/F29
30	=C29+(C29-C28)			Статистика хи-квадрат	=СУММ(G21:G29)
31				Ошибка	0,05
32				Число степеней свободы	=СЧЕТ(E21:E29)-3
33				Табл. Значение	=ХИ2ОБР(G31:G32)

34				Проверка условия	=TCKB(G33>G30; ”да”;”нет”)

Пример 3. Проведите анализ данных в рамках задачи проверки гипотезы о равенстве среднего некоторому определенному значению (для данных, взятых из нормально распределенной генеральной совокупности) с использованием средств **Вставка функций** MS Excel.

1. Перейдите на рабочий **Лист 3** книги и переименуйте его в **Тест среднее**.
2. Сформируйте массив данных, используя функцию генерации случайных чисел средства **Анализ данных** для нормального распределения со средним 0,15 и стандартным отклонением 0,1 при числе выборок 20.
3. Рассчитайте среднее и дисперсию выборки, используя стандартные функции MS Excel.
4. Сформируйте таблицу для проверки критерия, исходя из того, что критериальное значение вычисляется по формуле $t = \frac{(m - A)\sqrt{n}}{S^2}$, где А-предполагаемая величина среднего (например, 0,09), а критическое значение вычисляется для распределения Стьюдента со значимостью α и числом степеней свободы n-1. гипотеза о равенстве среднего отвергается, если по абсолютной величине критериальное значение больше верхней α/2-% точки распределения Стьюдента: $|t_{критериальное}| > T\{\ \alpha/2,\ n\text{-}1\}$. Для расчета распределения Стъюдента используйте функцию **СТЪЮДРАСПОБР**().

Применение пакета анализа MS Excel для решения задач проверки гипотез

Пример 4. проведите анализ данных в рамках задачи проверки гипотезы о принадлежности двух дисперсий одной генеральной совокупности (следовательно, их равенстве) по критерию Фишера с использованием средств **Анализа данных** MS Excel.

1. Перейдите на рабочий **Лист 4** книги и переименуйте его в **Тест. рав. дисп**.
2. Сформируйте массив исходных данных двух независимых выборок значений измерения веса опытных животных по образцу 5.

Образец 5

№ группы	Результаты измерений									
1	55	73	50	71	63	59	66	74	58	69
2	43	39	50	47	47	38	51	48	37	42

3. Запустите процедуру проверки гипотезы: **Сервис/Анализ данных/Двухвыборочный тест для дисперсии**/в одноименном окне укажите диапазоны ячеек для 1и 2 выборок в полях **Интервал переменной** …., введите уровень значимости **0,05** в поле **Альфа**, укажите верхнюю левую ячейку размещения результатов в поле **Выходной интервал** и **ОК**.
4. Проанализируйте полученные результаты. По условиям критерия нулевая гипотеза отвергается, если значение F статистики Фишера больше верхнего критического или меньше нижнего.
5. Постройте для обеих выборок гистограммы с полиномиальным трендом.

Пример 5. Проведите анализ данных в рамках задачи проверки гипотезы о равенстве средних при неравных дисперсиях и объемах выборок по критерию Стъюдента с использованием средств **Анализа данных** MS Excel.

1. Перейдите на рабочий **Лист 5** книги и переименуйте его в **Тест. рав. сред1**.
2. Сформируйте массив данных, используя функцию генерации случайных чисел средства **Анализ данных** для нормального распределения со средним 1 и стандартным отклонением 1 и 2 для двух выборок объемом 15 и 20 чисел соответственно, разместив их в столбцах **В** и **С**.
3. Запустите процедуру проверки гипотезы **Сервис/Анализ данных/Двухвыборочный t-тест с различными дисперсиями/** в одноименном окне

укажите диапазоны ячеек для 1и 2 выборок в полях **Интервал переменной** …., введите уровень значимости **0,05** в поле **Альфа**, укажите верхнюю левую ячейку размещения результатов в поле **Выходной интервал** и **ОК**.

4. Проанализируйте полученные результаты. По условиям критерия нулевая гипотеза отвергается, если значение t-статистики Стъюдента по абсолютной величине больше верхней точки распределения или критического значения.
5. Постройте для обеих выборок гистограммы с трендом скользящего среднего.

<u>Пример 6</u>. Проведите анализ данных в рамках задачи проверки гипотезы о равенстве средних при равных дисперсиях по критерию Стъюдента с использованием средств **Анализа данных** MS Excel.

1. Перейдите на рабочий **Лист 6** книги и переименуйте его в **Тест. рав. сред2**.
2. Сформируйте массив данных, используя функцию генерации случайных чисел средства **Анализ данных** для нормального распределения со средним 2 и стандартным отклонением 2 для двух выборок объемом 40 чисел каждая, разместив их в столбцах **B** и **C**.
3. Запустите процедуру проверки гипотезы **Сервис/Анализ данных/Двухвыборочный t-тест с одинаковыми дисперсиями/** в одноименном окне укажите диапазоны ячеек для 1и 2 выборок в полях **Интервал переменной** …., введите уровень значимости **0,05** в поле **Альфа**, укажите верхнюю левую ячейку размещения результатов в поле **Выходной интервал** и **ОК**. При вводе значения в поле **Гипотетическая средняя разность** проверяется гипотеза о разности значений средних двух выборок.
4. Проанализируйте полученные результаты и примите решение.
5. Постройте для обеих выборок гистограммы с полиномиальным трендом.

Приложение 1

Стандартные значения критерия Стъюдента

Число степеней свободы $u = n_1 + n_2 - 2$	Критерий Стъюдента t_{st} при вероятности безошибочного заключения *p*			
	0.1	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.952
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.684	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.732	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.723	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.714	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
∞	1.645	1.960	2.326	2.576

Приложение 2

Значения критерия Фишера F

U_2	Степень свободы для большей дисперсии U_1											
	3	4	5	6	8	10	12	16	24	30	50	∞
5%-ный уровень значимости $F_{0,05}$												
3	9,3	9,1	9,0	8,9	8,8	8,8	8,7	8,7	8,6	8,6	8,6	8,5
4	6,6	6,4	6,3	6,2	6,0	5,9	5,8	5,8	5,8	5,7	5,7	5,6
5	5,4	5,2	5,1	5,0	4,8	4,7	4,7	4,6	4,5	4,5	4,4	4,4
6	4,8	4,5	4,4	4,3	4,2	4,1	4,0	3,9	3,8	3,8	3,8	3,7
7	4,4	4,1	4,0	3,9	3,7	3,6	3,6	3,5	3,4	3,4	3,4	3,2
8	4,1	3,8	3,7	3,6	3,4	3,4	3,4	3,2	3,1	3,1	3,0	2,9
9	3,9	3,6	3,5	3,4	3,2	3,2	3,1	3,0	2,9	2,9	2,8	2,7
10	3,7	3,5	3,3	3,2	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5
12	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3
16	3,2	3,1	2,9	2,8	2,7	2,5	2,4	2,3	2,2	2,2	2,1	2,0
18	3,2	2,9	2,8	2,7	2,5	2,4	2,3	2,3	2,2	2,1	2,0	1,9
24	3,0	2,8	2,6	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,7
40	2,8	2,6	2,5	2,3	2,2	2,1	2,0	1,9	1,8	1,7	1,7	1,5
120	2,7	2,4	2,3	2,2	2,0	1,9	1,8	1,7	1,6	1,6	1,5	1,3
∞	2,6	2,4	2,2	2,1	1,9	1,8	1,8	1,6	1,5	1,5	1,4	1,0

1%-ный уровень значимости $F_{0,01}$												
3	29,5	28,7	28,2	27,9	27,5	27,0	26,8	26,6	26,5	26,5	26,4	26,1
4	16,7	16,0	15,5	15,2	14,8	14,5	14,4	14,2	13,9	13,8	13,7	13,5
5	12,1	11,4	11,0	10,7	10,3	10,1	9,9	9,7	9,5	9,4	9,2	9,0
6	9,8	9,2	8,8	8,5	8,1	7,9	7,7	7,5	7,3	7,2	7,1	6,9
7	8,4	7,8	7,5	7,2	6,8	6,6	6,5	6,3	6,1	6,0	5,8	5,6
8	7,6	7,0	6,6	6,4	6,0	5,8	5,7	5,5	5,3	5,2	5,1	4,9
9	7,0	6,4	6,1	5,8	5,5	5,3	5,1	4,9	4,7	4,6	4,5	4,3
10	6,6	6,0	5,6	5,4	5,1	4,9	4,7	4,5	4,3	4,3	4,1	3,9
12	6,0	5,4	5,1	4,8	4,5	4,3	4,2	4,0	3,8	3,7	3,6	3,4
16	5,3	4,8	4,4	4,2	3,9	3,7	3,6	3,4	3,2	3,1	3,0	2,8
20	4,9	4,4	4,1	3,9	3,6	3,4	3,2	3,0	2,9	2,8	2,6	2,4
30	4,5	4,0	3,7	3,5	3,2	3,0	2,8	2,7	2,5	2,4	2,2	2,0
60	4,1	3,6	3,3	3,1	2,8	2,6	2,5	2,3	2,1	2,0	1,9	1,6
120	3,9	3,5	3,2	3,0	2,7	2,4	2,3	2,2	1,9	1,9	1,7	1,4
∞	3,8	3,3	3,0	2,8	2,5	2,3	2,2	2,0	1,8	1,7	1,5	1,0

Приложение 3

Критические значения χ^2 для трех степеней доверительной вероятности.

Число степеней свободы, U	Уровень значимости			Число степеней свободы, U	Уровень значимости		
	0.95	0.99	0.999		0.95	0.99	0.999
1	3.8	6.6	10.8	26	38.9	45.6	54.1
2	6.0	9.2	13.8	27	40.1	47.0	55.5
3	7.8	11.3	16.3	28	41.3	48.3	56.9
4	9.5	13.3	18.5	29	42.6	49.6	58.3
5	11.1	15.1	20.5	30	43.8	50.9	59.7
6	12.6	16.8	22.5	32	46.2	53.5	62.4
7	14.1	18.5	24.3	34	48.6	56.0	65.2
8	15.5	20.1	26.1	36	51.0	58.6	67.9
9	16.9	21.7	27.9	38	53.4	61.1	70.7
10	18.3	23.2	29.6	40	55.8	63.7	73.4
11	19.7	24.7	31.3	42	58.1	66.2	76.1
12	21.0	26.2	32.9	44	60.5	68.7	78.7
13	22.4	27.7	34.5	46	62.8	71.2	81.4
14	23.7	29.1	36.1	48	65.2	73.7	84.0
15	25.0	30.6	37.7	50	67.5	76.2	86.7
16	26.3	32.0	39.3	55	73.3	82.3	93.2
17	27.6	33.4	40.8	60	79.1	88.4	99.6
18	28.9	34.8	42.3	65	89.8	94.4	106.0
19	30.1	36.2	43.8	70	90.5	100.4	112.3
20	31.4	37.6	45.3	75	96.2	106.4	118.5
21	32.7	38.9	46.8	80	101.9	112.3	124.8
22	33.9	40.3	48.3	85	107.5	118.2	131.0
23	35.2	41.6	49.7	90	113.1	124.1	137.1
24	36.4	43.0	51.2	95	118.7	130.0	143.3
25	37.7	44.3	52.5	100	124.3	135.8	149.4

Литература

1. Белько И.В. Теория вероятностей и математическая статистика. Примеры и задачи.; Учеб. пособие / И.В.Белько, Г.П.Свирид. Под ред. К.К. Кузьмича.- Мн.; Новое знание, 2004.- 254 с.
2. Доспехов Б.А. Методика полевого опыта. М.: Колос, 1979 .- 416 с.
3. Калинина В.И. Математическая статистика; Учеб. для студ. сред. спец. учеб. заведений / В.Н. Калинина, В.Ф. Панкин. – М.; Дрофа, 2002.- 336 с.
4. Лакин Г.Ф. Биометрия. М.: Высшая школа, 1980. – 293.
5. Математические методы в биологии \ отв. ред. Плохинский Н.А.. М.: Изд-во Моск. Ун-та, 1972. – 135с.
6. Рокицкий П.Ф. Биологическая статистика. Минск. «Вышэйшая школа», 1973.- 320 с.