Учреждение образования Гомельский государственный университет имени Франциска Скорины»

Факультет биологический
Кафедра зоологии, физиологии и генетики

СОГЛАСОВАНО

Заведующий кафедрой

_____Г. Г. Гончаренко
09 01 2019

СОГЛАСОВАНО

Декан факультета

_____В.С. Аверин
09 01 2019

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

Биометрия
для специальности I – 31 01 01 02 – «Биология»
(научно – педагогическая деятельность)
специализации
1-31 01 01-02 01 Зоология

Рассмотрено и утверждено на заседании
кафедры зоологии, физиологии и генетики
09 01 2019 г. протокол № 8

Составители:
старший преподаватель И.В.Кураченко
старший преподаватель С.А.Зятьков
член-корр.НАН Б, профессор, д.б.н. Г.Г.Гончаренко

Рассмотрено и утверждено
на заседании научно-методического совета
УО «Гомельский государственный университет им. Ф. Скорины»
22. 02. 2019 г., протокол № 5

Содержание
учебно-методического комплекса по дисциплине «Биометрия»
для специальности
1-31 01 01-02 – «Биология. Научно-педагогическая деятельность»

Пояснительная записка
учебно-методического комплекса по дисциплине
«Биометрия»

Учебно-методический комплекс предназначен для студентов 3 курса (4 курса ЗФ) специальности 1-31 01 01-02 «Биология (научно-педагогическая деятельность)».

Современная биология давно перестала быть исключительно описательной наукой. Сегодня ее существование и развитие невозможно без использования методов и подходов такой области математики как статистика. В связи этим курс лекций «Биометрия» является обязательным при подготовке специалистов биологического профиля.

Целью дисциплины обязательного компонента «Биометрия» является овладение студентами основами современных методов статистического анализа биологических данных.

В задачи курса входят: освоение студентами методов, позволяющих выявлять количественные закономерности в биологических явлениях; формирование у студентов навыков и умений компьютерной обработки экспериментальных данных, а также ознакомление с правилами корректного представления результатов исследований коллегам; математическое оформление статей биологического содержания и формирование способности к критическому анализу представляемых в публикациях данных; ознакомление с принципами построения математических моделей биологических явлений и процессов.

В курсе подробно рассматриваются традиционные методы анализа данных. Наряду с этим большое внимание уделяется непараметрическим методам, использование которых в практике биологических исследований постоянно возрастает. На примере кластерного и дискриминантного анализов, а также метода главных компонент слушатели знакомятся с элементами многомерной статистики. Большое количество часов в рамках курса отводится для лабораторных работ, в ходе которых студенты приобретают навыки и умения статистической обработки данных при помощи персонального компьютера.

Изложение курса студентам-биологам осуществляется в границах одного семестра на третьем курсе, что требует емкого и в тоже время богатого информацией и специфической терминологией материала.

Содержание разделов ЭУМК соответствует образовательным стандартам высшего образования данной специальности. Главная цель ЭУМК – оказание методической помощи студентам в систематизации учебного материала в процессе подготовки к итоговой аттестации по курсу «Биометрия».

Структура ЭУМК включает:

1. Учебно-методическое обеспечение дисциплины

1.1. Теоретический раздел (учебное издание для теоретического изучения дисциплины в объеме, установленном типовым учебным планом по специальности).

1.2. Практический раздел (материалы для проведения лабораторных занятий по дисциплине в соответствии с учебным планом).

2. Контроль контролируемой самостоятельной работы студентов (материалы текущей и итоговой аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля и др.).

3. Вспомогательный раздел.

3.1. Учебно-программные материалы (учебная программа для студентов дневной и заочной форм получения образования).

3.2. Информационно-аналитические материалы (список рекомендуемой литературы, перечень электронных образовательных ресурсов и их адреса и др.).

Работа с ЭУМК должна включать на первом этапе ознакомление с тематическим планом дисциплины, с тематикой лекций и лабораторных занятий, перечнях рассматриваемых вопросов и рекомендуемой для их изучения литературы. Для подготовки к лабораторным занятиям и промежуточным зачетам необходимо, в первую очередь, использовать материалы, представленные в разделе учебно-методическое обеспечение дисциплины, а также материалы для текущего контроля самостоятельной работы. В ходе подготовки к итоговой аттестации рекомендуется ознакомиться с требованиями к компетенциям по дисциплине, изложенными в учебной программе, структурой рейтинговой системы, а также перечнем вопросов к зачету. Для написания рефератов могут быть использованы информационно-аналитические материалы, указанные в соответствующем разделе ЭУМК.

В результате изучения дисциплины:

выпускник должен знать:

- анализ математической статистики;
- математические методы обработки результатов;
- принципы построения и использования математических моделей биологических процессов;

выпускник должен владеть:

- умениями и навыками расчета ошибки репрезентативности выборочных показателей и доверительных границ и доверительных интервалов выборочных параметров наблюдения, описания, идентификации, классификации живых организмов;
- математическими методами обработки результатов, понимать принципы построения и использования математических моделей биологических процессов;

- методами наблюдения, описания, классификации, экспериментального анализа.
- методами экспериментального анализа объектов профессиональной деятельности;

выпускник должен уметь использовать:

- знания о развитии математической статистики;
- выбор форм и методов обучения;

выпускник должен иметь опыт:

- планирования и анализа результатов количественных биологических экспериментов и наблюдений методами математической статистики;
- анализа методов обработки, представления и хранения информации.

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»

Кафедра зоологии, физиологии и генетики

Кураченко И.В., Зятьков С.А., Гончаренко Г.Г.

БИОМЕТРИЯ

Гомель 2019

СОДЕРЖАНИЕ

Лекция 1. ВВЕДЕНИЕ В КУРС. ДАННЫЕ В БИОЛОГИИ

1.1 Содержание науки. Биометрия − раздел вариационной статистики, с помощью методов которого производят обработку экспериментальных данных и наблюдений, а также планирование количественных экспериментов в биологических исследованиях; а также научная отрасль, связанная с разработкой и использованием статистических методов в научных исследованиях в медицине, здравоохранении, и эпидемиологии.

1.2 История. Биометрия сложилась в XIX веке, главным образом, благодаря трудам Ф. Гальтона и К. Пирсона. В 1920-30-х годах крупный вклад в развитие биометрии внес Р. Фишер.У истоков биометрии стоял Фрэнсис Гальтон (1822−1911). Первоначально Гальтон готовился стать врачом. Однако, обучаясь в Кембриджском университете, он увлекся естествознанием, метеорологией, антропологией, наследственностью и теорией эволюции. В его книге, посвященной природной наследственности, изданной в 1889 году им впервые было введено в употребление слово «biometry» и в это же время он разработал основы корреляционного анализа. Гальтон заложил основы новой науки и дал ей имя. Однако превратил её в стройную научную дисциплину математик Карл Пирсон (1857−1936). В 1884 году Пирсон получает кафедру прикладной математики в Лондонском университете, а в 1889 году знакомится с Гальтоном и его работами. Большую роль в жизни Пирсона сыграл зоолог Ф.Велдон. Помогая ему в анализе реальных зоологических данных, Пирсон ввел в 1893 г. понятие среднего квадратического отклонения и коэффициента вариации. Пытаясь математически оформить теорию наследственности Гальтона, Пирсон в 1898 г. разрабатывает основы множественной регрессии. В 1903 г. Пирсон разработал основы теории сопряженности признаков, а в 1905 г. опубликовал основы нелинейной корреляции и регрессии.

Следующий этап развития биометрии связан с именем великого английского статистика Рональда Фишера (1890−1962). Во время обучения в Кембриджском университете Фишер знакомится с трудами Менделя и Пирсона. В 1913−1915 годах Фишер работает статистиком на одном из предприятий, а в 1915−1919 годах преподает физику и математику в средней школе. С 1919 года Фишер начинает работу статистиком на опытной сельскохозяйственной станции в Ротамстеде, где он проработал до 1933 года. Затем с 1933 года по 1943 год Фишер работает профессором в Лондонском университете, а с 1943 года по 1957 год заведует кафедрой генетики в Кембридже. За эти годы им были разработаны теория выборочных распределений, методы дисперсионного и дискриминантного анализа, теории планирования экспериментов, метод максимального правдоподобия и многое другое, что составляет основу современной прикладной статистики и математической генетики.

1.3 Развитие представлений о статистике. Статистика − отрасль знаний, в которой излагаются общие вопросы сбора, измерения и анализа массовых статистических (количественных или качественных) данных.

Слово «статистика» происходит от латинского *status* − состояние дел[1]. В науку термин «статистика» ввел немецкий ученый Готфрид Ахенваль в 1746

году, предложив заменить название курса «Государствоведение», преподававшегося в университетах Германии, на «Статистику», положив тем самым начало развитию статистики как науки и учебной дисциплины. Несмотря на это, статистический учет вёлся намного раньше: проводились переписи населения в Древнем Китае, осуществлялось сравнение военного потенциала государств, велся учет имущества граждан в Древнем Риме и т. п.

Статистика разрабатывает специальную методологию исследования и обработки материалов: массовые статистические наблюдения, метод группировок, средних величин, индексов, балансовый метод, метод графических изображений и другие методы анализа статистических данных. Начало статистической практики относится примерно к времени возникновения государства. Первой опубликованной статистической информацией можно считать глиняные таблички Шумерского царства (III − II тысячелетия до н. э.).

Вначале под статистикой понимали описание экономического и политического состояния государства или его части. Например, к 1792 г. относится определение: «статистика описывает состояние государства в настоящее время или в некоторый известный момент в прошлом». И в настоящее время деятельность государственных статистических служб вполне укладывается в это определение.

Однако постепенно термин «статистика» стал использоваться более широко. По Наполеону Бонапарту, «статистика − это бюджет вещей». Тем самым статистические методы были признаны полезными не только для административного управления, но и для применения на уровне отдельного предприятия. Согласно формулировке 1833 г., «цель статистики заключается в представлении фактов в наиболее сжатой форме». Во 2-й половине XIX − начале XX веков сформировалась научная дисциплина −математическая статистика, являющаяся частью математики.

В XX веке статистику часто рассматривают, прежде всего, как самостоятельную научную дисциплину. Статистика есть совокупность методов и принципов, согласно которым проводится сбор, анализ, сравнение, представление и интерпретация числовых данных. В 1954 г. академик АН Борис Владимирович Гнеденко дал следующее определение: «Статистика состоит из трёх разделов:

1 сбор статистических сведений, то есть сведений, характеризующих отдельные единицы каких-либо массовых совокупностей;

2 статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе данных массового наблюдения;

3 разработка приёмов статистического наблюдения и анализа статистических данных. Последний раздел, собственно, и составляет содержание математической статистики».

Термин «статистика» употребляют ещё в двух смыслах. Во-первых, в обиходе под «статистикой» часто понимают набор количественных данных о каком-либо явлении или процессе. Во-вторых, статистикой называют функцию от результатов наблюдений, используемую для оценки характеристик и параметров распределений и проверки гипотез.

1.4 Краткая история статистических методов. Типовые примеры раннего этапа применения статистических методов описаны в Библии, в Ветхом Завете. Там, в частности, приводится число воинов в различных племенах. С математической точки зрения дело сводилось к подсчёту числа попаданий значений наблюдаемых признаков в определённые градации.

Сразу после возникновения теории вероятностей (Паскаль, Ферма, XVII век) вероятностные модели стали использоваться при обработке статистических данных. Например, изучалась частота рождения мальчиков и девочек, было установлено отличие вероятности рождения мальчика от 0.5, анализировались причины того, что в парижских приютах эта вероятность не та, что в самом Париже, и т. д.

В 1794 г. (по другим данным − в 1795 г.) немецкий математик Карл Гаусс формализовал один из методов современной математической статистики − метод наименьших квадратов. В XIX веке заметный вклад в развитие практической статистики внёс бельгиец Кетле, на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей, таких, как доля самоубийств среди всех смертей.

Первая треть XX века прошла под знаком параметрической статистики. Изучались методы, основанные на анализе данных из параметрических семейств распределений, описываемых кривыми семейства Пирсона. Наиболее популярным было нормальное распределение. Для проверки гипотез использовались критерии Пирсона, Стьюдента, Фишера. Были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

Разработанную в первой трети XX века теорию анализа данных называют параметрической статистикой, поскольку её основной объект изучения − это выборки из распределений, описываемых одним или небольшим числом параметров. Наиболее общим является семейство кривых Пирсона, задаваемых четырьмя параметрами. Как правило, нельзя указать каких-либо веских причин, по которым распределение результатов конкретных наблюдений должно входить в то или иное параметрическое семейство. Исключения хорошо известны: если вероятностная модель предусматривает суммирование независимых случайных величин, то сумму естественно описывать нормальным распределением; если же в модели рассматривается произведение таких величин, то итог, видимо, приближается логарифмически нормальным распределением и так далее.

Статистические методы −методы анализа статистических данных. Выделяют методы прикладной статистики, которые могут применяться во всех областях научных исследований и любых отраслях народного хозяйства, и другие статистические методы, применимость которых ограничена той или иной сферой. Имеются в виду такие методы, как статистический приемочный контроль, статистическое регулирование технологических процессов, надежность и испытания, планирование экспериментов.

Статистические методы анализа данных применяются практически во всех областях деятельности человека. Их используют всегда, когда необходимо

получить и обосновать какие-либо суждения о группе (объектов или субъектов) с некоторой внутренней неоднородностью.

Целесообразно выделить три вида научной и прикладной деятельности в области статистических методов анализа данных (по степени специфичности методов, сопряженной с погруженностью в конкретные проблемы):

а) разработка и исследование методов общего назначения, без учета специфики области применения;

б) разработка и исследование статистических моделей реальных явлений и процессов в соответствии с потребностями той или иной области деятельности;

в) применение статистических методов и моделей для статистического анализа конкретных данных.

Прикладная статистика − это наука о том, как обрабатывать данные произвольной природы. Математической основой прикладной статистики и статистических методов анализа является теория вероятностей и математическая статистика.

Описание вида данных и механизма их порождения − начало любого статистического исследования. Для описания данных применяют как детерминированные, так и вероятностные методы. С помощью детерминированных методов можно проанализировать только те данные, которые имеются в распоряжении исследователя. Например, с их помощью получены таблицы, рассчитанные органами официальной государственной статистики на основе представленных предприятиями и организациями статистических отчетов. Перенести полученные результаты на более широкую совокупность, использовать их для предсказания и управления можно лишь на основе вероятностно-статистического моделирования. Поэтому в математическую статистику часто включают лишь методы, опирающиеся на теорию вероятностей.

В простейшей ситуации статистические данные − это значения некоторого признака, свойственного изучаемым объектам. Значения могут быть количественными или представлять собой указание на категорию, к которой можно отнести объект. Во втором случае говорят о качественном признаке.

При измерении по нескольким количественным или качественным признакам в качестве статистических данных об объекте получаем вектор. Его можно рассматривать как новый вид данных. В таком случае выборка состоит из набора векторов. Есть часть координат − числа, а часть − качественные (категоризованные) данные, то говорим о векторе разнотипных данных.

Одним элементом выборки, то есть одним измерением, может быть и функция в целом. Например, описывающая динамику показателя, то есть его изменение во времени, − электрокардиограмма больного или амплитуда биений вала двигателя. Или временной ряд, описывающий динамику показателей определенной фирмы. Тогда выборка состоит из набора функций.

Элементами выборки могут быть и иные математические объекты. Например, бинарные отношения. Так, при опросах экспертов часто используют упорядочения (ранжировки) объектов экспертизы − образцов продукции, инвестиционных проектов, вариантов управленческих решений. В зависимости

от регламента экспертного исследования элементами выборки могут быть различные виды бинарных отношений (упорядочения, разбиения, толерантности), множества, нечёткие множества и т. д.

Итак, математическая природа элементов выборки в различных задачах прикладной статистики может быть самой разной. Однако можно выделить два класса статистических данных − числовые и нечисловые. Соответственно прикладная статистика разбивается на две части − числовую статистику и нечисловую статистику.

Числовые статистические данные − это числа, вектора, функции. Их можно складывать, умножать на коэффициенты. Поэтому в числовой статистике большое значение имеют разнообразные суммы. Математический аппарат анализа сумм случайных элементов выборки − это (классические) законы больших чисел и центральные предельные теоремы.

Нечисловые статистические данные − это категоризованные данные, вектора разнотипных признаков, бинарные отношения, множества, нечеткие множества и др. Их нельзя складывать и умножать на коэффициенты. Поэтому не имеет смысла говорить о суммах нечисловых статистических данных. Они являются элементами нечисловых математических пространств (множеств). Математический аппарат анализа нечисловых статистических данных основан на использовании расстояний между элементами (а также мер близости, показателей различия) в таких пространствах. С помощью расстояний определяются эмпирические и теоретические средние, доказываются законы больших чисел, строятся непараметрические оценки плотности распределения вероятностей, решаются задачи диагностики и кластерного анализа, и т. д.

В прикладных исследованиях используют статистические данные различных видов. Это связано, в частности, со способами их получения. Например, если испытания некоторых технических устройств продолжаются до определенного момента времени, то получаем т. н. цензурированные данные, состоящие из набора чисел − продолжительности работы ряда устройств до отказа, и информации о том, что остальные устройства продолжали работать в момент окончания испытания. Цензурированные данные часто используются при оценке и контроле надежности технических устройств.

Теория статистических методов нацелена на решение реальных задач. Поэтому в ней постоянно возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится математическими средствами, то есть путем доказательства теорем. Большую роль играет методологическая составляющая − как именно ставить задачи, какие предположения принять с целью дальнейшего математического изучения. Велика роль современных информационных технологий, в частности, компьютерного эксперимента.

Развитие вычислительной техники во второй половине XX века оказало значительное влияние на статистику. Ранее статистические модели были представлены преимущественно линейными моделями. Увеличение быстродействия ЭВМ и разработка соответствующих численных алгоритмов послужило причиной повышенного интереса к нелинейным моделям таким, как

искусственные нейронные сети, и привело к разработке сложных статистических моделей, например обобщенная линейная модель и иерархическая модель.

1. 5 Статистическое наблюдение — это массовое (оно охватывает большое число случаев проявления исследуемого явления для получения правдивых статистических данных) планомерное (проводится по разработанному плану, включающему вопросы методологии, организации сбора и контроля достоверности информации), систематическое (проводится систематически, либо непрерывно, либо регулярно), научно организованное (для повышения достоверности данных, которая зависит от программы наблюдения, содержания анкет, качества подготовки инструкций) наблюдение за явлениями и процессами социально-экономической жизни, которое заключается в сборе и регистрации отдельных признаков у каждой единицы совокупности.

Этапы статистического наблюдения

1. **Подготовка к статистическому наблюдению** (решение научно-методических и организационно-технических вопросов).

- определение цели и объекта наблюдения;
- определение состава признаков подлежащих изучению;
- разработка документов для сбора данных;

2. **Сбор информации**

- непосредственное заполнение статистических формуляров (бланки, анкеты);
- применение стандартных методов сбора (пробная площадка, ловушки Геро и пр.)

Статистическая информация — это первичные данные о предмете изучения, формирующиеся в процессе статистического наблюдения, которые затем подвергаются систематизации, сводке, анализу и обобщению.

3. **Первичная обработка данных**

4. **Статистический анализ обработанной информации.**

5. **Разработка предложений и рекомендаций по совершенствованию статистического наблюдения**

- заключается в анализе причин, которые привели к неверному заполнению статистических формуляров и разработке соответствующих предложений по совершенствованию наблюдения.

В результате статистического наблюдения должна быть получена объективная, сопоставимая, полная информация, позволяющая на последующих этапах исследования обеспечить научно-обоснованные выводы о характере и закономерностях развития изучаемого явления.

Виды статистического наблюдения

Статистические наблюдения подразделяются на виды по следующим признакам:

- по времени регистрации данных;
- по полноте охвата единиц совокупности;

Виды статистического наблюдения по времени регистрации:

Текущее (непрерывное) наблюдение - проводится для изучения текущих явлений и процессов. Регистрация фактов осуществляется по мере их свершения.

Прерывное наблюдение — проводится по мере необходимости, при этом допускаются временные разрывы в регистрации данных:

- **Периодическое** наблюдение — проводится через сравнительно равные интервалы времени.
- **Единовременное** наблюдение — осуществляется без соблюдения строгой периодичности его проведения.

По полноте охвата единиц совокупности различают следующие виды статистического наблюдения:

Сплошное наблюдение — представляет собой сбор и получение информации обо всех единицах изучаемой совокупности.

Несплошное наблюдение — основано на принципе случайного отбора единиц изучаемой совокупности, при этом в выборочной совокупности должны быть представлены все типы единиц, имеющихся в совокупности.

Несплошное наблюдение подразделяется на:

- **Выборочное наблюдение** - основано на случайном отборе единиц, которые подвергаются наблюдению.
- **Монографическое наблюдение** — заключается в обследовании отдельных единиц совокупности, характеризующихся редкими качественными свойствами.
- **Метод основного массива** — состоит в изучении самых существенных, наиболее крупных единиц совокупности, имеющих по основному признаку наибольший удельный вес в изучаемой совокупности.
- **Метод моментных наблюдений** — заключается в проведении наблюдений через случайные или постоянные интервалы времени с отметками о состоянии исследуемого объекта в тот или иной момент времени.

Способы статистического наблюдения

Непосредственное статистическое наблюдение — наблюдение, при котором сами регистраторы путем непосредственного замера, взвешивания, подсчета устанавливают факт подлежащий регистрации.

Документальное наблюдение — основано на использовании различного рода документов учетного характера.

Опрос - заключается в получении необходимой информации непосредственно от респондента.

Существуют следующие виды опроса:

Экспедиционный — регистраторы получают необходимую информацию от опрашиваемых лиц и сами фиксируют ее в формулярах.

Способ саморегистрации — формуляры заполняются самими респондентами, регистраторы только раздают бланки и объясняют правила их заполнения.

Корреспондентский — сведения в соответствующие органы сообщает штат добровольных корреспондентов.

Анкетный — сбор информации осуществляется в виде анкет, представляющих собой специальные вопросники, удобен в случаях, когда не требуется высокая точность результатов.

Явочный — заключается в предоставлении сведений в соответствующие органы в явочном порядке.

В зависимости от причин возникновения различают **ошибки регистрации и ошибки репрезентативности**. Ошибки регистрации характерны как для сплошного, так и для несплошного наблюдения, а ошибки репрезентативности — только для несплошного наблюдения. Ошибки регистрации, как и ошибки репрезентативности, могут быть **случайными и систематическими.**

Ошибки регистрации — представляют собой отклонения между значением показателя, полученного в ходе статистического наблюдения, и его фактическим значением. Ошибки регистрации бывают случайными (результат действий случайных факторов — перепутаны строки например) и систематическими (проявляются постоянно).

Ошибки репрезентативности — возникают, когда отобранная совокупность недостаточно точно воспроизводит исходную совокупность. Характерны для несплошного наблюдения и заключаются в отклонении величины показателя исследуемой части совокупности от его величины в генеральной совокупности.

Случайные ошибки — являются результатом действия случайных факторов.

Систематические ошибки — всегда имеют одинаковую направленность к увеличению или уменьшению показателя по каждой единице наблюдения, вследствие чего значение показателя по совокупности в целом будет включать накопленную ошибку.

Способы контроля:

- Счетный (арифметический) — проверка правильности арифметического расчета.
- Логический — основан на смысловой взаимосвязи между признаками.

Статистическая совокупность - множество единиц, обладающих массовостью, типичностью, качественной однородностью и наличием вариации.

Статистическая совокупность состоит из материально существующих объектов, является объектом статистического исследования.

Единица совокупности — каждая конкретная единица статистической совокупности.

Одна и та же статистическая совокупность может быть однородна по одному признаку и неоднородна по другому.

Качественная однородность — сходство всех единиц совокупности по какому-либо признаку и несходство по всем остальным.

В статистической совокупности отличия одной единицы совокупности от другой чаще имеют количественную природу. Количественные изменения значений признака разных единиц совокупности называются вариацией.

Вариация признака — количественное изменение признака (для количественного признака) при переходе от одной единицы совокупности к другой.

Признак - это свойство, характерная черта или иная особенность единиц, объектов и явлений, которая может быть наблюдаема или измерена. Признаки делятся на количественные и качественные. Многообразие и изменчивость величины признака у отдельных единиц совокупности называется **вариацией.**

Атрибутивные (качественные) признаки не поддаются числовому выражению (состав населения по полу). Количественные признаки имеют числовое выражение (состав населения по возрасту).

Показатель — это обобщающая количественно качественная характеристика какого-либо свойства единиц или совокупности в целом в конкретных условиях времени и места.

Система показателей — это совокупность показателей всесторонне отражающих изучаемое явление.

Например, изучается привес коров:

- Признак — вес
- Статистическая совокупность — коровы фермы
- Единица совокупности — каждая корова
- Качественная однородность — одного возраста
- Вариация признака — ряд цифр

1.6 Генеральная совокупность и выборка из нее

Основу статистического исследования составляет множество данных, полученных в результате измерения одного или нескольких признаков. Реально наблюдаемая совокупность объектов, статистически представленная рядом наблюдений $x_1, x_2, ..., x_n$ случайной величины X, является **выборкой**, а гипотетически существующая (домысливаемая) — **генеральной совокупностью**. Генеральная совокупность может быть конечной (число наблюдений **N = const**) или бесконечной (**N** = ∞), а выборка из генеральной совокупности — это всегда результат ограниченного ряда n наблюдений. Число наблюдений n, образующих выборку, называется **объемом выборки**. Если объем выборки n достаточно велик (**n** → ∞) выборка считается **большой**, в противном случае она называется выборкой **ограниченного объема**. Выборка считается **малой**, если при измерении одномерной случайной величины X объем выборки не превышает 30 (**n <= 30**), а при измерении одновременно нескольких (**k**) признаков в многомерном пространстве отношение **n** к **k** не превышает **10 (n/k < 10)**. Выборка образует **вариационный ряд**, если ее члены являются **порядковыми статистиками**, т. е. выборочные значения случайной величины **X** упорядочены по возрастанию (ранжированы), значения же признака называются **вариантами.**

Основные способы организации выборки

Достоверность статистических выводов и содержательная интерпретация результатов зависит от **репрезентативности** выборки, т.е. полноты и адекватности представления свойств генеральной совокупности, по отношению к которой эту выборку можно считать представительной. Изучение

статистических свойств совокупности можно организовать двумя способами: с помощью **сплошного** и **несплошного наблюдения.**

Сплошное наблюдение предусматривает обследование всех **единиц** изучаемой **совокупности**, а **несплошное (выборочное) наблюдение** — только его части.

Существуют пять основных способов организации выборочного наблюдения:

1. **простой случайный отбор**, при котором nобъектов случайно извлекаются из генеральной совокупности Nобъектов (например с помощью таблицы или датчика случайных чисел), причем каждая из возможных выборок имеют равную вероятность. Такие выборки называются **собственно-случайными**;

2. **простой отбор с помощью регулярной процедуры** осуществляется с помощью механической составляющей (например, даты, дня недели, номера квартиры, буквы алфавита и др.) и полученные таким способом выборки называются **механическими**;

3. **стратифицированный** отбор заключается в том, что генеральная совокупность объема Nподразделяется на подсовокупности или слои (страты) объема $N_1,N_2,...,N_r$так что $N_1+N_2+...+N_r=N$. Страты представляют собой однородные объекты с точки зрения статистических характеристик (например, население делится на страты по возрастным группам или социальной принадлежности; предприятия — по отраслям). В этом случае выборки называются **стратифицированными** (иначе, **расслоенными, типическими, районированными**);

4. методы **серийного** отбора используются для формирования **серийных** или **гнездовых выборок**. Они удобны в том случае, если необходимо обследовать сразу "блок" или серию объектов (например, партию товара, продукцию определенной серии или население при территориально-административном делении страны). Отбор серий можно осуществить собственно-случайным или механическим способом. При этом проводится сплошное обследование определенной партии товара, или целой территориальной единицы (жилого дома или квартала);

5. **комбинированный** (ступенчатый) отбор может сочетать в себе сразу несколько способов отбора (например, стратифицированный и случайный или случайный и механический); такая выборка называется **комбинированной.**

Виды отбора

По **виду** различаются индивидуальный, групповой и комбинированный отбор. При **индивидуальном отборе** в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при **групповом отборе** — качественно однородные группы (серии) единиц, а **комбинированный отбор** предполагает сочетание первого и второго видов.

По **методу** отбора различают **повторную и бесповторную** выборку.

Бесповторным называется отбор, при котором попавшая в выборку единица не возвращается в исходную совокупность и в дальнейшем выборе не участвует; при этом численность единиц генеральной совокупности **N** сокращается в процессе отбора.

При **повторном** отборе **попавшая** в выборку единица после регистрации возвращается в генеральную совокупность и таким образом сохраняет равную возможность наряду с другими единицами быть использованной в дальнейшей процедуре отбора; при этом численность единиц генеральной совокупности **N** остается неизменной (метод в социально-экономических исследованиях применяется редко). Однако, при большом **N (N → ∞)** формулы для **бесповторного** отбора приближаются к аналогичным для **повторного** отбора и практически чаще используются последние (**N = const**).

По своей природе распределения бывают **непрерывными** и **дискретными**. Наиболее известным непрерывным распределением является **нормальное**.

В зависимости от вида распределения и от способа отбора единиц совокупности по-разному вычисляются характеристики параметров распределения: теоретическое и эмпирическое распределения.

Долей выборки k_n называется отношение числа единиц выборочной совокупности к числу единиц генеральной совокупности:

$k_n = n/N$.

Выборочная доля w — это отношение единиц, обладающих изучаемым признаком **x** к объему выборки **n**:

$w = n_n/n$.

Пример. В партии товара, содержащей 1000 ед., при 5% выборке **доля выборки k_n** в абсолютной величине составляет 50 ед. (n = N*0,05); если же в этой выборке обнаружено 2 бракованных изделия, то **выборочная доля брака w** составит 0,04 (w = 2/50 = 0,04 или 4%).

Так как выборочная совокупность отлична от генеральной, то возникают **ошибки выборки**.

1.7 Шкалы измерений

Состояние объекта оценивается по критериям. В качестве критериев могут выступать: выживаемость животных, степень интоксикации, сохранение жизненно важных функций и т.д.

Оценки измеряются в той или иной шкале. *Шкала* (условно говоря, шкала – это множество возможных значений оценок по критериям) – числовая система, в которой отношения между различными свойствами изучаемых явлений, процессов переведены в свойства того или иного множества, как правило – множества чисел.

Различают несколько ***типов шкал:***

Во-первых, можно выделить ***дискретные шкалы*** (в которых множество возможных значений оцениваемой величины конечно – например, оценка в баллах – «1», «2», «3», «4», «5») и ***непрерывные шкалы*** (например, концентрация вещества в моль/л или активность фермента в сыворотке крови в мКат/л).

Во-вторых, выделяют ***шкалы отношений, интервальные шкалы, порядковые*** **(*ранговые*)** ***шкалы*** **и** ***номинальные шкалы*** (шкалы наименований).

Шкала отношений – самая мощная шкала. Она позволяет оценивать, во сколько раз один измеряемый объект больше (меньше) другого объекта, принимаемого за эталон, единицу. Для шкал отношений существует

естественное начало отсчета (нуль), но нет естественной единицы измерений. Шкалами отношений измеряются почти все физические величины – время, линейные размеры, площади, объемы, сила тока, мощность и т.д. В медико-биологических исследованиях шкала отношений будет иметь место, например, когда измеряется время появления того или иного признака после воздействия (порог времени, в секундах, минутах), интенсивность воздействия до появления какого-либо признака (порог силы воздействия в вольтах, рентгенах и т.п.). Естественно, к шкале отношений относятся все данные в биохимических и электрофизиологических исследованиях (концентрации веществ, вольтажи, временные показатели электрокардиограммы и т.п.). Сюда же, например, относятся и количество правильно или неправильно выполненных «заданий» в различных тестах по изучению высшей нервной деятельности у животных.

Шкала интервалов применяется достаточно редко и характеризуется тем, что для нее не существует ни естественного начала отсчета, ни естественной единицы измерения. Примером шкалы интервалов является шкала температур по Цельсию, Реомюру или Фаренгейту. Шкала Цельсия, как известно, была установлена следующим образом: за ноль была принята точка замерзания воды, за 100 градусов – точка ее кипения, и, соответственно, интервал температур между замерзанием и кипением воды поделен на 100 равных частей. Здесь уже утверждение, что температура 300С в три раза больше, чем 100С, будет неверным. В шкале интервалов сохраняется отношение длин интервалов. Можно сказать: температура в 300С отличается от температуры в 200С в два раза сильнее, чем температура в 150С отличается от температуры в 100С.

Порядковая шкала (шкала рангов) – шкала, относительно значений которой уже нельзя говорить ни о том, во сколько раз измеряемая величина больше (меньше) другой, ни на сколько она больше (меньше). Такая шкала только упорядочивает объекты, приписывая им те или иные баллы (результатом измерений является нестрогое упорядочение объектов). Например, так построена шкала твердости минералов Мооса: взят набор 10 эталонных минералов для определения относительной твердости методом царапанья. За 1 принят тальк, за 2 – гипс, за 3 – кальцит и так далее до 10 – алмаз. Любому минералу соответственно однозначно может быть приписана определенная твердость. Если исследуемый минерал, допустим, царапает кварц (7), но не царапает топаз (8), то соответственно его твердость будет равна 7. Аналогично построены шкалы силы ветра Бофорта и землетрясений Рихтера. Шкалы порядка широко используются в педагогике, психологии, медицине и других науках, не столь точных, как, скажем, физика и химия. В частности, повсеместно распространенная шкала школьных отметок в баллах (пятибалльная, двенадцатибалльная и т.д.) может быть отнесена к шкале порядка. В медикобиологических исследованиях шкалы порядка встречаются сплошь и рядом и подчас весьма искусно замаскированы. Например, для анализа свертывания крови используется тромботест: 0 – отсутствии свертывания в течение времени теста (а через минуту?), 1 – «слабые нити», 2 – желеподобный сгусток, 3 – сгусток, легко деформируемый, 4 – плотный, упругий, 5 – плотный, занимающий весь объем и т.п. Понятно, что интервалы между этими плохо отличимыми и очень субъективными позициями

произвольны. В этом случае фраза «Тромботест у исследуемых животных повышался в среднем с 3,3 до 3,7» выглядит абсурдной. Масса подобных шкал все еще встречается в экспериментальной токсикологии, экспериментальной хирургии, экспериментальной морфологии.

Частным случаем порядковой шкалы является *дихотомическая* шкала, в которой имеются всего две упорядоченные градации – например, «выжил после эксперимента», «не выжил».

Шкала наименований (номинальная шкала) фактически уже не связана с понятием «величина» и используется только с целью отличить один объект от другого: номер животного в группе или присвоенный ему уникальный шифр и т.п.

Лекция 2. ЭЛЕМЕНТЫ ТЕОРИИ ПЛАНИРОВАНИЯ ИССЛЕДОВАНИЙ

2.1 Цели и задачи науки. Предмет биометрии – изучение свойств массовых явлений в биологии. Эти явления обычно представляются сложными вследствие разнообразия (варьирования) отдельных индивидуумов или единиц. Чтобы получить правильное представление об изучаемых свойствах массовых явлений и дать им определенные количественные оценки, их подвергают совместному рассмотрению и анализу. Отдельные единицы или индивидуумы, обладающие некоторым общим свойством, объединяют в совокупности. Наблюдаемые единицы называют вариантами (данными, датами), а образуемую совокупность единиц – статистической совокупностью. Статистическая совокупность может быть образована по одному или по нескольким признакам. Она может состоять из одной или нескольких однородных в отношении изучаемого свойства групп. Однако часто бывает целесообразно подразделить отдельные наблюдаемые единицы на группы для достижения большей однородности их внутри этих групп.

Теорию и методы изучения свойств массовых явлений, вычисления и анализа их количественных характеристик изучает биометрия. Метод изучения массовых явлений основан на теории вероятностей. Теория вероятностей устанавливает закономерности событий, наступающих случайно и называемых случайными. Статистика предполагает анализ массовых явлений, имеющих также случайный характер в распределении значений отдельных единиц, составляющих явление.

Центральной задачей биометрии как метода исследования являются заключения, выходящие за рамки изученного материала, т. е. заключения о свойствах статистических совокупностей, принимая во внимание и неизученную их часть.

Всю статистическую совокупность, в отношении которой делают статистические обобщения и заключения, называют общей, или генеральной совокупностью, а часть ее, охваченную непосредственным наблюдением, называют выборочной совокупностью.

Вариационная статистика применяет метод оценки общей совокупности на основе изученных отдельных единиц или на основе выборочных совокупностей.

2.2 Статистические заключения. Статистические заключения о свойствах генеральных совокупностей по выборочным всегда имеют вероятностный характер, т. е. делаются с определенной степенью безошибочности и никогда не делаются с полной достоверностью.

Статистические заключения, как главная составная часть метода исследования массовых явлений, имеют свои отличительные черты. Статистические заключения делают с численно выраженной определенностью. Теоретической основой для их построения является раздел математики, изучающий закономерности случайных событий и называемый теорией вероятностей. Предпосылка, что результаты статистического наблюдения отобраны в случайном порядке из соответствующих генеральных

совокупностей, дает возможность в соответствии с теорией вероятностей оценить степень отклонения результатов наблюдения от соответствующих показателей генеральной совокупности. Таким образом, вероятностная основа вариационной статистики позволяет оценить степень точности получаемых результатов опыта. Основу изучения природных процессов составляет выявление причинно-следственных связей между явлениями экспериментальным путем.

2.3 Теория вероятностей. Осуществив по своему желанию одно или несколько первоначальных явлений (в дальнейшем они называются факторами), экспериментатор получает возможность изучать появляющиеся явления − следствия. Иногда в процессе эксперимента удается сделать случайное открытие, т. е. обнаружить явление − следствие, о котором ранее ничего не было известно. Но, как правило, экспериментатор заранее намечает явления-следствия, появление которых он ожидает. При этом самое сложное явление можно разбить на частные, мелкие явления, относительно которых остается выяснить: произошла они или не произошли. Например, обрабатывая семена на всхожесть определенным препаратом, экспериментатор мог поставить задачу оценить эффект различных его доз. В качестве эффекта могло быть принято число всхожих и невсхожих семян.

Измеряя массу какого-либо вещества, в качестве отдельных частных явлений можно рассматривать всевозможные априорные значения этой массы. Задача экспериментатора, таким образом, сводится к наблюдению того, какие из значений массы осуществились.

Явления, рассматриваемые с той точки зрения, осуществились они или не осуществились, называются событиями. Применительно к событиям ставится основная задача: предсказать, появится ли изучаемое событие при осуществлении некоторого наперед заданного комплекса факторов (явлений − причин). Событие, которое при заданном комплексе факторов обязательно произойдет называется достоверным. Событие, которое при заданном комплексе факторов не может произойти, называется невозможным событием. Суждения о достоверности или невозможности некоторого события являются категорическими суждениями. Такие суждения принято, считать окончательным результатом исследования. Отсюда возникает интерес к обратной задаче: указать комплексы факторов, при которых о заданном событии можно сделать категорические суждения.

Однако каждое событие − результат действия многих факторов, часть из которых иногда нельзя предсказать или организовать в опыте. В этом случае категорическое суждение о событии невозможно. Получается ситуация: заданные факторы благоприятствуют событию, и, следовательно, оно может произойти. С другой стороны, действия этих факторов недостаточно, чтобы гарантировать появление события, и, значит, оно может и не произойти.

Событие, которое при заданном комплексе факторов может либо произойти, либо не произойти, называется случайным событием. Случайные события связаны с действием не вошедших в организованный комплекс факторов, называемых случайными факторами в отличие от другой группы факторов, включаемых в комплекс и называемых основными, или

неслучайными.

Предположим, исследуется урожайность культур. Такие факторы, как технология возделывания, внесение различных доз удобрений и т.д. можно организовать в опыте, т. е. учесть. Эти факторы являются основными. Другая группа факторов является неизвестной, или не поддающейся учету. Эти факторы при статистическом анализе получили название случайных.

Для того чтобы выяснить, произойдет или не произойдет событие при заданном комплексе факторов, нужно осуществить этот комплекс, т. е. провести испытание. Испытанием является любой эксперимент, в результате которого производят наблюдения.

Предсказать результат единичного испытания можно только для достоверных или невозможных событий. Случайность же события не видна из единичного испытания. Любое случайное событие по единичному испытанию было бы оценено как достоверное, если оно произошло, и как невозможное − если не произошло. Такие оценки, однако, были бы сами случайными, как и результат единичного испытания. Теория оценки случайных событий строится на большом числе испытаний, т. е. для массовых событий.

Важным условием при этом является неизменность комплекса основных факторов. События, происходящие при одном и том же комплексе факторов, называются однородными. Установлено, что однородные случайные события в большой их массе подчиняются некоторым закономерностям. Эти закономерности получили название вероятностных.

Характер вероятностных закономерностей можно уяснить на следующих примерах.

Пример. При подбрасывании монеты возможны два события: выпадение монеты гербом или решкой. События с одинаковыми возможностями осуществления называются равновозможными. Так, при симметричной монете выпадение герба и цифры − равновозможны.

Однако, если бы было произведено, например, 1000 бросаний, и из них 700 раз выпал герб, то для следующей серии испытаний можно было бы предсказывать, что герб появится в 70% случаев. Причем такое отклонение от ожидаемых 700 появлений герба из 1000 бросаний можно было бы считать связанным с несимметричностью монеты.

Установленное в результате опыта отношение числа появления события к общему числу всех испытаний называется частотой события. В указанном примере с монетой частота выпадения герба равна 0,7.

Из примера можно заключить, что частота события, выступающая как некоторая статистическая закономерность, связана с внутренними характеристиками события. Частота является мерой этих внутренних характеристик события. Она тем надежнее, чем большее число испытаний было произведено. При очень большом числе испытаний частота почти перестает изменяться, приближаясь к некоторой величине. Эту величину и можно принять за интересующую нас числовую характеристику. Так, при бросании монеты 4, 12 и 24 тыс. раз частота появления герба соответственно равнялась 0,6080; 0,5016; 0,5005. Очевидно, что она здесь приближается к числу 0,5.

Числовая характеристика случайного события, обладающая тем

свойством, что для любой достаточно большой серии испытаний частота события лишь незначительно отличается от этой характеристики, называется вероятностью события.

Из этого рассмотрения устанавливаем, что вероятность является тем теоретическим пределом, к которому стремится частота событий при увеличении числа испытаний. Вероятность − идеальное выражение частоты событий.

Данное определение вероятности называется статистическим. Это определение не является достаточно строгим с точки зрения математики. По статистическому определению трудно изучать свойства вероятности.

Однако имеется и ряд положительных его свойств. Статистический подход позволяет находить вероятности событий, структура которых неизвестна. Например, только статистический подход позволил определить вероятность рождения мальчиков, равную 0,52 и девочек − 0,48.

Существуют два других, более удобных с формальной точки зрения, определения вероятности: классическое и геометрическое. Однако для них требуется знать структуру рассматриваемых событий.

Понятие о геометрическом определении вероятности можно получить из следующего примера испытаний.

Пример. Предположим, в некотором квадрате случайным образом выбирается точка. Какова вероятность, что она окажется в области *D*. Очевидно, что вероятность эта будет тем большей, чем больше область *D*. В качестве мерила вероятности выступает здесь площадь. Вероятность того, что случайная точка попадет в область *D* (осуществление события *D*) равна: $p(D)=S_D/S$, где S_D−площадь области *D; S* − площадь всего квадрата.

Геометрическое определение вероятности пригодно не только для плоскости, но и для прямой или пространства.

В первом случае основой для определения вероятности служит некоторый отрезок, а случайным событиям соответствуют его части. Вероятность вычисляется как отношение длины частей к общей длине отрезка. Во втором, случае основой к испытанию принимают некоторый куб, случайным событиям соответствуют различные тела, расположенные в кубе. Вероятность вычисляют как отношение объемов тел к объему куба.

Наибольший интерес представляет классическое определение вероятности. С этим определением связаны основные теоремы теории вероятностей.

Вероятность здесь определяется априори, до испытаний, исходя из определенной структуры случайных событий, т. е. из разбивки на равновозможные исходы.

Пример. Пусть при подбрасывании монеты появления герба или цифры будут изучаемыми событиями ***a*** и ***b***. Причем, если при одном бросании произойдет событие ***a***, то не произойдет другого события ***b***. Такие события называют несовместными. Каждое из событий называют исходом испытания. В силу равновозможности исходов в нашем испытании вероятность каждого события равна. При единичном бросании кубика с 6 гранями (имеющими, например, 1, 2, 3, 4, 5, 6 очков), вероятность появления любой одной грани

p= 1/6.

Исходы испытания являются простейшими случайными событиями. Можно рассматривать более сложные события, объединяющие несколько исходов. Например, при бросании игрального кубика мы можем интересоваться таким событием, как выпадение числа очков больше 2. В таком случае говорят, что появлению события с выпадением больше двух очков, т. е. с 3, 4, 5 и 6 очками, благоприятствуют четыре исхода из шести. Вероятность этого события *p=4/6*. Таким образом, мы подошли к классическому определению вероятности. Вероятностью случайного события называется отношение числа отходов, благоприятствующих событию, к числу всех возможных исходов.

2.4 Основные теоремы теории вероятностей. Если некоторое событие *A* может произойти при *n* испытаниях и *m* − число исходов, которые благоприятствуют наступлению события, то вероятность того, что данное событие произойдет, может быть определена как *P(A)=m/n*. Тогда, сумма вероятностей двух несовместных событий равна единице.

Сложение вероятностей. $P(A+B)=P(A)+P(B)=m_1/n + m_2/n$

Если в урне с 10 шарами 6 шаров черных, 3 белых и 1 зеленый, вероятности этих событий будут равны, соответственно, 6/10, 3/10 и 1/10.

Какова вероятность вынуть белый или зеленый шар?

Благоприятствует появлению белого шара 3/10 всех исходов, а зеленого шара − 1/10 исходов. Появлению либо белого, либо зеленого шара соответствует *p*=3/10+1/10=4/10=0,25, т. е. вероятность суммы двух несовместных (взаимоисключающих случайных) событий равна сумме их вероятностей.

Умножение вероятностей. Два события называются независимыми, когда наступление одного не оказывает влияния на наступление другого. Так, результат одного метания кости не влияет на результат следующего метания.

Вероятность сложного события (т. е. наступления двух событий независимых одно от другого равна произведению вероятностей отдельных событий. $P(AxB)=P(A)xP(B)=m_1/n$ x m_2/n

Например, вероятность выпадения очка, а затем двух очков, при двух последовательных бросаниях кубиков, равна *p*==1/6x1/6=1/36.

Вычисление вероятностей. Часто возникает необходимость одновременно складывать и умножать вероятности. Например, требуется определить вероятность выпадения 5 очков при одновременном бросании 2 кубиков. Искомая сумма вероятностей может получиться как результат одной из следующих 4-х комбинаций исходов:

кубик *a* 1, 2, 3, 4;

кубик *b* 4, 3, 2, 1

Вероятность получения одного очка на кубике *a* равна 1/6 и получения четырех очков на кубике *b* −также 1/6. Вероятность получения комбинации этих очков равна 1/36. Аналогично и вероятность трех других комбинаций равна 1/36. Но любой из этих четырех результатов, дающий в сумме 5 очков, будет считаться благоприятным исходом. Отсюда вероятность искомого исхода

p = 1/36 + 1/36 + 1/36 + 1/36 = 1/9.

Более общая форма вопроса о вероятности события является такой:

какова вероятность получения не менее, например, 8 очков при бросании 2 костей? Число очков, равное и более 8, рассматривается как благоприятный исход.

Рассчитаем вероятность каждого благоприятного результата:

Вероятность появления 12 очков	1/36
Вероятность появления 11 очков	2/36
Вероятность появления 10 очков	3/36
Вероятность появления 9 очков	4/36
Вероятность появления 8 очков	5/36
Сумма вероятностей	15/36

Вероятность выпадения по меньшей мере 8 очков при бросании 2 костей равна 15/36 или 5/12.

Биномиальное разложение и измерение вероятностей

Изложенные примеры исчисления вероятностей можно обобщить на основе следующей ниже иллюстрации вывода.

Если подбрасываются одновременно 2 монеты (***a, b***), то существуют 4 возможных случая выпадения герба Т и цифры Н:

ab ab ab ab

TT TH HT HH

В первом исходе имеем 2 герба. Принимая это за 2 благоприятных исхода, получим вероятность каждого из них ***p***, а сложного события (ТТ) $p*p=p^2$. В данном случае, при $p=1/2$ $p^2=l/4$.

Четвертый из возможных исходов НН представляет 2 неблагоприятных исхода с вероятностью $q*q=q^2=1/4$.

Каждый из двух других исходов является комбинацией одного благоприятного и одного неблагоприятного случаев.

Вероятность каждого из этих исходов равна $1/4=pq=l/2*l/2$, а обоих вместе ТН и НТ равна их сумме, т. е. $2pq=1/2$.

Обобщенным выражением процесса получения вероятностей различных сочетаний независимых событий, когда вероятности их известны, являются последовательные члены разложения бинома.

Для рассматриваемого примера из двух событий имеем:

$(p+q)^2 = p^2 + 2pq + q^2$. , При $p = 12$ получим $(1/2+ 1/2)^2= 1/4+ 1/2 + 1/4$.

Если 3 монеты *a, b, c* подбрасываются одновременно, получим 8 возможных комбинаций:

Abc abc abc abc abc abc abc abc

TTT TTH THH THT HTT HTH HHT HHH

Вероятность выпадения 3 гербов составит 1/8, 2 гербов (в сочетании с одним случаем цифры) равна 3/8, одного герба и 2 цифр − 3/8, ни одного герба−1/8. При 3 независимых событиях степень бинома равна 3.

Вероятности отдельных возможных исходов даются последовательными членами разложения:

$(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$.

При $p=q= 1/2$ имеем $(1/2 +1/2)^3 = 1/8 + 3/8 + 3/8 + 1/8$, т. е. то же, что и непосредственным подсчетом.

Если число независимых случайных событий n, то вероятность n, $n-1$, $n-2$ и т. д. благоприятных исходов равна последовательным членам разложения:

$(p + q)^n$

Если желаем получить вероятные численности разных исходов при данном числе испытаний n, применяем выражение:

$N(p+q)^n$.

Например, при числе испытаний $N=200$ и двух независимых событиях n в каждом испытании вероятные численности будут равны $200\ (p+q)^2=200(p^2+2pq+q^2)$. Если $p=q=1/2$, имеем последовательные вероятные численности: 50+100+50.

При подбрасывании монеты 200 раз ($N=200$) выпадения герба следует ожидать в 50 случаях, герба или цифры − в 100 случаях и цифры − 50 случаях.

При тех же p и N, но $n=3$ получим последовательные вероятные численности: 25+75+75+25, которые означают 3, 2, 1 наступление события и ненаступление его, причем сумма всех численностей равна N.

При 200 бросаниях трех монет ожидаем в 25 случаях выпадения 3 гербов (ТТТ), в 75 случаях выпадения 2 гербов и одной цифры (ТТН), в 75 случаях выпадения 2 цифр и одного герба (ННТ) и в 25 случаях − 3 цифр.

Итак, когда вероятности независимых событий известны априори, то можно определить вероятные численности любого данного числа n, $n-1$, $n-2$.... наступления события и ненаступления его. При этом неважно, равны или не равны p и q, лишь бы они оставались при испытаниях постоянными. Этот факт имеет большое значение в теории статистики и используется ниже.

При изучении природных явлений выделение элементарных событий и вообще расчленения причинного процесса, в результате которого происходят случайные события, обычно невозможно. Классический подход к определению вероятности здесь бессилен. Проблему определения вероятностей таких событий решают на основе статистического подхода.

Однако классический подход к определению вероятностей событий лежит в основе теории анализа случайных событий и теоретических (модельных) распределений исходов испытаний. В свою очередь теория математического анализа случайных событий и модели распределений исходов испытаний являются базой статистических методов, в частности, базой статистических заключений.

Лекция 3 ОПИСАТЕЛЬНАЯ СТАТИСТИКА

3.1 Характеристика совокупности. Всякое множество отдельных отличающихся друг от друга и в то же время сходных в некоторых существенных отношениях объектов составляет так называемую совокупность. (популяции рыжих полевок того или иного района, стадо коров данного хозяйства, потомство определенного быка, заготовляемые в области или крае

беличьи шкурки, растения на опытных делянках, группа цыплят, на которых ставится опыт по применению антибиотиков, мальки окуня в озере и т. д.) Понятие совокупности применимо не только к животным и растениям. Такими же совокупностями являются, например, дети, родившиеся в стране в течение какого-то года пли месяца, молекулы газа в том или другом объеме.

В состав совокупности входят различные члены, или единицы: для популяции животных − каждое отдельное животное, для стада коров единицей является каждая корова, для совокупности шкурок − каждая шкурка, для потомства быка − каждый теленок, от него полученный, для совокупности зерен гречихи − каждое отдельное зерно.

Обычно число единиц совокупности называют объемом совокупности и обозначают латинской буквой n. Единица совокупности может характеризоваться определенными признаками, например: коровы − удоями за лактацию, весом, мастью; молекулы газа − скоростями их движения и т. д. Каждый изучаемый признак принимает разные значения у различных единиц совокупности, он меняется в своем значении от одной единицы совокупности к другой. Это различие между единицами совокупности называется вариацией или дисперсией (т. е. рассеянием).

Мы говорим − признак варьирует. Это означает, что он принимает различные значения совокупностей. Так, совокупность из всех животных данной у разные членов совокупности, например, у коров данной породы, мышей опытной группы поросят одного помета и т д. Значение или меру признака единицы совокупности называют вариантой и обозначают буквой x. Значок i – порядковый номер варианты. Несмотря на различия между вариантами по значению изучаемого признака, совокупность этих вариант обладает однородностью. Беличьи шкурки неодинаковы по окраске, размеру, качеству меха, но они однородны, так как все они – шкурки особей одного и того же вида – белки обыкновенной.

Различают совокупности:

1. Генеральную
2. Выборочную (выборка).

Генеральная – теоретически бесконечная совокупность всех единиц или членов, которые могут быть отнесены к ней. Из-за бесконечно большого числа членов генеральную совокупность изучить практически невозможно. Поэтому из нее выбирают часть для непосредственного изучения, т.е выборку.

Существует несколько способов отбора вариант в выборку:

1. плановый отбор (групповой отбор; гнездовой (или серийный));
2. стихийный отбор (механический).

Единственное условие – однородность отбираемого в выборку материала.

Задачей изучения всякой совокупности является получение статистических (или, как иногда говорят, биометрических) характеристик, или показателей, которые позволяют судить о данной совокупности в целом, о различиях внутри нее и об отличии ее от других, сходных с ней или близких к ней совокупностей. Совокупность становится статистической тогда, когда в ее описание вносится количественный метод. Применение количественного

метода изучения совокупности и позволяет получать для нее статистические характеристики, с помощью которых получают основную информацию о совокупности.

3.2 Варьирующие признаки и их учет. При изучении единиц совокупности по признаку необходимо записать полученные данные и сгруппировать их. Способы группировки зависят от характера вариации изучаемых признаков.

Различают следующие типы вариации признаков:

- качественная;
- количественная

Если различия между вариантами выражаются в каких-то качествах, то такую вариацию называют качественной. Если совокупность животных характеризуют по масти, тогда каждая варианта должна получить качественную характеристику в соответствии с заранее принятыми обозначениями: черная, рыжая, черно-пестрая, черно-рыжая и т. д. В этом простейшем случае подсчет числа особей в каждой из выделенных групп дает представление о составе популяции в целом.

В других случаях различия между вариантами будут количественными. Количественная вариация может быть двух типов: прерывная (дискретная) и непрерывная. В первом случае различия между вариантами, отдельными значениями случайной переменной, выражаются целыми числами, между которыми нет и не может быть переходов. Например, количество детенышей в помете (поросят у свиноматок, щенков у серебристо-черных лисиц), число сосков у сриноматок, число лучей в плавниках рыб, количество лепестков в цветке, число позвонков у птиц и т. д. Для изучения подобного варьирования надо сосчитать у каждой единицы совокупности число изучаемых элементов и записать его на соответствующую карточку. При непрерывной вариации значения вариант не обязательно выражаются только целыми числами. Все зависит от того, какая степень точности принимается для характеристики данного количественного признака. Так, например, при изучении веса крупного рогатого скота можно ограничиться значениями вариант, выраженными в килограммах, отбросив граммы, но совершенно недостаточно округлять до килограммов веса рыб, так как грамм здесь имеет большое значение. В опытах же по изучению влияния гормонов на рост гребня у цыплят вес гребня придется измерять в миллиграммах. Молочную продуктивность за лактацию обычно выражают в килограммах, но общая картина удоев не изменится, если округлять ее до десятков килограммов. Оценка же жирности молока в процентах, выраженных целыми числами, явно недостаточна, ее надо давать с учетом десятых и даже сотых долей процента. Однако во всех этих и им подобных случаях существует непрерывная вариация, выражающаяся в том, что между вариантами возможны все переходы. При изучении непрерывной вариации надо все единицы совокупности характеризовать количественно с той степенью точности, которая заранее намечена и больше всего подходит в данном конкретном случае.

3.3 Группировка данных при качественной вариации. Чтобы проанализировать ту или иную совокупность, необходимо сгруппировать

полученные отдельные варианты и затем представить эту группировку в виде таблицы или ряда. При упорядочении полученных данных легко обработать их математически и вывести статистические показатели, которые будут исчерпывающе характеризовать изучаемую совокупность. Проблема группировки занимает большое место в статистике вообще (особенно в экономической), так как ошибочная группировка данных может привести к неправильным выводам о существе изучаемого явления.

Наиболее проста группировка при качественной вариации. Так, если норки различаются по окраске, то их распределение может быть выражено в количестве животных каждой окраски и в процентах, которые составляют норки каждой окраски от общего количества животных.

Частным случаем качественной вариации является альтернативная, когда в совокупности можно выделить только две группы. У членов одной группы присутствует определенное качество (или признак), у членов другой группы его нет. Так, при проверке на туберкулез животные распадаются на 2 группы − с положительной реакцией и с отрицательной. Одни коровы в данном стаде рогатые, другие − комолые и т. д.

Группировка данных при количественной дискретной вариации. При количественной вариации необходимо предварительно наметить для таблицы классы, охватывающие все полученные количественные данные от минимальных до максимальных. Это легко сделать при прерывной (дискретной) количественной изменчивости.

Допустим, что была изучена плодовитость 80 самок серебристо-черных лисиц, т. е. число родившихся у каждой самки щенков. Варианты x_1, x_2,, x_3, ..., x_n этой совокупности выражены цифрами, представленными в табл. 1.

Таблица 1

Количество щенков у 80 самок серебристо-черных лисиц

4	5	3	4	6	7	8	3	1	4
6	4	4	3	2	5	3	4	5	4
5	3	4	5	4	4	4	6	5	7
6	4	5	4	4	4	4	2	3	4
5	5	4	5	4	4	6	4	4	4
4	8	7	5	4	9	4	3	4	4
5	4	6	4	4	3	4	4	4	2
4	4	5	4	6	4	3	3	4	2

Группировку вариант лучше всего провести по значениям отдельных вариант. Минимальное число щенков 1, максимальное − 9. Отсюда естественно установить 9 классов: с 1 щенком, с 2, 3 и т. д.− и распределить все варианты по этим 9 классам. Наиболее простым способом разнесения вариант по классам является следующий.

Составляется таблица («классы» и «частоты») с намеченными 9 классами и в соответствующие горизонтальные строчки разносятся все варианты, начиная от первой. Обозначаются они так: первые четыре варианты данного класса − точками, а последующие − черточками, соединяющими четыре точки. (конвертик, домик, елочка).

Пример разноски:

Классы, x	Разноска	Частота, f
1	.	1
2	. .	
	. .	4
3	⊠	10
4	⊠ ⊠ ⊠ ⧅	39
5	⊠ ¨	
	·	13
6	⊏	7
7	. .	
	·	3
8	. .	2
9	.	1

Вторичная группировка данных при количественной дискретной вариации. В разобранном выше примере классов намечено столько, сколько было в изученной совокупности различных значений вариант (от 1 до 9 щенков). Однако такой способ будет нецелесообразным при очень большой вариации дискретного признака.

Так, например, у змеи *Lampropeltis getulus* количество хвостовых щитков варьировало от 40 до 58 (табл. 2). . .

Таблица 2

Количество хвостовых щитков у 60 экземпляров змеи *Lampropeltis getulus*

42	58	44	54	41	50	46	46	54	48	43	49
50	48	46	46	45	53	48	48	53	53	48	41
46	40	50	43	49	51	52	46	42	44	48	45
47	46	43	50	47	45	48	40	44	42	48	45
54	50	56	48	45	45	51	42	44	47	46	45

Если классы намечать по значениям каждой варианты, т. е. 40, 41 и т. д., то получится 19 классов, ряд окажется растянутым, труднообозримым, с перерывами в некоторых классах. Лучше наметить классы, охватывающие несколько значений вариант, например: 40−41, 42−43 и т. д. или 40−42, 43−45 и т. д. В первом случае вариационный ряд будет состоять из 10 классов, во втором − из 7. Имеем классовый промежуток – I, равен 3.

Таблица 3

Границы классов	Средний класс, x	Разноска	Частота, f
40-42	41	□	8
43-45	44	□	14
46-48	47	□□	20
49-51	50	□	9
52-54	53	□	7
55-57	56	.	1
58-60	59	.	1

3.4 Вариационный ряд и его графическое изображение. После

распределения вариант по классам получаются ряды, показывающие как часто встречаются варианты каждого класса и как варьируют признак от минимума до максимума. Т.о., ВР – двойной ряд чисел, показывающий распределение вариант по их частоте или встречаемости. По ВР можно судить не только о границах, но и о характере вариации.

Класс, обладающий наибольшей частотой, получил название модального, значения же крайних классов называют лимитами или пределами.

Всякий вариационный ряд можно изобразить графически. Графическое изображение вариационного ряда в общем виде получило название кривой распределения или вариационной кривой.

Существуют два способа графического изображения конкретных вариационных рядов. Первый из них, применяющийся при дискретной вариации, но в том случае, если классы намечены по отдельным значениям вариант, носит название полигона распределения. На оси абсцисс нанесены классы, на оси ординат−частоты. Высота каждого класса, пропорциональная частоте класса, отмечается кружком. При непрерывной вариации, если классы намечены по границам, на оси абсцисс наносят нижние границы классов, на оси ординат−частоты. Такой график носит название – гистограммы.

При статистической обработке материала возникает вопрос: сколько классов необходимо намечать? Это зависит от:

- объема совокупности;
- от величины вариационного размаха.

На практике можно руководствоваться примерно следующими правилами:

Количество вариант	Число классов
25−40	5−6
40−60	6−8
60−100	7−10
100−200	8−12
более 200	10−15

Вариационный ряд при непрерывной изменчивости также может быть изображен на графике. В этом случае нужно строить гистограмму, т. е. ступенчатую диаграмму.

Характер распределения вариант в вариационном ряду.

Изучая распределение вариант в вариационных ряду легко заметить некоторые общие закономерности, а именно:

1) большинство вариант располагается в средней части вариационного ряда или около середины вариационной кривой, здесь наблюдается максимум вариант, как бы их сгущение;

2) распределение вариант в обе стороны от этого максимума более или менее симметрично;

3) частота вариант постепенно убывает к краям вариационного ряда.

Эти закономерности в той или иной степени присущи любому вариационному ряду. В дальнейшем мы увидим, что закономерности вариационного ряда основываются на закономерностях случайной вариации, изучаемых теорией вероятностей.

Лекция 4 ОПИСАТЕЛЬНАЯ СТАТИСТИКА. СРЕДНИЕ ВЕЛИЧИНЫ

4.1 Две группы показателей для характеристики вариационных рядов. В предыдущей лекции мы рассмотрели способы сведения данных, составляющих статистические совокупности, в вариационные ряды. Каждый вариационный ряд и его графическое изображение − это как бы «сгущение» исходного фактического материала, превращение его в наглядную форму. Однако этого недостаточно. Очень важно получить характеристики для совокупности, которые были бы выражены цифровыми показателями. С их помощью можно было бы сравнивать разные ряды. Одним из простейших способов количественной характеристики вариационного ряда является указание на его размах, т. е. на верхнюю и нижнюю его границы, которые обычно называют лимитами. Если, например, известно, что вариационный ряд по молочной продуктивности одного стада коров имеет размах от 2000 до 4000 кг, а другого−от 2500 до 6800 кг, то, казалось бы, можно сделать вывод о более высоком качестве второго стада. Однако лимиты не указывают на то, как распределяются по изученному признаку отдельные члены совокупности. Вот почему для характеристики совокупности нужны такие показатели, которые отражали бы свойства всех ее членов.

Вариационные ряды могут различаться: а) по тому значению признака, вокруг которого концентрируется большинство вариант. Это значение признака отражает как бы уровень развития признака в данной совокупности, или, иначе, центральную тенденцию ряда, т. е. типичное для ряда; б) по степени вариации вариант вокруг уровня, по степени отклонения от центральной тенденции ряда.

Соответственно этому статистические показатели разделяются на две группы:

- показатели, которые характеризуют центральную тенденцию, или уровень ряда,
- показатели, измеряющие степень вариации.

К первой группе относятся различные средние величины: мода, медиана, средняя арифметическая, средняя геометрическая. Ко второй − вариационный размах, среднее абсолютное отклонение, среднее квадратическое отклонение, варианса (дисперсия), коэффициенты асимметрии и вариации. Существуют еще и другие показатели, но их мы не будем рассматривать, так как они редко применяются в биологической статистике.

Мода и медиана. При изучении распределения самок лисиц по числу щенков в помете обнаружилось, что 39 самок из общего числа 80 имели по 4 щенка, т. е. класс «4 щенка» обладал наибольшей частотой. Такой класс был назван модальным. Значение же модального класса называют модой. Мода обозначается символом Мо. Величина моды является как бы типичной для всей совокупности. Действительно, в нашем примере почти половина самок из 80 имела в помете именно 4 щенка.

Для ряда распределения змей по числу хвостовых щитков (табл. 2) модальным является класс «46−48 щитков». А так как класс здесь охватывает несколько значений вариант, то для его характеристики надо вычислить среднее значение класса. Оно равно 46 + 48/2= 47. В таком случае Мо = 47

щиткам.

К__числу средних величин относится также медиана. Медиана− это значение варианты, находящейся точно в середине ряда (обозначается Ме).

Чтобы найти такую варианту, надо сначала расположить все варианты по порядку от минимальных их значений до максимальных. Такое расположение вариант называют ранжировкой. Чтобы определить Ме при четном числе вариант, надо взять значения двух соседних срединных вариант, например при n = 80 значения вариант с порядковыми номерами 40 и 41, и разделить их сумму на 2. В примере, представленном в табл. 4, обе эти варианты будут иметь значения «4 щенка», следовательно, Ме данного ряда = 40.

Медиана и мода дают известное представление о совокупности в целом. Они характеризуют своего рода типичное в данной совокупности (конечно, речь идет только о каком-то определенном признаке).

Использование моды и медианы в биологии в настоящее время довольно ограничено, но в некоторых случаях без них очень трудно обойтись, в частности, если полученные данные не являются чисто количественными, а поэтому не могут быть представлены в виде точного вариационного ряда. Так, например, тяжесть заболевания подопытных животных или их упитанность можно условно оценивать степенями: слабая, удовлетворительная, средняя, высокая−или баллами 1, 2, 3 и т. д. Тогда мода или медиана могут достаточно хорошо характеризовать типичное в совокупности.

Обычно же, когда изучаемая совокупность достаточно_однородна и вариация внутри нее чисто количественная, выгоднее пользоваться, другими средними величинами.

2. Средняя арифметическая и ее свойства. Нахождение средней арифметической − это в сущности замена индивидуальных варьирующих значений признаков отдельных членов совокупности некоторой уравненной величиной при сохранении основных свойств всех членов совокупности. Этому условию в наибольшей степени удовлетворяет так называемая «средняя арифметическая, обозначаемая» - X (ранее обозначали M).

Представим себе, что ряд членов совокупности, т. е. ряд значений случайной переменной x_1, x_2, …, x_i заменим таким же рядом из одинаковых величин x, т. е. x, x, x, ..., x (n раз).

Тогда сумма всех вариант совокупности $x_1+x_2+x_3+\ldots+x_n$ будет равна

X+X+X+…+X (n раз), т. е. nX. Сумму всех вариант совокупности можно сокращенно обозначить x, (x − обозначает значение любой варианты; греческая буква Σ − большая сигма − обозначает суммирование; конкретные суммы часто обозначают также латинской буквой S). Тогда

$\Sigma x, = nx$, откуда

$X= \Sigma x/n$

Мы получили наиболее общую и в то же время наиболее простую формулу средней арифметической. Для того чтобы вычислить среднюю арифметическую, достаточно сложить значения всех вариант и сумму разделить на общее число вариант. В простейших случаях так и делают. Очевидно, в таких случаях можно пользоваться данными, полученными непосредственно при анализе членов совокупности, не прибегая к группировке

вариант.

Однако при большом количестве вариант этот прямой способ определения средней арифметической по указанным формулам оказывается не столь удобным. Кроме того, при его применении нет возможности вычислить некоторые другие биометрические показатели. Поэтому на практике часто пользуются окольными методами вычисления средней арифметической на основе уже сгруппированных данных. Эти методы будут разобраны позднее.

Свойства средней арифметической:

1. Если каждую из вариант совокупности, для которой вычисляется средняя арифметическая, увеличить или уменьшить на одну и ту же величину, то и средняя арифметическая соответственно увеличится или уменьшится на столько же.

X1 /A; X2 /A; X3 /A; X4 /A; Xi /A; то Σxi/A

2. Алгебраическая сумма отклонений отдельных вариант от средней арифметической (т. е. разностей между каждым кон¬кретным значением признака и средней арифметической) рав¬няется нулю:

Σ(xi - X)=0.

3. Сумма квадратов отклонений от средней арифметической меньше суммы квадратов отклонений от любой другой вели¬чины A не равной X, т. е.

Σ (xi-X)2 < Σ (xi−A)2, если A не равно X.

4.2 Значение средней арифметической и ее сущность. Средняя арифметическая, как и некоторые другие средние, известна издавна. Она имеет очень большое значение в науке и технике. Нет буквально ни одной биологической работы, в которой не встречались бы в той или другой форме средние арифметические. Средняя арифметическая является обобщающей величиной, которая как бы впитывает в себя все особенности данной совокупности или ряда распределения. Она отражает уровень всей совокупности в целом, дает сводную, обобщенную характеристику данного изучаемого признака.

Цифровое значение средней арифметической как таковое может не встретиться ни в одном конкретном случае в совокупности. Может оказаться, что ни одна варианта не будет ей равной. Если среднее число щенков у серебристо-черных лисиц равно 4,7, то, очевидно, фактическое число щенков никак не может быть дробным. В этом смысле средняя арифметическая является абстрактной величиной. Но в то же время она и конкретна. Она выражается в тех же единицах измерения, что и варианты ряда. При определении средней арифметической взаимопогашаются, отметаются случайные колебания, отклонения от центральной тенденции, от уровня вариационного ряда и выступает общий закон явления. Вскрывается типичное для всей совокупности в целом.

В то же время нужно предостеречь от возможных ошибок в понимании средней арифметической.

- Средняя арифметическая характеризует всю совокупность в целом, а не отдельные члены совокупности. Среднее число щенков в помете лисиц 4,7 относится только ко всей группе, каждая же отдельная лисица характеризуется своим числом щенков в помете−от 1 до 9.

• Средняя имеет смысл только по отношению к качественно однородной совокупности. Так, нельзя вычислять средний вес животных для группы, включающей и молодняк разных возрастов и взрослых животных. Надо взять каждую возрастную группу отдельно и для них вычислить X.

• Поскольку средняя арифметическая относится к данной совокупности, перенесение ее на явления, выходящие за ее рамки, рискованно, без специального анализа вопроса о правомерности такого перенесения.

• Средняя относится лишь к отдельным изучаемым признакам и не может быть автоматически перенесена на их сумму.

4.3 Измерение вариации. Вариационный размах и средние отклонения. Средняя арифметическая указывает на то, какое значение признака наиболее характерно для данной совокупности. Но сама по себе она еще недостаточна для характеристики совокупности, так как главной особенностью совокупности является наличие разнообразия между ее членами, т. е. вариации. Если бы не было вариации, то информацию о совокупности можно было бы получить по одному члену совокупности. При наличии же вариации эта информация должна быть основана на учете характера и степени вариации.

Учет вариации того или другого признака в совокупности имеет очень большое значение для биолога, так как всякая вариация в популяции животных или растений в конечном счете отражает различия между организмами − в их наследственной природе и в тех условиях, при которых они выращивались. Приемы работы с животными должны меняться в зависимости от характера их вариации. Без оценки вариации невозможно и сравнение двух совокупностей.

Два стада коров могут иметь очень близкие средние удои, но в одном величины удоев сильно различаются, в другом же коровы представляют собой довольно однородную группу с небольшим размахом колебаний. Определение вариационного размаха, т. е. разницы между максимальным и минимальным значениями вариант, может в известной степени указывать на степень вариации, но оно недостаточно. Во-первых, крайние величины в рядах не очень устойчивы, и при изменении количества изучаемых особей они легко сдвигаются. Во-вторых, при одних и тех же пределах вариации распределение вариант в рядах может быть различным. Вот почему для характеристики различий между отдельными значениями случайной переменной x, иначе говоря, вариации между членами совокупности нужен такой показатель, который обобщал бы колеблемость всех вариант. Для этого надо сравнивать варианты илу друг с другом, или с какой-то одной постоянной величиной. В качестве последней лучше всего взять среднюю арифметическую.

Варианса и среднее квадратическое отклонение. Более совершенными показателями, характеризующими вариацию, являются средний квадрат отклонений вариант от средней арифметической, иначе называемый вариансой,** и среднее квадратическое отклонение, или, иначе, стандартное отклонение. Вариансу обозначают σ2 (греческая буква сигма) или s 2 (латинская буква эс), а среднее квадратическое отклонение−σ.

Словами это можно формулировать так: варианса − это сумма отклонений отдельных значений вариант от средней арифметической, деленная на общее количество вариант, а среднее квадратическое отклонение − корень

квадратный из этого частного. Хотя после извлечения корня квадратного получаются значения со знаками плюс и минус, обычно берут только положительное значение.

Степени свободы. Величина n − 1 получила особое название− число степеней свободы (точнее, число степеней свободы вариации). Мы будем обозначать ее буквами df. Так как во многих разделах статистики приходится пользоваться числом степеней свободы, то следует объяснить его значение.

Существуют различные способы вычисления статистических показателей:

а) прямой через значения вариант (без ВР, при малом n);

б) прямой через значения вариант для ВР;

в) непрямой способ (способ условной средней)

Из всего сказанного видно, что для определения статистических показателей требуется довольно большая вычислительная работа, но объем ее может быть сокращен правильным выбором метода, наиболее подходящего для обработки данного материала, и применением имеющихся технических средств для вычислений ЭВМ, лучше всего пользоваться прямым способом вычислений, так как он дает наиболее точные результаты. Непрямому же способу в силу искусственной разбивки материала на классы всегда сопутствует известная неточность.

Средняя геометрическая. Средняя арифметическая − наиболее часто применяемый статистический показатель, в том числе в биологии. Однако в некоторых случаях (например, при изучении темпов роста организмов или роста целых популяций приходится пользоваться другой средней величиной − средней геометрической.

Формула для ее вычисления следующая:

$$X_g = \sqrt{x1 \cdot x2 \cdot x_n} = \sqrt{\Pi x_i}$$

Очевидно, что при ее определении надо исключать варианты выражающиеся нулем или отрицательным числом.

На практике вычисление средней геометрической производится с помощью логарифмов по следующей рабочей формуле

$$\text{Log } X_g = 1/n (\log x_1 + \log x_2 + \log x_n)$$

т. е. логарифм средней геометрической равен арифметической средней суммы логарифмов отдельных значений x. По значению log X затем определяется величина x .

Лекция 5 СТАТИСТИЧЕСКАЯ ГИПОТЕЗА. ВЫБОРОЧНЫЙ МЕТОД

5.1 Проблема достоверности в статистике. Приемы и методы, изложенные в предыдущих лекциях дают возможность исчерпывающе охарактеризовать биологические совокупности. Каждая совокупность может

быть представлена в виде ряда распределения. Для ряда распределения можно определить статистические показатели, указывающие на наиболее типичный уровень развития изучаемого в совокупности признака и на степень вариации отдельных единиц совокупности вокруг этого уровня.

Большинство из них − именованные величины (средняя арифметическая, мода, медиана, среднее квадратическое отклонение), некоторые выражаются в процентах (коэффициент вариации) или, наконец, являются неименованными числами (варианса, коэффициент асимметрии). Но так как все они − статистические величины, то есть основаны на изучении массовых явлений, возникает очень важный теоретически и практически вопрос о том, насколько они достоверны.

Проблема достоверности занимает видное место в статистической теории. Выборочные и генеральные совокупности. Напомним, что генеральная совокупность − это вся подлежащая изучению совокупность данных объектов. В пределе она рассматривается как состоящая из бесконечно большого количества отдельных единиц. Та часть объектов, которая подвергается исследованию, называется выборочной совокупность или просто выборкой.

Оба типа совокупностей в общем характеризуются одинаковыми закономерностями случайной вариации. Для их характеристик могут быть вычислены статистические показатели: средняя арифметическая и среднее квадратическое отклонение. Среднюю арифметическую мы обозначали ранее символом x. Условимся теперь что x обозначает среднюю арифметическую выборочной совокупности. Среднюю арифметическую генеральной совокупности будем обозначать μ. Каково же соотношение между x и μ?

Допустим, что для совокупности, состоящей из 168 коров симментальской породы, была получена средняя арифметическая глубины груди 73,8 см. 168 коров представляют собой выборку из генеральной совокупности, охватывающей популяцию всех коров симментальской породы. Если бы мы взяли ряд выбора из популяции симментальской породы, то обнаружилось бы, что x этих выборок будут различными. Одни из x будут несколько больше чем 73,8 см, другие − меньше.

Очень важно, что распределение выборочных средних при достаточном их количестве близко к нормальному, поэтому к нему относятся указанные в предыдущей лекции закономерности. Оказывается, что отдельные значения средних арифметических выборок (x) варьируют вокруг средней арифметической генеральной совокупности. Вариация же выборочных средних вокруг μ может быть измерена своим средним квадратическим отклонением, своей сигмой. Эта сигма получила название сред¬ней ошибки или средней квадратической ошибки. Иногда ее называют также стандартной ошибкой. Именно она указывает на степень близости x и μ.

Вопрос 2. Формула для средней ошибки. Средняя ошибка для x может быть вычислена по формуле

$$m_x = \delta/\sqrt{n}$$

В знаменателе формулы под корнем n − объем выборочной совокупности. Это значит, что величина средней ошибки обратно

пропорциональна численности выборочной совокупности.

В примере с глубиной груди у симментальских коров n = 168 и δ == 2,45. Отсюда средняя ошибка для средней арифметической глубины груди изученных 168 симментальских коров

$\delta = 2{,}45/168 = 0{,}17$

5.2 Средняя ошибка − ошибка выборочности Термин «ошибка» часто вводит в заблуждение начинающих, которые предполагают, что она является результатом недостаточной аккуратности в работе. Это не так. Средняя ошибка − это статистическая ошибка. Она не имеет ничего общего с ошибкой точности. Само собою разумеется, что все измерения (веса и промеров рыб, удоев коров и жирности их молока, настригов шерсти овец и ее длины) надо делать точно и добросовестно. Но статистические показатели для выборочной совокупности всегда имеют так называемые ошибки выборочности (их также называют ошибками репрезентативности), которые представляют собой среднюю величину расхождения между средними значениями изучаемых признаков в выборках и генеральной совокупности. Так как

$m_x = \delta/\sqrt{n}$

то, очевидно, что размер определяемой средней ошибки зависит от сигмы выборочной популяции и от ее объема. Чем лучше взята выборка и, чем больше ее размеры, тем меньше и средняя ошибка, тем меньше расхождение между значениями признаков в выборочных и генеральной совокупностях.

Биолог почти всегда имеет дело с выборками − и при проведении опытов с животными или растениями, и при изучении материала, взятого из природы, генеральные же совокупности остаются неизвестными. Поэтому он должен постоянно помнить о том риске, который сопутствует его выводам. Часто эти выводы основываются на изучении небольшого материала, поэтому полученные в опытах или наблюдениях статистические показатели могут иметь значительные статистические ошибки. Легко видеть, что в силу колеблемости выборочных средних вокруг средней генеральной совокупности один какой-либо опыт может дать результат, отклоняющийся от истинного на 2 или даже 3 ошибки. Но при значительном количестве опытов их результаты будут группироваться близко к центру распределения генеральной совокупности, т. е. к μ , что дает возможность уверенно сделать правильный вывод.

Некоторая погрешность органически присуща результатам всякого наблюдения, проведенного на основе выборки. Эту погрешность и измеряет средняя ошибка, которая поэтому и называется ошибкой выборочности (или, иначе, ошибкой репрезентативности). Вместе с тем совершенно необходимо, чтобы выборочная совокупность достаточно хорошо отображала генеральную совокупность, иначе суждение о генеральной совокупности по выборке будет неправильным, несмотря на правильность статистических вычислений. Добиться правильного отображения генеральной совокупности можно при одном непременном условии − отборе вариант для выборки на основе случайности. Чем в большей степени этот отбор будет случайным, тем более правильными будут выводы, делаемые на основе выборочной совокупности. Именно тогда можно полагаться на результаты выборочного наблюдения.

Наиболее простой способ получения случайных выборок −отбирать

экземпляры с помощью таблицы случайных чисел. На принципе случайности основываются различные схемы отбора вариант для выборки: случайная бесповторная выборка, когда взятые для выборки варианты уже не возвращаются обратно в генеральную совокупность, случайная повторная выборка с возвратом взятых для выборки вариант обратно в генеральную совокупность и т. д. Все они подробно рассматриваются в специальных пособиях.

5.3 Закон больших чисел. В связи между статистическими показателями выборочных и генеральных совокупностей выражается так называемый закон больших чисел. В наиболее общем виде этот закон заключается в том, что чем больше число n некоторых случайных величин, тем их средняя арифметическая ближе к средней арифметической генеральной совокупности, тем меньше разница между x и μ. По мере увеличения n вероятность осуществления приближения x к μ становится все большей, стремясь при $n = \infty$ к единице, т. е. к полной достоверности.

В этом заключается теорема одного из основоположников математической статистики русского математика П. Л. Чебышева.

Так как всякое явление, как правило, складывается из массы единичных, случайных явлений, то закон больших чисел выступает как реальный закон объективной действительности. Именно он лежит в основе нормального распределения вариант в вариационном ряду, т. е. распределения значений случайной переменной xi вокруг X, а также в основе распределения выборочных X вокруг μ.

Выборочные средние, для которых вычисляются средние ошибки, являются такими же случайными величинами, как и значения вариант в обычном вариационном ряду. С возрастанием объемов выборок их вариация вокруг генеральной средней становится все меньше. Средняя же арифметическая из всех выборочных средних должна быть равна средней арифметической генеральной совокупности, т. е. μ .

Таким образом, основное содержание закона больших чисел состоит в том, что при увеличении n отдельных выборок происходит взаимное погашение индивидуальных отклонений от некоторого уровня, характерного для всей совокупности в целом. Именно тогда проявляется закономерность, лежащая в основе биологического процесса. Закон больших чисел − одно из выражений диалектической связи между случайностью и необходимостью.

Распределение X малых выборок. Когда выборки являются достаточно большими по объему, распределение их средних арифметических является нормальным. Однако если выборки малы ($n < 30$), то возникает большое сомнение в возможности суждения по таким выборкам о генеральной совокупности. В значение t может вкрасться значительная неточность.

В биологических исследованиях нередко приходится встречаться с выборочными совокупностями, состоящими из очень ограниченного количества вариант или наблюдений.

Возникает вопрос о том, каковы в этих случаях закономерности распределения выборочных средних арифметических. Ответ на него практически дал английский математик Госсет, который писал под

псевдонимом Стьюдент. Поэтому изученное им распределение вероятностей получило название t-распределения по Стьюденту.

Теоретическое обоснование закона распределения, открытого Стьюдентом, было дано Фишером. Существенно то, что оно может быть использовано и при очень малых количествах вариант.

Критерий t по Стьюденту − Фишеру представляет собой следующее:

$t = X - \mu / m_x$

Оказалось, что распределение значений t отличается от нормального, при этом тем сильнее, чем меньше n. Поэтому и вероятности нахождения выборочных средних в пределах определенных значений n значительно снижаются по сравнению с нормальным распределением. В практической работе надо исходить из определенных уровней значимости, поэтому были составлены рабочие таблицы, по которым можно определять минимальное значение, обязательно требующееся для данной вероятности (табл. III, Рокицкий).

5.4 Определение необходимого объема выборочной совокупности. В практике биологических исследований часто возникает вопрос о том, сколько животных (или растений) данного вида надо взять, чтобы получить достаточно правильное представление о популяции вида (по изучаемому признаку). Вообще говоря, следует стремиться к большему числу наблюдений, однако очевидно, что численность выборки не может возрастать бесконечно. Она должна иметь какие-то рациональные границы, которые будут зависеть прежде всего от желаемой точности наблюдения, т. е. допустимого расхождения между средней арифметической (по данному признаку) выборки и средней арифметической генеральной совокупности, а также от заданной вероятности и от степени однородности популяции. Желаемая точность (обозначим ее Δ)− это возможное при принятой вероятности отклонение X от μ, т. е.

$\Delta = tm$.

А так как $m=\delta/n$. то $\Delta = t\,\delta/n$. Отсюда $n=t\,\delta / \Delta$

Значение t определяется ожидаемой вероятностью результата выборочного обследования. При p = 0,997 t должно быть равно 3. При p = 0,95 можно ограничиться t= 2. Величина Δ берется заранее. Так, например, изучая вес зайцев, можно принять, что желаемая точность должна быть в пределах 0,2 кг, т. е. Δ == 0,2 кг.

Несколько труднее решить вопрос о величине среднего квадратического отклонения изучаемой популяции вида, заранее неизвестной. В качестве ее приблизительной оценки можно взять сигму по данным проводившихся ранее исследований или попытаться вычислить ее по максимальным и минимальным значениям изучаемого признака, имея в виду, что вариационный размах должен охватывать примерно шесть средних квадратических отклонений.

ЛЕКЦИЯ 6 СТАТИСТИЧЕСКАЯ ГИПОТЕЗА.
РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРОЧНЫХ ПОКАЗАТЕЛЕЙ

6.1 Оценка достоверности статистических показателей с помощью средней ошибки. Оценка достоверности Х. Роль средней, или статистической, ошибки в статистическом анализе очень велика. С одной стороны, как было показано выше, она позволяет определить границы для показателей генеральной совокупности, например для у, а с другой стороны, дает возможность оценить степень достоверности самих статистических показателей, в частности средней арифметической данной выборочной совокупности.

Что же следует понимать под достоверностью средней арифметической? Фактическая средняя арифметическая всегда является выборочной. Поэтому для суждения о ее достоверности надо сравнить ее со средней арифметической генеральной совокупности. Мерилом достоверности явится нормированное отклонение, для вычисления которого можно использовать приведенную выше формулу.

Возникает вопрос о том, откуда же взять величину р.? Возможны два случая. В первом у. представляет собой определенную, отличающуюся от нуля, величину, значение которой можно примерно предположить по другим данным. Допустим, что изучали жирность молока 10 коров. Были получены следующие показатели; x= 3,7%; σ =0,28%; m=0,09%. Если при этом ранее изучали жирность молока в других выборках и получали различные значения выборочных средних, то можно вычислить среднюю из этих средних. Допустим, что она оказалась равна 4,0%. Можно принять ее за мю. Тогда t=3,7-4,0/0,009=3,3

При малом n (= 8) следует проверить достоверность по табл. II.(Рокицкий) Вероятность достоверности (p= 0,987) вполне достаточная.

В общем можно сказать, что x, вычисленные для большинства биологических показателей даже на сравнительно малых по размерам выборочных совокупностях, чаще всего будут достаточно достоверными, если только ряд не слишком растянут. Однако может получиться иначе, если приходится оперировать экспериментальными данными, в которых фигурируют какие-либо условные или относительные величины, часть последних может иметь и отрицательный знак. Тогда установление достоверности x совершенно необходимо.

6.1 Нулевая гипотеза. Метод средней ошибки позволяет сравнивать между собой любые две группы животных или растений, например: две выборочные совокупности, взятые из природной, неизученной популяции; выборку из какой-то уже известной группы и группу, из которой эта выборка взята; опытную и контрольную группы при постановке опытов − и установить, насколько достоверны различия между их статистическими показателями (средними арифметическими, вариансами и др.).

Общие принципы сравнения основываются на анализе так называемой нулевой гипотезы. Согласно этой гипотезе, первоначально принимается, что между данными показателями (или группами, на основе которых они

получены) достоверного различия нет, т. е. что обе группы вместе составляют один и тот же однородный материал, одну совокупность. Статистический анализ должен привести или к отклонению нулевой гипотезы, если доказана достоверность полученных различий, или к ее сохранению, если достоверность различий не доказана, т. е. различия признаны случайными. Но так как все статистические показатели и различия между ними характеризуются определенными уровнями значимости, то отбрасывание нулевой гипотезы должно быть связано с принятием определенного уровня значимости. Так, если признан необходимым уровень значимости 0,01 и если вероятность достоверности данного статистического показателя или разницы между показателями не удовлетворяет этому условию, т. е. она ниже 0,99 (например, 0,97, 0,91, 0,88), то нет оснований для отбрасывания нулевой гипотезы. Ее надо считать правильной по крайней мере до тех пор, пока новые данные не дадут возможности ее опровергнуть, доказав, что существующие различия не являются чисто случайными.

Конечно, и в том случае, когда нулевая гипотеза считается опровергнутой, какой-то шанс, что она в действительности верна, остается. При уровне значимости 0,01 этот шанс составляет 1 на 100, т. е. в 1 % случаев отбрасывание нулевой гипотезы было ошибкой. Если достигнут уровень значимости не 0,01, а 0,001, то уверенность в том, что нулевая гипотеза действительно отвергнута правильно, резко возрастает (лишь 1шанс на 1000 случаев, что она все же верна). При P = 0,05 уверенность правильности вывода составляет лишь 95 случаев из 100, а в 5 возможен неправильный вывод.

Таким образом, если полученные данные характеризуются уровнем значимости P<0,05, то нет оснований отклонять нулевую гипотезу. Если P > 0,05 – нулевая гипотеза опровергнута.

Но значительно неопределеннее положение вещей, если результаты анализа или сравнения удовлетворяют уровню значимости 0,05, но не удовлетворяют уровню значимости 0,01. Надежное суждение оказывается невозможным. Очевидно, что в таких случаях должны быть проведены дополнительные опыты, чтобы решить, следует ли отбрасывать нулевую гипотезу. Вообще надо иметь в виду, что сохранение нулевой гипотезы еще не означает ее правильности. Может оказаться все же, что она неправильна. Сохранение же нулевой гипотезы оставляет вопрос открытым.

Приведенная выше оценка достоверности средней арифметической выборочной совокупности также являлась проверкой нулевой гипотезы. Согласно нулевой гипотезе, X=0. Надо было доказать, что X достоверно отличается от нуля. При достаточном доказательстве, удовлетворяющем принятому уровню значимости, нулевая гипотеза отбрасывается, т. е. признается достоверность X. Если это не у дается сделать, остается правильной нулевая гипотеза (недостоверность x) впредь до новых опытов.

6.3 Оценка достоверности разницы между средними арифметическими двух выборочных совокупностей. Если была получена разница между средними арифметическими двух генеральных совокупностей, то, очевидно, не может стоять вопрос о статистической ошибке этой разницы. Эта разница всегда достоверна, даже если она и очень мала. Иное дело, если

сравниваются две выборочные совокупности, например: две группы морских свинок, подвергавшихся воздействию химических веществ или физических факторов, две группы коров, сравниваемые по удою и взятые из одной породы, хозяйства н т. д. В этих случаях разница между средними имеет свою статистическую ошибку, с которой ее можно сравнить и установить, достоверна эта разница или нет. Нулевая гипотеза в данном случае будет сводиться к тому, что две изучаемые выборочные совокупности происходят из одной и той же генеральной совокупности и что разница между их средними арифметическими случайна, т. е. лежит в пределах ошибки выборочности.

Чтобы иметь право отвергнуть нулевую гипотезу, надо доказать, что разница между средними арифметическими достоверна, т.е. удовлетворяет требуемому уровню значимости. Для установления достоверности разницы между средними арифметическими надо воспользоваться нормированным отклонением. Нормированное отклонение примет следующую форму:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{(\bar{x}_1 - \bar{x}_2)}}.$$

На самом деле формула для t должна быть несколько сложнее, а именно;

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{(\bar{x}_1 - \bar{x}_2)}}.$$

Но, так как надо исходить из нулевой гипотезы о том, что две выборочные средние арифметические взяты из одной генеральной совокупности, то $\mu_1 = \mu_2$ и правая часть числителя обращается в нуль.

Числителем является разница между средними арифметическими двух групп (знак разницы не имеет значения). Ее можно обозначить сокращенно буквой d. В знаменателе же−средняя ошибка этой разницы, т. е. $m_{x1} - m_{x2}$ или более сокращенно m_d. Тогда

$$t = \frac{d}{s_d}.$$

Существует два способа определения средней ошибки разницы. Первый из них применяется, когда обе сравниваемые группы обладают достаточно большой численностью, большей чем по 30 особей в каждой. Средняя ошибка разницы определяется тогда по формуле

$$s_d = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}.$$

Допустим, что мы хотим сравнить по удою 2 группы коров. В одной группе n_1==50. В другой n_2=40. Средние удой и ошибка для первой группы: $X_1 \pm m_{x1}$== 2100 ± 120 кг; для второй группы: $X_2 \pm m_{x2}$= 2635 ±140 кг. Разница между средними удоями 2 групп

$$d = \bar{x}_2 - \bar{x}_1 = 2635 - 2100 = 535 \text{ кг.}$$

Ошибка разницы

$$s_d = \sqrt{s_{x_1}^2 + s_{x_2}^2} = \sqrt{140^2 + 120^2} = 184 \text{ кг.}$$

Таким образом, $d \pm m_d$ = 535 ± 184 кг, а t= 2,91,

По таблице нормального интеграла вероятности (табл. I, Рокицкий) находим, что в этом случае вероятность достоверности очень велика - 0,9963.

При отсутствии таблицы можно исходить из правила трех сигм: если разница превышает свою ошибку почти в три раза, она достоверна с вероятностью не менее 0,991. Но из сказанного выше очевидно, что в таком высоком значении t нет надобности. Если n > 30, то t=2,58 гарантирует достоверность разницы с вероятностью 0,99.

При сравнении двух групп с малыми, и, особенно с неодинаковыми обьемами, ошибка разницы определяется по формуле:

$$s_d = \sqrt{\frac{\Sigma (x_1 - \bar{x}_1)^2 + \Sigma (x_2 - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)} \left(\frac{n_1 + n_2}{n_1 \cdot n_2} \right)}.$$

Смысл этой формулы заключается в том, что нельзя пользоваться просто готовыми средними ошибками, вычисленными заранее для двух сравниваемых групп, как это было при применении формулы, а нужно сначала сложить суммы квадратов отклонений по обеим группам, т.е. т. е. получить объединенную сумму квадратов отклонений, затем определить вариансу объединенных рядов (путем деления объединенной суммы квадратов на сумму чисел степеней свободы обеих групп) и, наконец, после умножения на $n_1 + n_2$ / n_1 x n_2 и извлечения квадратного корня получить ошибку разницы.

Для иллюстрации возьмем следующий дример. На двух группах крыс был поставлен опыт по сравнению влияния разных рационов на рост. Крысы 1 группы (12 шт.) получали рацион с высоким содержанием белка, крысы второй (7) – с низким. Привесы за 56 дней опыта для каждой крысы составляли (в г): первая группа − 134, 146, 104, 119, 124, 161, 107, 83, 113, 129, 97. 123;

вторая группа−70, 118, 101, 85, 107, 132,94.

После обработки данных с помощью одной из формул для сумм квадратов получим: d= $X_1 - X_2$ = 19 г.; $\sum(x - X_1)^2$ =5302, $\sum(x - X_2)^2$ =2575, тогда общая сумма квадратов равна 7584, а степени свободы df=17. Применив указанные выше формулы получим t=1,89. По табл. III (Рокицкий) находим, что (при df = 17 и уровне значимости 0,05) t должно быть не менее 2,11 , полученное значение t ниже табличного. Для уточнения вероятности достоверной разницы воспользуемся табл. II. Из нее видно, что t = 1,89 соответствует вероятности 0,92, т,е. уровень значимости 0,08. Т.о. можно считать, что разные рационы не привели к разделению популяции крыс по

привесам на две достоверно отличающиеся друг от друга популяции, иначе говорят нулевая гипотеза не может быть отвергнута. Конечно, опытные группы были слишком малы. Возможно, что при их увеличений будет получена более достоверная разница между группами крыс, находившимися на разных рационах кормления.

Лекция 7. ОСНОВЫ ДИСПЕРСИОННОГО АНАЛИЗА

7.1 Сущность и метод дисперсионного анализа. Ранее были рассмотрены методы оценки различия двух выборок путем сравнения их средних μ_1 и μ_2 и стандартных отклонений. В исследованиях часто приходится иметь дело не с двумя, а с большим числом выборок. Обычно эти выборки относятся к различным совокупностям. Например, это могут быть группы растений, получивших разные удобрения или уход, когда в опыте ставится цель статистически оценить эффект мероприятия. В начале 1950-х годов Р. А. Фишер разработал критерий и метод для такой оценки. Это привело к значительному последующему развитию теории планирования опыта и статистической оценки его эффекта.

Статистический смысл задачи по оценке эффекта мероприятия в многогрупповом опыте состоит в проверке значимости различия в групповых средних оцениваемого на основе сравнения дисперсий.

Для раскрытия сущности метода оценки эффекта мероприятия, т. е. дисперсионного анализа, рассмотрим сначала анализ нескольких выборок, взятых из общей совокупности. Такой опыт называют условным экспериментом.

Дж. У. Снедекор (1961) произвел 4 выборки (a=4) из общей совокупности данных по привесу 511 животных. Каждая из групп включала n=5 наблюдений (повторений). Средняя для совокупности μ=30, а дисперсия σ^2=100. Результаты опыта приведены в табл. 1.

Таблица 1. Привесы (в фунтах) 4 групп по 5 животных в группе.

Группа	Привес X	Сумма ΣX	Сред-ние μ	ΣX^2	$\frac{\Sigma(X)^2}{n}$	Σx^2
1	40,, 24, 46, 20, 35	165	33	5917	5445	472
2	29, 27, 20, 39, 45	160	32	5516	5120	396
3	11, 31, 17, 37, 39	135	27	4261	3645	616
4	17, 21, 28, 33, 21	120	24	3044	2880	164
По опыту в целом		580	29	18738	16820	1918

Данные таблицы позволяют получить три оценки дисперсии в совокупности σ^2==100. Первая оценка получается на основе всех 20 наблюдений.

$$\sigma = \frac{\sum x^2}{N-1} = \frac{1918}{19} = 100.9, (N = a \cdot n)$$

Вторая оценка получается из сумм квадратов внутри четырех групп. Она отражает варьирование «отдельных групп».

$$\sigma_1 = \frac{\sum x_1^2 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2}{a \cdot n - n} = \frac{472 + 396 + 616 + 164}{20 - 4} = 103$$

Групповые средние приводят к третьей оценке дисперсии совокупностей.

Средний квадрат средних будет равен:

$$\frac{(\mu_1-\mu)^2+(\mu_2-\mu)^2+(\mu_3-\mu)^2+(\mu_4-\mu)^2}{n-1}=\frac{(33-29)^2+(32-29)^2+(27-29)^2+(24-29)^2}{4-1}=18$$

Число 18 является оценкой $\sigma^2/5$, т. е. оценкой 20.

Каждая средняя представляет 5 наблюдений. Следовательно, третья оценка σ^2 будет равна $\sigma^2=18\cdot5=90$. Она основана на 4 групповых средних при n–11=4–1=3 степенях свободы. Сумма квадратов всех групповых средних составит 90*3=270.

Результаты произведенного подразделения общего варьирования на части и его анализ называют дисперсионным анализом (табл. 2).

Таблица 2. Дисперсионный анализ данных о привесе животных

Источник варьирования	Число степеней свободы	Сумма квадратов	Средний квадрат
Объекты отдельных групп	16	1648	103
Групповые средние	3	270	90
Итого	19	1918	100,9

1 Сумма всех наблюдений:

2 ΣX =40+ 24+... +21=580.

3 Общая сумма квадратов:

$$\sum x^2=\sum X^2-\frac{\left(\sum X\right)^2}{N}=40^2+24^2+\ldots+21^2-\frac{580^2}{20}=1918$$

4 Сумма квадратов для групповых средних:

$$\frac{\sum\left(\sum X\right)^2}{n}-\frac{\left(\sum X\right)^2}{a\cdot n}=\frac{165^2+160^2+\ldots+120^2}{5}-\frac{580^2}{20}=17090-16820=270$$

Сравнение среднего квадрата групповых средних (90) и среднего квадрата для объектов внутри отдельных групп (103) показывает незначительное их расхождение.

Прежде чем делать окончательные выводы, приведем схему расчетов и таблицу анализа в общепринятом виде.

Результат пунктов 2, 3 вносят в таблицу и на их основе получают данные для объектов (табл. 3).

Таблица 3. Дисперсионный анализ данных р привесе животных (общепринятая форма)

Источник варьирования	ν	Σx^2	σ
Общее	19	1918	–
Групповые средние (факториальное)	3	270	90
Объекты отдельных групп (случайное)	16	1648	103

7.2 Дисперсионный анализ случайных выборок из двух или большего числа совокупностей. В большинстве приложений дисперсионного анализа

изучаемые варианты опыта (например, данные дозы удобрения) влияют на средние. Группы становятся выборками из различных совокупностей. Считается, что эти совокупности имеют различные средние μ, но общую дисперсию, не зависимую от вариантов опыта. При дисперсионном анализе средний квадрат для объектов оценивает σ^2, как ранее было показано, но средний квадрат групповых средних оказывается преувеличенным в связи с различиями между μ. Табл. 4 и 5 представляют данные такого эксперимента.

Таблица 4. Высота тополевых саженцев, полученных из черенков особей с разными потомственными данными (от высоты каждого саженца отнято 50 см)

Группа	Высота, см						Сумма	Средняя
1	64	72	68	77	56	95	432	72
2	78	91	97	82	85	77	510	85
3	75	93	78	71	63	76	456	76
4	55	66	51	64	70	66	372	62

Вычисления:

1 $\sum X = 64^2 + 78^2 + \ldots + 66^2 = 1770$

2 $\sum x^2 = 64^2 + 78^2 + \ldots + 66^2 - \frac{1770^2}{24} = 3586.5$

3 Для средних:

$$\frac{432^2 + 510^2 + \ldots + 372^2}{6} - \frac{1770^2}{24} = 1636.5$$

Таблица 5. Дисперсионный анализ данных о высоте саженцев

Источник варьирования	Число степеней свободы	Сумма квадратов	Средний квадрат
Общее	23	3586,5	
Между группами (факториальное)	3	1636,5	545,5
Варианты (случайное)	20	1950	97,5

7.3 Критерий *F*–отношение дисперсий. Заключение о равенстве μ. Полученные данные, приводят к вопросу: обусловливается ли значительное различие между средними квадратами σ_1^2 и σ_2^2 обычным варьированием случайных выборок из одной совокупности или оно настолько велико, что следует его приписать влиянию выборочных средних. Соответствующая такой постановке вопроса нулевая гипотеза такова. H_0: $\mu_1 = \mu_2 = \ldots = \mu_0$ (средние групп одинаковы). Для ответа на подобные вопросы Р. А. Фишер предложил критерий − отношение дисперсий, распределение которого получено на основе случайных выборок из одной общей совокупности. Выше применение критерия *F* рассматривалось для проверки различия в дисперсиях двух малочисленных выборок.

Дж. У. Снедекор знакомит с распределением, полученным на основе 100 выборок по 10 наблюдений в каждой, взятых из уже упоминавшейся общей совокупности по привесу животных. Для каждой выборки по методу, изложенному выше, найдены *F:*

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

Распределение 100 значений *P* (число степеней свободы 9 и 90):

Интервал *F*	0–	0,25–	0,50–	0,75–	1,00–	1,25–
Число случаев	7	16	16	26	11	8
Интервал *F*	1,50–	1,75–	2,00–	2,25–	2,50–	2,75–
Число случаев	5	2	4	2	2	1

Распределение *F* несимметрично. 65 значений *F* меньше 1. Однако среднее значение $\overline{F} = 0.96$, т. е. близко к ожидаемой единице. 5% значений *F* превосходят 2,25, а 1 % выше 2,75.

Такой таблицей распределения *P* можно пользоваться для практических целей. Можно, например, сказать, что при выборках в 10 единиц значение *F*>2,75 может встретиться вследствие случайных причин 1 раз на 100 случаев.

На основе исследований Р. А. Фишера получено теоретическое распределение *F*-критерия для разных уровней значимости и для различного числа степеней свободы.

В таблицах приложений практически всех изданий по статистическим методам приведен 5%-ный уровень в распределении *F*.

При числе степеней свободы *v*=3 и *v*=20 имеем 5%-ный уровень критерия *F*=3,10. Полученное в опыте с саженцами отношение дисперсий $F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{545.5}{97.5} = 5.6 > F_{0.05}$. Оно превышает даже $F_{0.01}$=4,9.

На основании сопоставления *F,* полученного в опыте, с табличными значениями можно сказать, что вследствие случайных причин из одной общей совокупности имеется менее одной возможности из 100 получить выборку, дающую значение *F* больше, чем наблюденное. Очевидно, что данные анализируемой выборки принадлежат к совокупности с различными μ. Следовательно, должен быть дан положительный ответ на поставленный выше вопрос о влиянии материнских наследственных качеств на рост нового поколения. Нулевая гипотеза H_0: $\mu_1 = \mu_2 = ... = \mu_0$ отвергается.

Такой вывод получен на основе установленного значимо более высокого варьирования между групповыми средними, измеряемого σ_1^2 по сравнению с варьированием высот растений внутри групп, измеряемым σ_2^2.

7.4 Дисперсионный анализ с классификацией по двум признакам. В рассмотренном выше примере с высотой саженцев была использована классификация только по одному признаку. Дисперсионный анализ применим и при классификации по нескольким признакам. Ниже рассмотрим пример группировки по двум признакам (факторам), значимость которых проверяют. Имеем следующие результаты наблюдений *X* относительно влияния удобрений (B_1 и B_2) на почвах с разным качественным составом (A_1 и A_2) (табл. 6).

Таблица 6. Результаты наблюдений *X*

Удобрение	Почва		
	A_1	A_2	

B_1	8, 12 $\mu_{11}=10;\ldots\sum x_{11}^2=8$	1, 3 $\mu_{21}=2;\ldots\sum x_{21}^2=2$	$\mu_{B1}=\frac{24}{4}=6$	$\sum x_B^2=0$
B_2	3, 4, 5 $\mu_{12}=4;\ldots\sum x_{12}^2=2$	6, 8, 10 $\mu_{22}=8;\ldots\sum x_{12}^2=8$	$\mu_{B2}=\frac{36}{6}=6$	
Вся группа	$\mu_{A1}=\frac{32}{5}=6.4$	$\mu_{A2}=\frac{28}{5}=5.6$	$\mu=6;\ldots\sum x^2=108$	
	$\sum x_A^2=1.6$			

Числа 8, 12, 3, 4, 5, 1, 3, 6, 8, 10 –значения результативного признака – *X*.

$\mu_{11}, \mu_{12}, \mu_{21}, \ldots\mu_{22}$ – частные средние в клетках; они получены по формуле:

$$\mu=\frac{\sum X_i}{n}$$

μ_{A1}, μ_{A2} –средние для 1 и 2-й групп почв;

μ_{B1}, μ_{B2}– то же, для соответствующих групп удобрений.

$\sum x_{11}^2 \ldots \sum x_{22}^2$ –суммы квадратов отклонений вариант от средних в клетках.

Проверяемые гипотезы. В опытах, подобных рассматриваемому, интересуют вопросы:

1 Различаются ли значимо по своему эффекту на рост растений почвы A_1 и A_2?
2 Значительно ли различен эффект двух удобрений B_1 и B_2?
3 Влияют ли удобрения на рост растений в одинаковой мере на обоях почвах?

Ответ на первый вопрос содержат средние для двух групп почв μ_{A1}=6,4 и μ_{A2}= 5,6.

Различия этого рода, связанные с неотъемлемыми качественными факторами среды, в литературе о дисперсионном, анализе называют *эффектом среды*.

Ответ на второй вопрос содержится в итогах двух строк $\mu_{B1} = \mu_{B2}$=6,0.

Различия, связанные с процессом производства, в данном случае с удобрением, называют *эффектом обработки*.

Ответ на третий вопрос следует искать в средних по клеткам $\mu_{11}, \mu_{12}, \mu_{21}, \ldots\mu_{22}$. Видно, что удобрение B_1 на почве A_1 привело к средней μ_{11}= 10, тогда как на почве A_2 средняя μ_{21}=2. Удобрение B_2 характеризуется обратным указанному результатом: μ_{12}=4; μ_{22}=8. Ответ на третий вопрос выявляет взаимодействие, факторов *AB*.

В поисках заслуживающего доверия ответа на поставленные 3 вопроса выдвигаются 3 нулевые гипотезы:

гипотеза H_a– средние столбцов не отличаются друг от друга

гипотеза H_b– средние строк не отличаются друг от друга

гипотеза H_{ab}– взаимодействие *ab* отсутствует.

Компоненты общей суммы квадратов.

Общая сумма квадратов:

$$\sum x=\sum\left(8^2+10^2+\ldots+6^2+8^2+10^2\right)-\frac{60^2}{10}=468-360=108\,.$$

Эту сумму квадратов разделяем на компоненты, измеряющие влияние

двух испытываемых факторов, их взаимодействие, а также влияние большого числа случайных факторов, т. е. «компонента ошибки» – меры колебаний вследствие игры случая.

Сумма квадратов, соответствующая каждому из принципов классификации, вычисляется так же, как и при однофакторном комплексе, – как сумма квадратов отклонений каждой групповой средней от общей средней (с учетом веса n_i каждой средней):

для фактора почвы – $\sum x_A^2 = (6.4-6)^2 \cdot 5 + (5.6-6)^2 \cdot 5 = 1.6$

для фактора удобрения – $\sum x_B^2 = (6-6)^2 \cdot 5 + (6-6)^2 \cdot 5 = 0$.

«Компонент ошибки», независимый от двух положенных в основу классификации принципов, представляет собой сумму квадратов внутри всех четырех клеток.

$\sum\sum x^2 = 8+2+2)8 = 20$

Эта сумма квадратов, разделенная на соответствующее число степеней свободы, принимается в качестве меры влияния случайных факторов.

Сумма трех компонентов $\sum x_A^2 + \sum x_B^2 + \sum\sum x^2 = 1.6+0+20 = 21.6$.

Вычитая этот результат из общей $\sum x^2 = 108$, получим остаток равный 86,4. Этот остаток можно определить как «остаточную межгрупповую изменчивость». Он будет измерять взаимодействие *AB*.

Для степеней свободы найденных 4-х компонентов имеем следующие зависимости (*a* − число столбцов, *b* − число строк).

Степень свободы

Между строками *b–1*

Между столбцами *a−1*

Для взаимодействия *(a−1) (b−1)*

Внутри клеток *N–ab*

Для итога *N–1*

Дисперсионный анализ показан в табл. 7.

Данные анализа подтверждают нулевую гипотезу $H_a=0$; $H_b=0$, но не согласуются с нулевой гипотезой $H_{ab}=0$. Эта гипотеза отвергается на *1%*-м уровне значимости, т. е. при вероятности p=0,99.

Таблица 7. Дисперсионный анализ

Источник варьирования	Степень свободы v	Сумма квадратов $\sum x^2$	Средний квадрат (дисперсия) σ	F
Фактор *A* (почвы)	1	1,6	1,6	$0,5 < F_{005} = 6,0$
Фактор *B* (удобрение)	1	0		
Взаимодействие *(AB)*	1	86,4	86,4	$26,2 > F_{005} > F_{001} = 13.4$
Внутри клеток (ошибка)	6	20	3,3	–
Итого	9	108		

Из этих результатов анализа делаем вывод, что 2 вида удобрения B_1 и B_2 тесно взаимодействуют с почвой, т. е. Производят эффект в зависимости от почв. Можно сказать и так почвы A_1и A_2 по-разному реагируют на удобрения.

Лекция 8. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

8.1 Понятие о корреляции. Изложенные в предыдущих главах методы анализа дают возможность изучать вариацию животных по каждому отдельному признаку − весу, промерам, плодовитости и т. д. Однако в ряде случаев важно знать, какова зависимость между вариацией двух или даже нескольких признаков изменяются ли два признака самостоятельно, независимо от друга или, может быть, вариация одного признака в какой степени связана с вариацией другого.

Существуют две категории связей, или зависимостей между признаками: функциональные и корреляционные, или статистические. При функциональных зависимостях каждому значению одной переменной величины соответствует одно вполне определённое значение другой переменной. Такие зависимости наблюдаются в математике и физике. Различные измерительные приборы основаны на функциональных зависимостях. Так, высота ртутного столбика в термометре дает точный и однозначный ответ о температуре воздуха или воды. Между радиусом окружности К и ее длиной С существуй функциональная зависимость по известной из элементарной геометрии формуле $C=2\pi R$. Иначе говоря, каждому значений Х соответствует строго определенное значение У. Точно так же накал нити в электрической лампочке определяется напряжением

Наряду с функциональными существуют статистические связи, при которых численному значению одной переменной соответствует много значений другой переменной. Например, между количеством внесенных на поле удобрений и урожайностью пшеницы существует бесспорная зависимость. Это не значит, что определенному количеству удобрений соответствует строго определенная величина урожая. В формировании урожая на данном участке поля много влияет факторов (состава и структуры почвы, способа внесения удобрений, глубина их заделки, различий в методах посева). Во многих исследованиях требуется изучить несколько признаков в их взаимной связи. Если вести такое исследование по отношению к двум признакам, то можно заметить, что изменчивость одного признака находится в некотором соответствии с изменчивостью другого.

В некоторых случаях такая зависимость проявляется настолько сильно, что при изменении первого признака на определенную величину всегда изменяется и второй признак на определенную величину, поэтому каждому значению первого признака всегда соответствует совершенно определенное, единственное значение второго признака. Такие связи называются функциональными.

Встречаются функциональные связи в физических и математических обобщениях. Площадь треугольника точно определяется его высотой и основанием, длина окружности − радиусом, скорость падения есть функция времени падения и ускорения силы тяжести, скорость протекания определенной химической реакции находится в зависимости от температуры.

Необходимо учесть, что функциональные связи встречаются только в идеальных условиях, когда предполагается, что никаких посторонних влияний

нет.

При изучении живых объектов − диких и культурных растений, животных, микроорганизмов − приходится иметь дело со связями другого рода. Живой организм развивается в связи с условиями его жизни, под действием бесконечно большого числа факторов, которые по-разному определяют развитие разных признаков.

У живых объектов связь между любыми двумя признаками настолько часто и сильно нарушается и модифицируется, что не всегда даже может быть легко обнаружена. У растений, животных и микроорганизмов связь между признаками обычно проявляется особым образом. Каждому определенному значению первого признака соответствует не одно значение второго признака, а целое распределение этих значений при вполне определенных основных показателях этого частного распределения− средней величины и степени разнообразия. Такая связь называется корреляционной связью или просто корреляцией.

Корреляционная связь, например, между весом животных и их длиной выражается в том, что каждому значению длины соответствует определенное распределение веса (а не одно значение веса), и с увеличением длины увеличивается и средний вес животных.

Корреляционная связь не является точной зависимостью одного признака от другого, поэтому она может иметь различную степень − от полной независимости до очень сильной связи. Кроме того, характер связи между разными признаками может быть различен. Поэтому возникла необходимость определять форму, направление и степень корреляционных связей.

По форме корреляция может быть прямолинейной и криволинейной, по направлению − прямой и обратной. Степень корреляции измеряется различными показателями, введенными для установления силы связи между количественными и качественными признаками. Такими показателями являются коэффициент корреляции r, корреляционное отношение η.

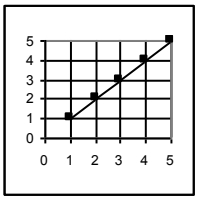
Изобразить корреляционную связь двух признаков можно тремя способами:

- При помощи корреляционного ряда, состоящего из ряда пар значений, из которых одно относится к первому признаку, а другое в этой паре относится ко второму признаку, связанному с первым. На рис. 7.1 показаны схемы корреляционных рядов при пяти степенях корреляционной связи.

5					**1**
4				**1**	
3			**1**		
2		**1**			
1	**1**				
	1	2	3	4	5

X_1	1	2	3	4	5
X_2	1	2	3	4	5

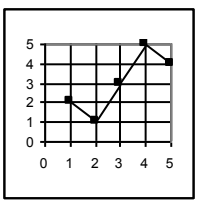
Прямая полная связь; r=+1,0



5				**1**	
4					**1**
3			**1**		
2	**1**				
1		**1**			
	1	2	3	4	5

X_1	1	2	3	4	5
X_2	2	1	3	5	4

Прямая частичная связь; r=+0,8



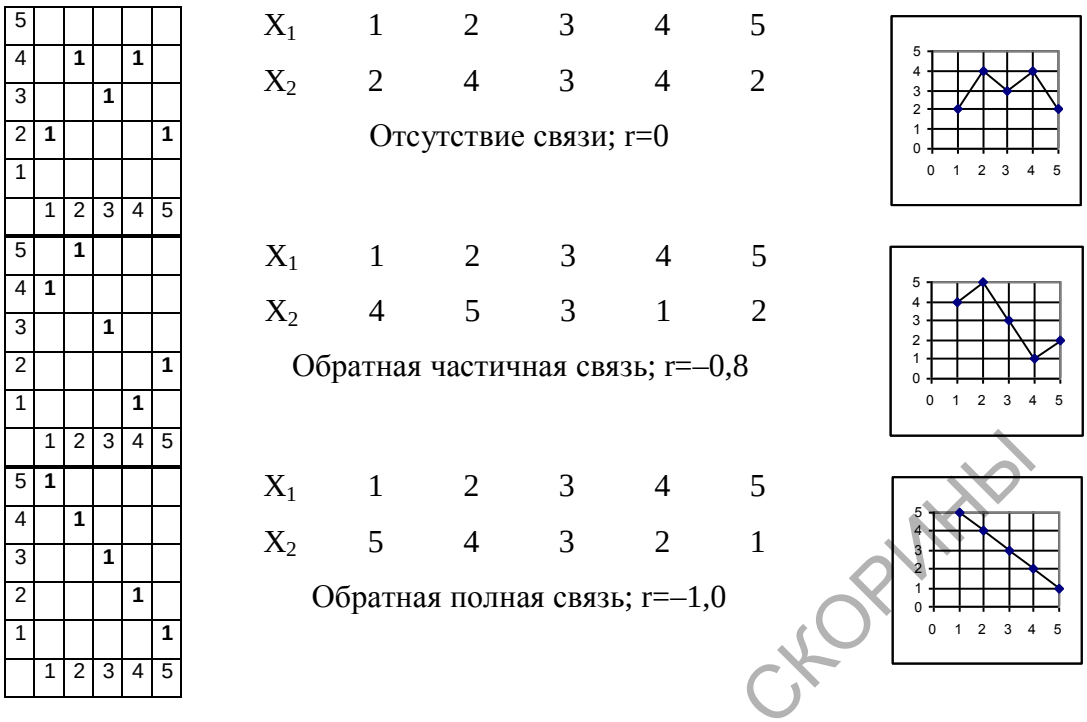


Рис. 1 Схема прямолинейных корреляционных связей

- При помощи корреляционной решетки, в которой каждой особи соответствует определенная клетка. На рис. 7.1 показана схема корреляционных решеток для пяти степеней корреляционной связи между двумя признаками Значения первого признака нанесены по оси абсцисс, значения второго − по оси ординат.
- При помощи линии регрессии, абсциссы которой пропорциональны значениям первого признака, а ординаты − значениям второго признака, корреляционно связанного с первым. На рис. 7.1 показаны схемы линий регрессии для пяти степеней корреляционной связи между двумя признаками.

8.2 Коэффициент корреляции. Коэффициент корреляции измеряет степень и определяет направление прямолинейных связей.

Прямолинейная связь между признаками − это такая связь, при которой равномерным изменениям первого признака соответствуют равномерные (в среднем) изменения второго признака при незначительных и беспорядочных отклонениях от этой равномерности. Например, при увеличении длины тела на каждый сантиметр ширина увеличивается в среднем на 0,7 см

При графическом изображении прямолинейных связей (см рис. 7.1) (если по оси абсцисс отложить значения первого признака, по оси ординат − второго и полученные точки соединить) получается прямая или такая кривая, среднее течение которой проходит по прямой

При изображении прямолинейных корреляционных связей в форме корреляционных решеток (см рис. 1) частоты внутри располагаются в форме эллипса Большая ось этого эллипса проходит или по диагонали от угла наименьших значений (при положительной корреляционной связи), или по диагонали от угла, где сходятся наименьшие значения одного признака и наибольшие значения другого, к противоположному углу (при отрицательной корреляционной связи).

При измерении степени связи между разными признаками приходится сравнивать величины, выраженные в разных единицах измерения. Например, при измерении связи между весом животного и его длиной надо сопоставить килограммы веса с сантиметрами длины. В других случаях изменения объема сопоставляются с изменениями возраста, изменения веса руна в килограммах с изменениями содержания в нем жиропота в процентах, длина ног в сантиметрах со скоростью бега в минутах и т. д.

Проводить такие сравнения оказалось возможным путем использования нормированного отклонения, вычисляемого по формуле:

$$\bar{x}_i = \frac{X_i - \mu}{\sigma}.$$

Нормированное отклонение служит универсальной и неименованной мерой развития признаков. Эти свойства нормированного отклонения и позволили сконструировать основной показатель корреляционной связи − коэффициент корреляции.

Основная формула, которая вскрывает сущность этого показателя, имеет совсем простую структуру:

$$r = \frac{\sum \bar{x}_1 \cdot \bar{x}_2}{\nu}$$

где r – коэффициент корреляции;

$\bar{x}_1 \cdot \bar{x}_2$ – нормированные отклонения дат по первому и второму признаку;

ν – число степеней свободы, равное в данном случае числу сравниваемых пар без одной.

Сумма произведений нормированных отклонений, входящая в формулу для коэффициента корреляции, обладает следующими тремя особыми свойствами

Если оба признака изменяются параллельно, то сумма произведений их нормированных отклонений дает положительную величину. Если при увеличении одного признака другой уменьшается, то приходится умножать положительные числа на отрицательные и вся сумма произведений нормированных отклонений дает отрицательную величину. Поэтому коэффициент корреляции может определять направление связи: при прямых связях он положителен, а при обратных связях отрицателен.

При полных связях, когда изменения обоих признаков строго соответствуют друг другу и корреляционная связь превращается в функциональную, сумма произведений нормированных отклонений становится равной числу степеней свободы:

$$\sum \bar{x}_1 \cdot \bar{x}_2 = \nu = n - 1$$

Поэтому максимальное значение коэффициента корреляции равно 1; для положительных, или прямых связей:

$$r_{\max} = \frac{\sum \bar{x}_1 \cdot \bar{x}_2}{\nu} = \frac{+\nu}{\nu} = +1.0$$

для отрицательных, или обратных связей:

$$r_{\min} = \frac{\sum \bar{x}_1 \cdot \bar{x}_2}{\nu} = \frac{-\nu}{\nu} = -1.0$$

- При полном отсутствии корреляционной связи между признаками сумма произведений нормированных отклонений равна нулю, и поэтому коэффициент корреляции в этих случаях тоже равен нулю:

$$r_{\min} = \frac{\sum \bar{x}_1 \cdot \bar{x}_2}{\nu} = \frac{0}{\nu} = 0$$

Предельные значения коэффициента корреляции (*r=+1,0; r=0,0; r= –1,0*) на практике встречаются крайне редко.

Пять основных видов прямолинейной корреляционной связи, соответствующие коэффициентам корреляции +1,0; +0,8; 0,0; –0,8 и –1,0, показаны на рис. 7.1.

Основная формула коэффициента корреляции хорошо вскрывает сущность этого показателя, но для работы крайне неудобна, особенно при многочисленных группах. Поэтому разработаны разнообразные рабочие формулы для практических расчетов в разных условиях – для малых и больших групп при малозначных и многозначных вариантах.

Все эти формулы дают одинаковый результат и применение любой из них обусловливается только удобством и простотой необходимых вычислений.

Наиболее приемлемы в биологических работах две формулы, предложенные дли малых групп:

$$r = \frac{\sum X_1 \cdot X_2 - \frac{\sum X_1 \sum X_2}{n}}{\sigma_1 \cdot \sigma_2}$$

где X_1, X_2 − даты первого и второго признаков; N− число сравниваемых пар дат, или число объектов, у которых измерено по два признака;

σ_1, σ_2− стандартные отклонения по первому признаку и по второму признаку.

Применяется коэффициент корреляции в тех случаях, когда необходимо знать направление и силу связи между признаками, причем заранее известно, что эта связь может считаться прямолинейной, или когда требуется выяснить степень именно прямолинейной связи. При этом лучше проводить два этапа исследования: 1) рассмотрение корреляционной решетки; 2) расчет коэффициента корреляции или по этой же решетке, или непосредственно по датам.

Уже самый вид корреляционной решетки позволяет приблизительно установить направление и степень прямолинейных связей, а также характер криволинейных связей. При известном опыте по виду корреляционной решетки можно получить первое представление об особенностях и силе связи между изучаемыми признаками. Облегчает решение этой задачи схема степеней прямолинейной корреляции, показанная в табл. 7.1. В этой схеме приведены стандартные корреляционные распределения 50 особей при различных степенях прямолинейной связи по девяти градациям от *r=+1,0 до r=-1,0.*

Схемой степеней прямолинейной корреляции можно пользоваться как эталоном для первоначального ориентировочного отнесения изучаемой связи к одной из условных степеней («сильная», «средняя», «слабая») только по одному виду корреляционной решетки. В некоторых случаях такая грубая

оценка бывает достаточна для выяснения предварительных вопросов исследования.

Таблица 1. Схема степеней прямолинейной корреляции

Прямая корреляция сильная

			2	2
		5	2	2
	5	8	5	
2	5	5		
2	2			

r=+0.75

Прямая корреляция средняя

		1	2	1
	2	4	4	2
1	4	8	4	1
2	4	4	2	
1	2	1		

r=+0.5

Прямая корреляция слабая

	1	1	1	1
1	2	3	5	1
1	3	10	3	1
1	5	3	2	1
1	1	1	1	

r=+0.25

Прямая корреляция полная

				4
			12	
		18		
	12			
4				

r=+1.0

Отсутствие корреляции

	1	2	1	
1	3	4	3	1
2	4	6	4	2
1	3	4	3	1
	1	2	1	

r=+0.0

Обратная корреляция полная

4				
	12			
		18		
			12	
				4

r=-1.0

Обратная корреляция слабая

1	1	1	1	
1	5	3	2	1
1	3	10	3	1
1	2	3	5	1
	1	1	1	1

r=-0.25

Обратная корреляция средняя

1	2	1		
2	4	4	2	
1	4	8	4	1
	2	4	4	2
		1	2	1

r=-0.5

Обратная корреляция сильная

2	2			
2	5	5		
	5	8	5	
		5	5	2
			2	2

r=-0.75

8.3 Ошибка коэффициента корреляции

Как и всякая выборочная величина, коэффициент корреляции имеет свою ошибку репрезентативности, вычисляемую для больших выборок по формуле:

$$s_r = \frac{1-(\bar{r})^2}{\sqrt{n-1}},$$

Где $\bar{r}$ −коэффициент корреляции в генеральной совокупности, из которой взята выборка;

n − численность выборки, т. е. число пар значений, по которым вычислялся выборочный коэффициент корреляции.

Поскольку в числителе формулы ошибки выборочного коэффициента корреляции стоит квадрат генерального коэффициента корреляции, то эта формула может применяться лишь в исключительных случаях, когда заранее известна или предполагается степень корреляции в генеральной совокупности.

Пример. Для проверки гипотезы о том, что коэффициент корреляции между детьми и родителями r =+0,5, была сопоставлена плодовитость 226 лисиц и их дочерей в соответствующем возрасте и в сходных условиях. Коэффициент корреляции оказался равным +0,45. Подтверждает или

опровергает этот результат гипотезу?

В данном случае разность между выборочным и генеральным коэффициентами *d=+0,45–(+0,50)=–0,05*, а ее ошибка равна ошибке выборочного коэффициента, так как генеральные величины не имеют ошибок репрезентативности. Для вычисления ошибки коэффициента корреляции имеется возможность применить точную формулу с генеральным коэффициентом в числителе:

$$s_r = \frac{1-0.5^2}{\sqrt{225}} = \frac{0.75}{15} = 0.05$$

Оказалось, что критерий достоверности разности $t_{(r-r)} = \frac{0.05}{0.05} = 1$ не превышает даже первого порога достоверности ($t_1=2,0$ $\beta_1 = 0,95$).

Гипотеза в данном исследовании не опровергнута, так как эмпирический коэффициент корреляции недостоверно отличается от гипотетического.

В большинстве исследований значение коэффициента корреляции в генеральной совокупности неизвестно, поэтому вместо точного значения ошибки коэффициента корреляции берут приближенное значение:

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

Где *r*–выборочное значение коэффициента корреляции,

n–число сравниваемых пар данных или число объектов, у которых измерены два признака.

Ошибка коэффициента корреляции используется для определения: 1) достоверности выборочного коэффициента корреляции; 2) доверительных границ генерального коэффициента корреляции; 3) достоверности разности двух выборочных коэффициентов корреляции; 4) достоверности разности между выборочным и генеральным коэффициентом корреляции.

8.4 Достоверность выборочного коэффициента корреляции

Критерий выборочного коэффициента корреляции определяется по формуле:

$$t_r = \frac{r}{s_r} \geq t_{st}\{\nu = n-2\}$$

где t_{st} – критерий достоверности коэффициента корреляции;

r– выборочный коэффициент корреляции;

n–число коррелированных пар дат;

t_{st}– стандартное значение критерия Стьюдента, находимое по таблице для установленного числа степеней свободы и порога вероятности безошибочных прогнозов.

При $t \geq t_{st}$ коэффициент корреляции достоверен. В этом случае с определенной вероятностью можно считать, что между коррелируемыми признаками имеется связь и в генеральной совокупности такая же по знаку, какая получилась в выборке (прямая или обратная).

При $t < t_{st}$ выборочный коэффициент корреляции недостоверен, что не дает возможности сделать какое-либо заключение о связи признаков в генеральной совокупности. Для выяснения этого вопроса требуется провести

повторные исследования на более многочисленном материале.

Пример. При проверке гипотезы о связи крупноплодности с жирномолочностью был рассчитан коэффициент корреляции между процентом жира в молоке у 50 коров и весом при рождении телят от этих же коров. Получено:

коэффициент корреляции *r=+0,21*;

его ошибка $s_r = \sqrt{\frac{1-0.21^2}{50-2}} = 0.14$;

критерий достоверности:

$t_r = \frac{0.21}{0.14} = 1.5$; ν=48; t_{st}= {2,0-2,7-3,5}.

Выборочный коэффициент оказался явно недостоверным. На основе проведенного исследования нельзя ожидать связи между крупноплодностью и жирномолочностью у всех коров вообще.

Определение достоверности коэффициента корреляции можно значительно упростить, используя свойства особой функции предложенной Фишером:

$$z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

При помощи этой функции можно заранее определить, при каком объеме выборки коэффициент корреляции определенной величины будет достоверен по требуемому порогу вероятности безошибочных прогнозов, по следующей формуле:

$$\hat{n} = \frac{t^2}{z^2} + 3$$

где $\hat{n}$ − количество пар значений, достаточное для достоверности выборочного коэффициента корреляции,

t − критерий Стьюдента для каждого из трех порогов вероятности безошибочных прогнозов (β_1 = 0,95, β_2 = 0,99, β_3 = 0,999), для больших групп: t_1 = 1,96, t_2=2,58, t_3 = 3,30.

z − функция Фишера $z = \frac{1}{2}\ln\frac{1+r}{1-r}$

По этой формуле рассчитано значение z и количество пар значений, достаточное для достоверности выборочного коэффициента корреляции для каждого из трех порогов вероятности безошибочных прогнозов.

В примере в выборке объемом n = 50 получен коэффициент корреляции ***r***= +0,21.

При r = 0,21, рассчитаны три числа: 87−149−242. Это значит, что выборочный коэффициент корреляции, равный r=0,21, может стать достоверным в том случае, если объем выборки (число коррелируемых пар данных) будет: для первого порога вероятности 87, для второго− 149, для третьего − 242. Так как фактический объем выборки n =50 далеко не достигает первого, максимальною порога, то полученный коэффициент корреляции оказался недостоверным, что было найдено и обычным способом.

8.5 Доверительные границы коэффициента корреляции

Доверительные границы генерального значения коэффициента

корреляции находятся общим способом по формуле:

$\bar{r} = r \pm \Delta$,

где $\bar{r}$ и r − генеральное и выборочное значения коэффициента корреляции;

$\Delta = t^* s_r$ − возможная погрешность при определении генерального параметра;

t_{st} − критерий Стьюдента при числе степеней свободы $\nu = n-2$;

s_r − ошибка коэффициента корреляции.

Пример. При разработке способов определения веса устриц определенного вида по их длине было измерено и взвешено 200 экземпляров и определен коэффициент корреляции между весом и длиной $r = +0{,}85$.

Ошибка этого коэффициента

$$s_r = \sqrt{\frac{1-0.85^2}{200-2}} = 0.037$$

Число степеней свободы и критерий Стьюдента

$\nu = n - 2 = 198$, $t_{st} = \{2{,}0 - 2{,}6 - 3{,}3\}$.

Возможная погрешность при прогнозе генерального параметра

$\Delta = t^* s_r = 2{,}0 * 0{,}037 = 0{,}074$.

Доверительные границы:

$r = +0.85 \pm 0.074$[−не более + 0.85+0.074 = 0.92; −не менее 0.85-0.074 = 0.78]

Даже минимальная граница (гарантированный минимум) оказалась достаточно высокой. Это указывает на возможность практического использования вскрытой закономерности путем разработки формулы регрессии для определения веса устриц по их длине с практически достаточной точностью.

Достоверность разности двух коэффициентов корреляции

Достоверность разности коэффициентов корреляции определяется так же, как и достоверность разности средних, по обычной формуле

$$t_d = \frac{d}{s_d} \geq t_{st}\{\nu = n_1 + n_2 - 4\},$$

где t_d −критерий достоверности разности коэффициентов корреляции;

$d = r_1 - r_2$−разность коэффициентов корреляции;

$s_d = \sqrt{s_1^2 \cdot s_2^2}$ −ошибка разности, равная корню квадратному из суммы квадратов ошибок обоих сравниваемых коэффициентов корреляции; $s^2 = \frac{1-r^2}{n-2}$;

t_{st} − стандартные значения критерия Стьюдента;

ν − число степеней свободы для разности коэффициентов корреляции, равное сумме чисел степеней свободы обоих коэффициентов:

$\nu = n_1 - 2 + n_2 - 2 = n_1 + n_2 - 4$.

Пример. При разработке способов определения высоты дерева по его обхвату (на высоте груди измеряющего) получены коэффициенты корреляции между этими признаками для двух пород деревьев:

$n_1 = 200$, $r_1 = 0{,}60$, $s_1^2 = \frac{1-0.6^2}{198} = 0.0032$;

$n_2 = 150$, $r_2 = 0{,}80$, $s_2^2 = \frac{1-0.8^2}{148} = 0.0024$.

Для выяснения возможности применения единой формулы пересчета обхвата на высоту потребовалось выяснить: достоверно ли различие связи высоты с обхватом между двумя изучаемыми породами деревьев. Получены следующие результаты:

$d = 0{,}80 - 0{,}60 = 0{,}20$;

$s_d^2 = 0{,}0032 + 0{,}0024 = 0{,}0056$, $s_d = \sqrt{0{,}0056} = 0{,}075$;

$t_d = \frac{0.200}{0.075} = \underline{\underline{2.7}}$, $\nu = 200 + 150 - 4 = 346$, $t_{st} = \{2{,}0 - 2{,}6 - 3.3\}$.

Оказалось, что сравниваемые породы достаточно достоверно (по второму порогу вероятности) различаются по степени связи между высотой и обхватом дерева. Поэтому для этих пород нельзя пользоваться единой формулой пересчета обхвата на высоту.

Лекция 9. РЕГРЕССИОННЫЙ АНАЛИЗ

9.1 Многообразие методов изучения связи. Известно, что различные зависимости широко распространены как в органической, так и в неорганической природе. Их изучение проводилось уже давно и привело к разработке большого количества методов их математической характеристики. И первым из них являлся разобранный в предыдущей лекции корреляционный метод, или метод корреляций.

Коэффициент корреляции указывает лишь на степень связи в вариации двух переменных величин или, как иногда говорят, на меру тесноты этой связи, но не дает возможности судить о том, как количественно меняется одна величина по мере изменения другой. На этот последний вопрос позволяет ответить другой метод определения связи между варьирующими признаками, носящий название метода регрессии.

В современной статистике, в том числе биологической, коэффициентами корреляции пользуются реже, чем прежде, Метод же регрессии приобретает все большее значение. Анализ взаимоотношения двух изменчивых величин с помощью метода регрессии часто может дать очень ценные результаты, особенно в практическом отношении. В некоторых случаях для освещения различных сторон вопроса надо применять и корреляционный, и регрессионный методы анализа.

При простой корреляции изучается зависимость между изменчивостью двух признаков x и y. С помощью регрессии ставится дополнительно задача установить, как количественно изменяется одна величина при изменении другой на единицу. Так как изменчивых величин две, то регрессия, очевидно, может быть двусторонней:

1. определение изменения y по изменению x
2. определение изменения x по изменению y.

В этом заключается главное отличие метода регрессии от метода корреляции. Регрессия может быть выражена несколькими способами:

- путем построения так называемых эмпирических линий регрессии,
- путем составления уравнений регрессии и построения теоретических линий регрессии,
- с помощью вычисления коэффициента регрессии.

Первые два способа позволяют выразить регрессию графически. Для построения эмпирических линий регрессии можно воспользоваться обычной корреляционной решеткой. Но в ней следует заменить границы классов средними значениями классов.

9.2 Коэффициент прямолинейной регрессии

Прямолинейная корреляция отличается тем, что при этой форме связи каждому из одинаковых изменений первого признака соответствует вполне определенное и тоже одинаковое в среднем изменение другого признака, связанного с первым или зависящего от первого.

Та величина, на которую в среднем изменяется второй признак, при изменении первого на единицу измерения, называется коэффициентом

регрессии. Рассчитывается он по формуле:

$$R_{2/1} = \frac{\sigma_2}{\sigma_1} \cdot r_{12},$$

где $R_{1/2}$ − коэффициент регрессии второго признака по первому;

σ_2 − среднее квадратическое отклонение второго признака, который изменяется в связи с изменением первого;

σ_1 − среднее квадратическое отклонение первого признака, в связи с изменением которого изменяется второй признак;

r_{12} − коэффициент корреляции между первым и вторым признаками.

Ошибка коэффициента регрессии равна ошибке коэффициента корреляции, умноженной на отношение сигм:

$$s_R = \frac{\sigma_2}{\sigma_1} \cdot s_r = \frac{\sigma_2}{\sigma_1} \cdot \sqrt{\frac{1-r^2}{n-2}}.$$

Критерий достоверности коэффициента регрессии равен критерию достоверности коэффициента корреляции:

$$t_R = \frac{R}{s_R} = \frac{\frac{\sigma_2}{\sigma_1} \cdot r_{12}}{\frac{\sigma_2}{\sigma_1} \cdot s_r} = \frac{r}{s_r} = t_r,$$

Пример. Для разработки способа определения веса лошадей без взвешивания по обхвату груди было взвешено 1618 лошадей и у каждой из них измерен обхват груди. Получены следующие показатели: x − обхват груди, n = 1618, μ_x=174 см, σ_x=7,9 см;

y− вес, n=1618, μ_y=424 кг, σ_y = 56,8 кг.

Коэффициент корреляции $r_{x/y}$=*+0,89±0,011*.

Коэффициент регрессии веса по обхвату равен:

$$R_{y/x} = \frac{\sigma_y}{\sigma_x} \cdot r_{y/x} = \frac{56.8}{7.9}(+0.89) = +6.4.$$

Ошибка коэффициента регрессии веса лошадей по обхвату их груди равна:

$$s_R = \frac{\sigma_y}{\sigma_x} \cdot s_r = \frac{56.8}{7.9} \cdot 0.011 = 0.08.$$

Достоверность этого коэффициента регрессии определяется следующим образом:

$$t_R = \frac{6.4}{0.08} = \underline{\underline{80.0}},\ \nu=1618-2=1616,$$

t_{st}={2.0−2.6−3.3}

Возможная максимальная погрешность при прогнозе генерального параметра

$\Delta = t_m$ = 2,0*0,08 = 0,16.

Доверительные границы

$$\overline{R}_{y/x} == +6,4 \pm 0,16 = \{6,24 - 6,56\}.$$

Таким образом, можно ожидать, что при увеличении (или уменьшении) обхвата груди на 1 см вес лошадей увеличится (или уменьшится) в среднем на

R=+6,4 кг при гарантированном минимуме изменения +6,24 кг и возможном максимуме +6,56 кг, если учитывать изменения признаков в обе стороны от их средней величины.

Коэффициент прямолинейной регрессии показывает, на сколько от своей средней отклоняется второй признак, если первый признак от своей средней отклонился на единицу измерения. Это можно выразить следующей формулой:

$(X_2-\mu_2)=R_{2/1}\,(X_1-\mu_1)$

Обозначая X_1 через x, X_2 через y, $R_{1/2}$ через b и произведя необходимые преобразования этого выражения, можно получить рабочую формулу прямолинейной регрессии:

$y=a+bx$

$$\left\{\frac{a=\mu_y-b\mu_x}{b=R_{y/x}}\right\}.$$

По этой формуле, зная значение x (аргумент), можно определить значение y (функция) без непосредственного его измерения: нужно аргумент x помножить на коэффициент регрессии и к полученному произведению прибавить (или отнять) свободный член a.

Для предыдущего примера (определение веса лошадей по обхвату груди) уравнение регрессии может быть выведено следующим образом:

$a = \mu_y - R_{y/x} * \mu_x = 424 - (+6{,}4)\ 174 = -690,$

$b = R_{y/x} = +6{,}4,$

$y = a + bx = -690 + 6{,}4x = 6{,}4x - 690.$

Следовательно, чтобы определить (без взвешивания живой вес лошади по этому способу, надо обхват груди лошади умножить на постоянный коэффициент 6,4 и из полученного произведения вычесть постоянное число –690.

На основе уравнения прямолинейной регрессии можно заранее рассчитать значение функции для каждого значения аргумента.

По обхвату груди можно определить живой вес лошадей.

Если эти цифры нанести на график, по оси абсцисс которого отложить через равные интервалы значения аргумента (обхвата), а по оси ординат − значения функции (веса), то получится номограмма для определения веса лошадей без взвешивания и без вычислений.

Ошибки элементов уравнения прямолинейной регрессии.

В уравнении простой прямолинейной регрессии:

$y_x = a + bx$

возникают три ошибки репрезентативности.

4 Ошибка коэффициента регрессии:

$$s_b = \frac{\sigma_y}{\sigma_x}\cdot\sqrt{\frac{1-r^2}{n-2}} = \frac{\sigma_y}{\sigma_x}\cdot s_r$$

5 Ошибка уравнения регрессии, т. е. ошибка средней величины функции для каждого значения аргумента:

$$m_{y_x} = \sigma_y\cdot\sqrt{\frac{1-r^2}{n-2}}$$

По данным примера

$s_{y_x} = 56.8 \cdot 0.011 = 0.62$.

Следовательно, максимальная погрешность в определении уровня точек линии регрессии при первом пороге вероятности безошибочных прогнозов (β_1 = 0,95, t_1=2,0) будет равна:

$\Delta = t*s = 2*0{,}62 = \pm 1{,}24$ кг.

6 Ошибка индивидуальных определений функции:

$s_y = \sigma_y \sqrt{1 - r^2}$

Для примера:

$s_y = 56.8\sqrt{1 - 0.89^2} = 26.2$.

Следовательно, индивидуальная погрешность в определении веса лошадей по обхвату груди по найденной формуле регрессии, принимая первый порог вероятности безошибочных прогнозов (β_1 = 0,95, t_1 = 2,0), в крайних случаях не будет превышать

$\Delta = 2 * 26 = \pm 52$ кг.

ЛИТЕРАТУРА

Рекомендуемая литература (основная):

1. Лакин Г.Ф. «Биометрия». М. Высшая школа, 1990.
2. Бейли Н. «Математика в биологии и медицине». М., Мир, 1970.
3. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975.
4. Гланц С. Медико-биологическая статистика. М.: Практика, 1998.
5. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. – М.: Финансы и статистика, 1982.
6. Носов В.Н. «Компьютерная биометрика». МГУ, 1990.
7. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. – СПб.: Питер, 2003.
8. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. – М.: МедиаСфера, 2002.
9. Рокицкий П. Ф. Биологическая статистика, 1976(1980).

Рекомендуемая литература (дополнительная):

1. Плохинский Н.А. Биометрия. - М.: МГУ, 1970. – 368 с.
2. Свалов Н.Н. Вариационная статистика. - М.: Лесная промышленность, 1977. – 177 с.
3. Справочник по прикладной статистике. В 2-х т. / Под ред. Э. Лойда, У. Ледермана, Ю.Н. Тюрина. – М.: Финансы и статистика, Т.1: 1989; Т.2: 1990.
4. Глас Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976.
5. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: ИНФРА-М, Финансы и статистка, 1995.
6. Боровиков В.П. Популярное введение в программу STATISTICA.- М.: КомпьютерПресс, 1998.
7. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистические функции MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2003.
8. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel. – К.: МОРИОН, 2000.
9. Макарова Н.В., Трофимец В.Я. «Статистика в Excel». М. Финансы и статистика, 2002.
10. Глотов Н. В., Животовский Л. А.,Хованов Н. В., Хромов-Борисов Н. Н. Биометрия. Л., 1982.
11. Терентьев П. В. Истоки биометрии. Из истории биологии. М., 1971.

Глоссарий по дисциплине

Алгоритм - полностью определенный, конечный набор шагов, операций или процедур, которые приводят к конкретному результату.

Альтернативная вариация – простейший случай качественной вариации, когда совокупность состоит только из двух групп: одной, имеющий данный признак, а другой – его не имеющий.

Аппарат Гальтона – устройство, предназначенное для наглядной демонстрации распределения вариант в виде вариационного ряда, частоты в котором следуют коэффициентам разложения бинома Ньютона.

Апостериорные сравнения - Обычно, получив при проведении дисперсионного анализа статистически значимое значение F-критерия, мы хотели бы узнать, какая из групп вызвала этот эффект, т.е. какие из групп значительно отличаются от других. Конечно, мы могли бы вычислить последовательность обычных t-критериев для сравнения всех возможных пар средних. Однако такая процедура будет основана на случайности. Получаемые уровни вероятности будут завышать значимость различия между средними. Например, предположим, что мы получили 20 выборок по 10 случайно выбранных чисел каждая, а затем вычислили 20 средних. После этого возьмем группу (выборку) с наибольшим средним и сравнить ее с выборкой с наименьшим средним. t-критерий для независимых выборок проверяет, являются ли два средних значимо отличающимися друг от друга, в предположении, что рассматриваются всего две выборки. Метод апостериорных сравнений, наоборот, предполагает наличие более чем двух выборок. Этот метод используется для проверки гипотез и разведочного анализа.

Асимметрия или **коэффициент асимметрии** - (термин был впервые введен Пирсоном, 1895) является мерой несимметричности распределения. Если этот коэффициент отчетливо отличается от 0, распределение является асимметричным. Плотность нормального распределения симметрична относительно среднего.

Биноминальное распределение – распределение, при котором вероятности появления отдельных значений x_i выражаются величинами, соответствующие коэффициентам разложения бинома Ньютона.

Варианта – значение или мера признака для единицы совокупности.

Варианса (средний квадрат отклонений вариант от средней арифметической) σ^2 – это сумма квадратов отклонений отдельных значений данной переменной от средней арифметической, деленная на число вариант.

Вариация (дисперсия) - различие между единицами совокупности.

Вариационный ряд – ряд, в котором показано, как часто встречаются варианты каждого класса и как варьируют признаки от минимальной величины до максимальной.

Вероятность – возможность осуществления определенного события в некотором количестве случаев из общего числа возможных, или, иначе говоря, степень уверенности в том, что событие произойдет.

Вероятностный или **стохастический процесс** – процесс осуществления явления на основе известной его возможности или вероятности.

Взаимодействия - эффект *взаимодействия* возникает, когда зависимость между двумя или более переменными изменяется под воздействием одной или нескольких других переменных. Другими словами, сила или знак (направление взаимодействия) зависимости между двумя или более переменными зависит от значения принимаемого некоторыми другими переменными. Термин *взаимодействие* был впервые использован в работе Фишера (Fisher, 1926). Отметим, что слово "зависит" в данном контексте не означает причинной зависимости, а просто отражает тот факт, что в зависимости от рассматриваемого подмножества наблюдений (от значения модифицирующей переменной или переменных) характер зависимости будет меняться (модифицироваться).

Внутриклассовый коэффициент корреляции - значение внутриклассового коэффициента корреляции для популяции является мерой однородности наблюдений внутри классов случайного фактора относительно изменчивости наблюдений между классами. Он равен нулю только в случае, когда оцениваемый эффект случайного фактора равен нулю, и достигает единицы только если оцениваемый эффект ошибки равен нулю, при условии, что общая дисперсия наблюдений отлична от нуля. *Внутриклассовый коэффициент корреляции* может быть измерен с помощью метода оценивания компонент дисперсии.

Временной ряд - это последовательность измерений в последовательные моменты времени. Анализ *временных рядов* включает широкий спектр разведочных процедур и исследовательских методов, которые ставят две основные цели: (a) определение природы временного ряда и (b) прогнозирование (предсказание будущих значений временного ряда по настоящим и прошлым значениям). Обе эти цели требуют, чтобы модель ряда была идентифицирована и, более или менее, формально описана. Как только модель определена, вы можете с ее помощью интерпретировать рассматриваемые данные (например, использовать в вашей теории для понимания сезонного изменения цен на товары, если занимаетесь экономикой). Не обращая внимания на глубину понимания и справедливость теории, вы можете экстраполировать затем ряд на основе найденной модели, т.е. предсказать его будущие значения.

Выборочная совокупность – сравнительно небольшая по объему совокупность, входящая в состав генеральной.

Генеральная совокупность – теоретически бесконечно большая или приближающаяся к бесконечности совокупность.

Групповое программное обеспечение - это программное обеспечение, которое дает возможность группе пользователей, использующих компьютерную сеть, одновременно работать над конкретным проектом. Оно содержит средства для организации связи (электронную почту), для совместной обработки документов, проведения анализа, создания отчетов и статистической обработки данных, а также календарного планирования и наблюдения. При этом обрабатываемые документы могут содержать информацию любого типа: текст, картинки или мультимедийный формат.

Дискриминантный анализ - используется для принятия решения о том,

какие переменные дискриминируют или разделяют объекты на две или более естественно возникающих групп (его используют как метод проверки гипотез или как метод разведочного анализа).

Дисперсионный анализ – позволяет оценивать значимость влияния отдельных факторов, а также их относительную роль в общей изменчивости. **Д. а.** был разработан английским математиком и биологом Р. Фишером.

Доверительные вероятности – вероятность, при достижении которой можно с большой степенью уверенности заключить определенный вывод. В биологии используются доверительные вероятности: 0,95 и 0,99. Понятие **Д.В.** было введено Р. Фишером.

Доверительные границы или **доверительный интервал** - используются для оценки той или иной величины, указывают те границы, в которых она может находиться при разных вероятностях.

Доля выборки – отношение n/N, где n – численность выборочной совокупности, а N – численность генеральной совокупности. Используется для получения более точного значения средней ошибки.

Желаемая точность – допустимое расхождение между средней арифметической (по данному признаку) выборки и средней арифметической генеральной совокупности.

Закон больших чисел – выражает связь между статистическими показателями выборочных и генеральных совокупностей, заключается в том, что чем больше число n некоторых случайных величин, тем их средняя арифметическая ближе к средней арифметической генеральной совокупности.

Интервальная шкала - эта шкала измерений позволяет не только упорядочить наблюдения, но и количественно выразить расстояния между ними (при этом на шкале не обязательно присутствует *абсолютная* нулевая отметка).

Интерполяция - восстановление значения функции в промежуточной точке по известным ее значениям в соседних точках.

Категоризация, группировка, разбиение на подмножества - одним из наиболее важных, общих, а также мощных аналитических методов заключается в разделении (разбиении) данных на несколько подмножеств и последующее сравнение структуры данных в полученных подмножествах. У этого общего метода имеется много различных названий (в том числе: разбиение, группировка, категоризация, расщепление, разветвление и условный анализ), и он используется как для разведочного анализа данных, так и для проверки гипотез.

Качественная изменчивость – изменчивость, различия между вариантами которой выражаются в каких-либо качествах.

Классификация - отнесение наблюдения к одному из нескольких, заранее известных классов (представленных значениями номинальной выходной переменной).

Кластерный анализ - термин *кластерный анализ* (впервые ввел Tryon, 1939) в действительности включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих

областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры, т.е. развернуть таксономии и определить кластеры схожих объектов. Например, биологи ставят цель разбить животных на различные виды, чтобы содержательно описать различия между ними. В соответствии с современной системой, принятой в биологии, человек принадлежит к приматам, млекопитающим, амниотам, позвоночным и животным. Заметьте, что в этой классификации, чем выше уровень агрегации, тем меньше сходства между членами в соответствующем классе. Человек имеет больше сходства с другими приматами (т.е. с обезьянами), чем с "отдаленными" членами семейства млекопитающих (например, собаками) и т.д.

Ковариация - показатель, являющийся связующим звеном между корреляционным и регрессионным методами анализа.

Количественная дискретная (прерывная) изменчивость – изменчивость, при которой различия между вариантами отдельными значениями случайной переменной, выражаются целыми числами, между которыми нет и не может быть переходов.

Количественная непрерывная изменчивость – вариация, при которой значения вариант выражаются как целыми, так и дробными числами.

Комплексные числа - это множество чисел, которое включает все действительные и мнимые числа. Комплексное число представляется выражением вида $a + ib$, где a и b - действительные числа, i - мнимая единица,

Компоненты дисперсии (в смешанной модели дисперсионного анализа). Термин *компоненты дисперсии* используется в контексте дисперсионного анализа и планирования эксперимента, включающего случайные эффекты, для обозначения оценки (доли) дисперсии, которая связана с этими эффектами.

Корреляция - это мера связи между двумя переменными. Коэффициент корреляции может изменяться от -1.00 до +1.00. Значение -1.00 означает полностью отрицательную корреляцию, значение +1.00 означает полностью положительную корреляцию. Значение 0.00 означает отсутствие корреляции.

Корреляция Пирсона - наиболее часто используемый коэффициент корреляции Пирсона r (Pearson, 1896) называется также *линейной корреляцией* (термин корреляция впервые ввел Galton, 1888), т.к. измеряет степень линейных связей между переменными. Можно сказать, что корреляция определяет степень, с которой значения двух переменных пропорциональны друг другу. Важно, что значение коэффициента корреляции не зависит от масштаба измерения. Например, корреляция между ростом и весом будет одной и той же, независимо от того, проводились измерения в дюймах и фунтах или в сантиметрах и килограммах. Пропорциональность означает просто линейную зависимость. Корреляция высокая, если на графике зависимость можно представить прямой линией (с положительным или отрицательным углом наклона). Проведенная прямая называется прямой регрессии или прямой, построенной методом наименьших квадратов. Последний термин связан с тем, что сумма квадратов расстояний (вычисленная по оси Y) от наблюдаемых точек до прямой является

минимальной из всех возможных. Заметим, что использование квадратов расстояний приводит к тому, что на оценки параметров сильно влияют выбросы. Корреляция Пирсона предполагает, что две рассматриваемые переменные измерены, по крайней мере, в интервальной шкале.

Корреляционные или статистические связи – связи, при которых численному значению одной переменной соответствует много значений другой переменной.

Коэффициент вариации – применяется при сравнении вариации различных признаков, представляет собой отношение σ к x, выраженное в процентах.

Коэффициент детерминации - это квадрат корреляции Пирсона между двумя переменными. Он выражает количество дисперсии, общей между двумя переменными.

Коэффициент корреляции *r* – указывает на степень связи в вариации двух переменных величин или на меру тесноты этой связи.

Коэффициент регрессии - количественная мера регрессии, вычисляемая если известны сигмы обоих вариационных рядов по признакам x и y, и коэффициенты корреляции между ними.

Кривая распределения (вариационная кривая) – графическое изображение вариационного ряда.

Критерий соответствия хи-квадрат χ^2 – показатель, определяющий степень соответствия фактических данных теоретически ожидаемым, или согласие фактических данных с предложенной гипотезой.

Критерий Стьюдента *t* – применяется при малых выборках ($n \leq 30$), характеризует отклонение выборочных средних от генеральной средней. Устанавливает тот факт, что среднее квадратическое отклонение для малых выборок постоянно отличается от того, которое ожидалось бы при нормальном распределении.

Круговая диаграмма - последовательность значений переменной изображается в виде последовательных круговых секторов (термин "круговая диаграмма" был впервые использован Хаскеллом в 1922 г.); размер каждого сектора пропорционален соответствующему значению. Значения должны быть больше 0 (нулевое и отрицательные значения не могут быть представлены в виде круговых секторов). Круговая диаграмма интерпретирует данные самым непосредственным образом: одно наблюдение соответствует одному сектору.

Лимиты (пределы) – значения крайних классов, верхняя и нижняя граница вариационного ряда.

Метод регрессии – метод, позволяющий установить, как количественно меняется одна величина при изменении другой на единицу.

Медиана – значение варианты, находящееся точно в середине ряда.

Множественная корреляция – зависимость изменения величины x от одновременного изменения величин y,z и т.д.

Мода – значение модального класса, являющееся как бы типичной для всей совокупности.

Модальный класс – класс, обладающий наибольшей частотой.

Номинальные переменные - переменные, которые могут принимать конечное множество значений, например, *Пол* = {*Муж, Жен*}.

Нормальная вариационная кривая – симметричная плавная кривая, при которой верхние границы ломанной линии полигона сливаются в гладкую кривую линию.

Нормированное отклонение *t* – представляет собой отклонение тех или других вариант от их средней арифметической, выраженное в долях среднего квадратического отклонения.

Нулевая гипотеза - согласно этой гипотезе, первоначально принимается, что между данными показателями (или группами, на основе которых они получены) достоверного различия нет, т.е. что обе группы вместе составляют один и тот же однородный материал, одну совокупность.

Общность - это доля дисперсии, которая является общей для данной и всех остальных переменных. Доля дисперсии, которая является характерной для данной переменной (иногда называется характерностью) получается после вычитанием общности из дисперсии переменной. Другими словами дисперсия переменной есть общность плюс характерность. Обычно вначале в качестве оценки общности используют коэффициент множественной корреляции выбранной переменной со всеми другими.

Объем совокупности – число единиц совокупности.

Отрицательная корреляция - обратная зависимость между признаками: увеличение одного признака соответственно связано с уменьшением другого.

Ошибка выборочности или **ошибка репрезентативности** - представляют собой среднюю величину расхождения между средними значениями изучаемых признаков в выборках и генеральной совокупности.

Ошибка выборочности коэффициента корреляции – мера расхождения между коэффициентами корреляции для выборочной и генеральной совокупности.

Полигон распределения – графическое изображение конкретных вариационных рядов, применяющееся при дискретной вариации.

Положительная корреляция – прямая зависимость между признаками: при увеличении одного увеличивается и другой.

Поправка на непрерывность Иейтса – применяется при вычислении χ^2 в случае если исследуются малочисленные группы.

Ранжировка – расположение всех вариант по порядку от минимальных до максимальных значений.

Распределение Пуассона или **пуасоново распределение** – в биологии применяется для анализа редко наблюдаемые явления.

Симметричное распределение - если вы разобьете распределение пополам в точке среднего (или медианы), то распределения значений с двух сторон от этой центральной точки будут "зеркальным отображением" друг друга.

Случайная переменная – величина, изменяющаяся под влиянием многих случайных причин, которая может принимать разные значения.

Совокупность - всякое множество отдельных отличающихся друг от друга и в то же время сходных в некоторых существенных отношениях объектов.

Среднее - показывает "центральное положение" (центр) переменной и рассматривается совместно с доверительным интервалом. Обычно интерес представляют показатели (например, среднее), дающие информацию о популяции в целом. Чем больше размер выборки, тем более надежна оценка среднего. Чем больше изменчивость данных (больше разброс), тем оценка менее надежна.

Средняя арифметическая $\bar{x}$ – некоторая уравненная величина, отражающая основные свойсива всех членов совокупности.

Средняя геометрическая – статистический показатель, применяемый в случае, если возрастание данного признака происходит умножением пропорционально степени.

Стандартная ошибка - термин стандартная ошибка среднего был впервые введен Юлом (Yule, 1897). Эта величина характеризует стандартное отклонение выборочного среднего, рассчитанное по выборке размера n из генеральной совокупности, и зависит от дисперсии генеральной совокупности (сигма) и объема выборки (n).

Стандартное отклонение - (термин был впервые введен Пирсоном, 1894), это широко используемая мера разброса или вариабельности (изменчивости) данных.

Таблицы сопряженности – таблицы, в которых предусматривается распределение групп по признакам, сопряженность или связь между которыми нужно будет установить.

Теоретические (априорные) вероятности – вероятности, которые знают заранее до проведения опыта.

Уровень значимости – обозначает вероятность получения случайного отклонения от установленных с определенной вероятностью результатов. Вероятности 0,95 (95%) соответствует уровень значимости 0,05% (5%). При вероятности 0,99% (99%) уровень значимости 0,01 (1%).

Функциональная зависимость – зависимость, при которой, каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной.

Частная корреляция - корреляция между двумя переменными, вычисленная после устранения влияния всех других переменных, называется частной корреляцией

Число степеней свободы df – величина n-1.

Эмпирические (апостериорные) вероятности – вероятности, которые получены после проведения опыта.

Биологический факультет

Кафедра зоологии, физиологии и генетики

КУРАЧЕНКО И.В., ЗЯТЬКОВ С.А., ГОНЧАРЕНКО Г.Г.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ К ЛАБОРАТОРНЫМ ЗАНЯТИЯМ ПО КУРСУ

«Биометрия»

для студентов специальности 1 31 01 01 02

Биология (научно-педагогическая деятельность)

Гомель, 2019

Лабораторное занятие 1 «Первичная и вторичная группировка экспериментальных данных»

Материалы и оборудование: данные замеров статистических величин; миллиметровка; калькуляторы.

Цель. Образование выборочных, сгруппированных статистических совокупностей и их графическое отображение.

Пояснения к заданиям:

Статистическая совокупность подвергается ранжированию, которое заключается в следующем:

а) находится минимальная и максимальная варианты (лимиты);

б) определяют вариационный размах: $\rho = x_{max} - x_{min}$

в) при p≤11, проводят первичную группировку, т.е. i=1

г) ранжирование данных и заполнение рабочей таблицы.

Ход работы

Задание 1. Провести группировку данных по качественным признакам по следующей схеме: объект, предмет, вариация, объём совокупности, число классов.

Задача. На звероводческой ферме выращивают норок:стандартные коричневые – 120 особей, сапфировых – 180 особей, серебристо-голубых – 160 особей, черных – 40 особей. Определить долю особей каждого из окрасов, изобразить диаграмму распределения норок по окрасу. Сделать обоснованный вывод.

Задание 2. Провести первичную группировку по экспериментальным данным о длине левого уха (в см) у 70 кроликов-мериносов:

12 11 13 14 10 13 13 13 14 12 12 12 14 13 13 14 13 14 12 15 12 11
13 10 12 13 12 11 12 14 11 10 15 12 11 11 13 13 12 15 11 12 13 11
12 12 14 16 12 14 12 11 14 12 14 11 13 12 14 11 14 12 14 11 10 16
11 12 12 12.

Данные представить в виде рабочей таблицы

Классы, x_i	Разноска	Частота, f_{xi}

Задание 3. Составить безынтервальный вариационный ряд, найти моду и медиану. Вариационный ряд отобразить графически (на оси ОХ отметить значения классов, на оси ОУ – значения частот) и сделать обоснованный вывод.

Задание 4. Решить упражнение №1,2,5 из учебника Рокицкого П.Ф. (стр.20-23).

Вопросы для самоконтроля

1 Что такое выборочная и сгруппированная статистическая совокупность?

2 Как образуется классовый интервал?

3 Что такое кумулята, огива и как они отображаются графически?

4 Как построить полигон распределения статистических частот, гистограмму, диаграмму?

Лабораторное занятие 2 «Совокупность и вариационный ряд»

Материалы и оборудование: плоды акации, ветви хвойных деревьев, задачник, калькуляторы.

Цель: Образование выборочных статистических совокупностей методом вторичной группировки и их графическое отображение.

Пояснения к заданиям:

Пояснения к заданиям:

Статистической совокупностью называют некоторое множество относительно однородных предметов или объектов, объединяемых по выбранному признаку. Теоретически бесконечно большая или приближающаяся к бесконечности совокупность всех единиц или членов вариационного ряда называют ***генеральной.*** Генеральная совокупность может состоять из такого большого количества единиц, что изучить их все не представляется возможным. Поэтому приходится иметь дело со сравнительно небольшими, ***выборочными*** совокупностями. Выборочная совокупность наиболее полно отражающая свойства генеральной называется репрезентативной, например, при изучении морфометрических показателей рептилий исключаются рептилии с анатомированными хвостами, при изучении роста деревьев в высоту исключаются деревья, сломанные бурей, поврежденные огнем и т.д. При образовании выборки используется метод случайного отбора, то есть выдерживается принцип объективности.

Статистическая совокупность подвергается упорядовачиванию, при вторичной группировке, которое заключаются в следующем:

а) находится минимальная и максимальная варианты (лимиты);

б) определяют вариационный размах: $\rho = x_{max} - x_{min}$, по величине которого судят о группировке данных: если p≤11, проводят первичную группировку, т.е. i=1; если p≥11, весь диапазон значений признака разбивается на «классовые промежутки» и величину интервала определяют по формуле:

$$i = \frac{Xmax - Xmin}{1 + 3.32\log n}$$

в) Находится нижняя граница первого интервала: $l=x_{max} - i/2$

Ход работы:

Задание 1. Провести вторичную группировку по экспериментальным данным о длине тела у 85 экземпляров густеры озера Швакшта (в мм):

143 143 128 130 143 127 143 157 120 119 94 145 138 118 134 95
148 144 120 140 140 120 138 142 153 130 138 153 135 124 130 148
150 138 130 137 135 134 135 136 142 124 114 142 139 111 133 165
164 127 126 145 126 145 125 132 134 172 139 137 138 137 137 133
151 139 139 117 141 131 100 107 140 129 132 125 120 142 158 141
124 154 154 139 117

Данные представить в виде рабочей таблицы

Границы классов	Срединные значения классов, x_i	Разноска	Частота, f_{xi}

Задание 2. Составить интервальный вариационный ряд. Найти моду и медиану. Вариационный ряд отобразить графически (на оси ОХ отметить нижние значения классов, на оси ОУ – значения частот) и сделать обоснованный вывод.

Задание 3. Изучен живой вес 70 телят ярославских помесей при рождении (в кг):

27 32 32 31 32 28 37 35 26 28 32 28 35 36 28 39 43 28 33 36 34 26
32 33 36 30 35 36 28 37 43 32 32 23 26 26 36 28 27 35 37 34 40 32
33 32 35 32 28 26 37 27 31 35 37 31 29 30 26 29 29 31 32 35 41 40
31 36 29 33

Составить вариационный ряд. Найти моду и медиану.

Задание 4. Провести подсчет семян в 30 плодах акации, сгруппировать данные в безынтервальный вариационный ряд, определить моду и медиану. Построить полигон распределения. Собрать данные группы, увеличив многократно объем совокупности и сделать обоснованный вывод.

Данные представить в виде рабочей таблицы

Классы, X_i	Разноска	Частота, f_{xi}

Вопросы для самоконтроля

1 Что такое выборочная и сгруппированная статистическая совокупность?

2 Что такое репрезентативность?

3 Типы отбора выборок?

4 Что такое классовый промежуток?

5 Сформулировать закон больших чисел??

6 В каких случаях проводится вторичная группировка?
7 Закономерности распределения вариант в вариационном ряду?

Лабораторное занятие 3 «Закономерности распределений»

Материалы и оборудование: мешочки с семенами кукурузы, калькуляторы.

Цель: экспериментальное доказательство теорем сложения и умножения вероятностей

Пояснения к заданиям:

Для того чтобы выяснить, произойдет или не произойдет событие при заданном комплексе факторов, нужно осуществить этот комплекс, т. е. провести испытание. Испытанием является любой эксперимент, в результате которого производят наблюдения. События, происходящие при одном и том же комплексе факторов, называются однородными. Установлено, что однородные случайные события в большой их массе подчиняются некоторым закономерностям. Эти закономерности получили название вероятностных. События с одинаковыми возможностями осуществления называются равновозможными. Числовая характеристика случайного события, обладающая тем свойством, что для любой достаточно большой серии испытаний частота события лишь незначительно отличается от этой характеристики, называется вероятностью события. Исходы испытания являются простейшими случайными событиями. Вероятностью случайного события называется отношение числа отходов, благоприятствующих событию, к числу всех возможных исходов.

Если некоторое событие может произойти при n испытаниях и a — число исходов, которые благоприятствуют наступлению события, а b — не благоприятствуют, то вероятность того, что событие произойдет, может быть определена как $p=a/n$. Вероятность того, что событие не произойдет, будет $q=b/n$. (p+ q =1)

Основные теоремы теории вероятностей

Вероятность суммы двух несовместных, независимых событий равна сумме их вероятностей

P(A+B)=P(A)+P(B).

Вероятность сложного события (т. е. наступления двух событий независимых одно от другого) равна произведению вероятностей отдельных событий

P(AxB)=P(A)xP(B).

Вероятности отдельных возможных исходов даются последовательными членами разложения Бинома Ньютона:

$(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$.

Ход работы:

Задание 1. Для доказательства теоремы сложения двух независимых случайных событий провести 20 извлечений по одному семени из мешочка, в котором находится по 10 зеленых, белых и желтых семян кукурузы. Результаты испытаний занести в таблицу:

События	Разноска	Число случаев

Желтое		m_1
Зеленое		m_2
Белое		m_3

Задание 2. Определить эмпирические вероятности данных событий. Найти вероятность того, что вынутое наугад семя окажется окрашенным.

Задание 3. Определить теоретические вероятности данных событий и найти отклонение $P_{эмп} - P_{теор}$

Задание 4. Решить упражнения:

1. В урне m белых и n черных шаров. Какова вероятность того, что вынутый наугад шар окажется: а) белым, б)черным.
2. Стрелок сделав 200 выстрелов, попал в цель 190 раз. Какова вероятность попадания в цель? Сколько будут попаданий в цель, если стрелок сделает 300 выстрелов?
3. При бросании двух кубиков, какова вероятность выпадения суммы цифр, равной 7? Не менее семи?
4. При бросании трех монет, какова вероятность выпадения гербом: одной монеты? Двух? Трех? Не менее одной монеты?
5. При бросании трех кубиков, какова вероятность выпадения суммы цифр, не менее семи?
6. В городе N с населением в 100000 жителей родилось 8000 новорожденных? Какова вероятность рождения детей: абсолютная, удельная?
7. В урне 4 белых и 7 красных шаров. Какова вероятность того,что вынутый шар окажется белым? Красным?
8. В мешочке пять букв М, О, Л, О, Т. Какова вероятность того, что доставая по одной карточке наугад получится слово «Молот»? «Том»?
9. В колоде 36 карт. Какова вероятность того, что вынутая карта окажется тузом?
10. В лотерее 4 выигрышных и 96 безвыигрышных билета. Какова вероятность, что два билета окажутся выигрышными?
11. В пассажирском поезде 12 вагонов. Какова вероятность того, что двое друзей независимо друг от друга окажутся в одном вагоне?

Задание 5. С помощью треугольника Паскаля и бинома Ньютона получить вероятности распределения различных комбинаций детей по полу в семьях, имеющих четырех детей.

Вопросы для самоконтроля

1 Что такое вероятность?
2 Как осуществить расчет вероятности случайного события?
3 Теорема сложения вероятностей? Примеры.
4 Теорема умножения вероятностей? Примеры
5 Бином Ньютона. Формула последнего множителя?
6 Достоверные события? Независимые, несовместимые события?

Лабораторное занятие 4 «Средние величины»

Материалы и оборудование: данные статистических величин и расчетов по лабораторным работам тем 1, 2; калькуляторы.

Цель: Определение статистических показателей выборочной совокупности

Пояснения к заданиям:

Наиболее распространенным и используемым показателем является ***среднеарифметическая величина***. Она рассчитывается по формуле:

$$\overline{X} = (x_i + x_2 + \ldots + x_n)/N = (\Sigma x_i)/N,$$

где N - общее число вариант; Σ - знак суммирования;

x_i – значения вариант.

Средняя квадратическая величина используется при вычислении средней из величин объема, запаса, площади. Рассчитывается по формуле:

$$\overline{X}_q = \sqrt{(\sum x_i^2)/N}$$

где x_i^2 - квадраты замеряемых величин – объем, площадь и т. п.; N - общее число деревьев в выборке; x_i – значения вариант.

Средняя геометрическая M_g (или $\overline{X_g}$) используется для расчета среднего темпа роста изучаемого признака. Она известна также как средняя логарифмическая, так как ее логарифм есть арифметическая средняя логарифмов составляющих величин.

Вычисляется по формуле:

$$\overline{X_g} = \sqrt{x_i \cdot x_2 \cdot x_{3\ldots} \ldots x_n},$$

где x_1, $x_{2, \ldots,}$ x_n - темпы роста (величины, показывающие, во сколько раз увеличивался признак от периода к периоду); n - число периодов.

При $n>2$ формулу удобнее применять в логарифмическом виде:

$$\ln \overline{X_g} = \frac{1}{n}(\ln x_1 + \ln x_2 + \ldots + \ln x_n) = (1/n) \cdot (\sum \ln x_i).$$

откуда $\overline{X}_g = e^{\ln}$, где e – основание натуральных логарифмов, равно 2,72.

Средняя гармоническая используется для вычисления средней величины отношений двух варьирующих величин. Определяется по формуле:

$$\overline{X_h} = N/(\sum 1/x_i),$$

где N - число значений; x_i – значения соотношений величин

Рассеяние вариант выборки относительно средней характеризуется:

центральным отклонением;
дисперсией;
среднеквадратическим отклонением;
коэффициентом вариации.

Их вычисляют по формулам:

центральное отклонение $\alpha = x_i - \bar{x}$,

дисперсия $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{k} (x_i - \bar{x})^2$

среднеквадратическое отклонение $\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{k} (x_i - \bar{x})^2}$

коэффициент вариации $C_\chi = \frac{\sigma_\chi \cdot 100}{\bar{\chi}}$

При C_χ <30% - выборка имеет большую степень концентрации вариант возле величины. При 30%≤C_χ≤100% - степень концентрации допустимая. При C_χ >100% - делается вывод о неоднородности выборки.

Ход работы:

Задание 1. Рассчитать средние величины для данных о весе 11 поросят при рождении: 1,0 1,2 1,5 1,3 1,3 1,4 1,4 1,1 1,0 0,9 2,5. Вариационный ряд не составлять, произвести расчеты прямым способом.

Задание 2. Рассчитать средние величины для данных о длине левого уха (в см) у 70 кроликов-мериносов:

12 11 13 14 10 13 13 13 14 12 12 12 14 13 13 14 13 14 12 15 12 11
13 10 12 13 12 11 12 14 11 10 15 12 11 11 13 13 12 15 11 12 13 11
12 12 14 16 12 14 12 11 14 12 14 11 13 12 14 11 14 12 14 11 10 16
11 12 12 12.

Проведя первичную группировку, рассчитать средние величины следующими способами:

а) прямым способом через значения вариант по таблице

x_i	разноска	f_i	x_i^2	$f x_i$	$f x_i^2$
				$\sum f x_i =$	$\sum f x_i^2 =$

б) прямым способом через центральные отклонения по таблице

x_i	разноска	f_i	$f\,x_i$	$x_i - X$	$f\,(x_i - X)$	$f\,(x_i - X)^2$
			$\sum f\,x_i =$			$\sum f\,(x_i - X)^2 =$

в) способом условной средней (A=Mo, тогда $a=x_i - A$) по таблице

x_i	разноска	f_i	a	fa	fa^2
				$\sum f\,a =$	$\sum fa^2 =$

Задание 3. Изучен живой вес 70 телят ярославских помесей при рождении (в кг):

27 32 32 31 32 28 37 35 26 28 32 28 35 36 28 39 43 28 33 36 34 26
32 33 36 30 35 36 28 37 43 32 32 23 26 26 36 28 27 35 37 34 40 32
33 32 35 32 28 26 37 27 31 35 37 31 29 30 26 29 29 31 32 35 41 40
31 36 29 33

Данные представить в виде рабочей таблицы и произвести расчеты способом условной средней:

Границы классов	Срединные значения классов, x_i	x_i	разноска	f_i	α	f α	$f\,\alpha^2$
						$\sum f\,\alpha =$	$\sum f\,\alpha^2 =$

Вопросы для самоконтроля

1 На какие группы делятся статистические показатели?

2 Что такое основное отклонение, дисперсия, коэффициент вариации и как они вычисляются?

3 Что такое средняя арифметическая, средняя квадратическая, средняя гармоническая, средняя геометрическая и как они вычисляются?

4 Свойства средней арифметической?

Лабораторное занятие 5 «Статистический анализ частот распределений»

Материал и оборудование: таблицы, калькуляторы

Цель: определение нормальности распределения данных с помощью критерия χ^2

Пояснения к заданиям:

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:

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

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.

Ход работы:

Задание. Для предложенных примеров произвести расчет критериев

- достоверности
- Фишера
- хи-квадрат

и оценить их величину.

Задача 1. Из 100 вакцинированных заболело 8 человек. Определить уровень заболеваемости в исследуемой группе людей и сравнить с теоретической 0,12. Провести анализ: составить диаграмму заболеваемости, вычислить статистические характеристики (p, δ_p, m_p, P, C_v), дать оценку достоверности (доверительный интервал при трех уровнях значимости; H_0; критерий Стьюдента). Сделать обоснованный вывод.

Задача 2. Получены данные о распределении бычков и телочек в совхозе «Восток» за 2002 год: телочек – 1256, бычков – 1857. Соответствует ли распределение бычков и телочек соотношению 1:1. Провести анализ Провести анализ: составить диаграмму заболеваемости, вычислить статистические характеристики (p, δ_p, m_p, P, C_v), дать оценку достоверности (доверительный интервал при трех уровнях значимости; H_0; критерий Стьюдента). Сделать обоснованный вывод.

Задача 3. Получены данные о распределении самок и самцов плодовой мушки: самок – 126, самцов – 250. Соответствует ли данное распределение соотношению 1:1. Провести анализ: составить диаграмму заболеваемости, вычислить статистические характеристики (p, δ_p, m_p, P, C_v), дать оценку достоверности (доверительный интервал при трех уровнях значимости; H_0; критерий Стьюдента). Сделать обоснованный вывод.

Вопросы для самоконтроля

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

Лабораторное занятие 6 «Показатели вариации»

Материалы и оборудование: данные статистических величин и расчетов по лабораторным работам тем 1, 2, 5; таблицы; калькуляторы.

Цель: Определение основных ошибок статистических показателей и их доверительных интервалов; оценить асимметрию и эксцесс, правило трех сигм.

Пояснения к заданиям:

Выборочная совокупность довольно точно воспроизводит свойства и соотношения в генеральной совокупности, но не абсолютно точно вследствие вариации изучаемых признаков. Поэтому между статистическими показателями выборочной совокупности и действительными значениями этих показателей генеральной совокупности всегда будут некоторые расхождения, которые являются случайными ошибками выборки (иначе - случайными ошибками репрезентативности) и называются ***основными ошибками*** того или иного статистического показателя. На основании величины этой основной ошибки и значения соответствующего показателя выборки можно судить о действительном значении данного показателя в генеральной совокупности. Так, с вероятностью равной 0,68 (в 68% случаев из ста), можно утверждать, что расхождение между действительным значением данного показателя в генеральной совокупности и вычисленным его значением для выборки не превышает однократного значения основной ошибки этого показателя (со знаком плюс или минус); предельное же расхождение не превышает трехкратного значения основной ошибки (о чем можно утверждать с вероятностью 0,997 или 99,7% случаев из ста).

Основные ошибки статистических показателей вычисляются по формулам:

Ошибка среднего значения (или ошибка стандартная, выборочности, статистическая) $m_x = \frac{\sigma}{\sqrt{n}}$; $m_x = \frac{\sigma}{\sqrt{n-1}}$ при малых объемах (n≤30).

Отклонение распределений фактических данных от нормального типа характеризуется основными моментами - r3,, r4 , которые показывают асимметричность коэффициент асимметрии - A и крутость коэффициент эксцессов распределений - E:

$A= r_3= \mu_3 / (\sqrt{\mu_2})^3$; $E=r_4-3= \mu_4/(\mu_2)^2-3$

Или

$A= \sum f(x_i - X)^3 / n\cdot\sigma^3$ $E= \sum f(x_i - X)^4 / n\cdot\sigma^4 - 3$

Ошибка показателя асимметрии проводится по формуле: $m_A = \sqrt{6/N}$. Ошибка показателя эксцесса равна удвоенной ошибке показателя асимметрии. Оценку достоверности асимметрии и эксцесса проводят по формулам: $t= A /m_a \geq 3$ и $t= E /m_e \geq 3$

Ход работы:

Задание 1. Произвести расчет средних величин, нормированных отклонений по данным о длине правого уха (в см) у 60 серебристо-черных лисиц:

12 10 14 14 13 12 12 12 15 13 11 12 12 14 12
12 13 14 11 13 14 12 13 12 12 14 12 14 13 13
12 13 12 12 13 12 11 11 12 13 14 12 14 12 14
10 11 10 11 15 11 16 11 16 11 11 11 12 15 14

Применить прямой способ через центральные отклонения по таблице

x_i	разноска	f_i	$f\,x_i$	$x_i - X$	$f\,(x_i - X)$	$f\,(x_i - X)^2$	$f\,(x_i - X)^3$	$f\,(x_i - X)^4$	t	f/n %
			$\sum f\,x_i =$			$\sum =$	$\sum =$	$\sum =$		

Задание 2. Построить полигон распределения, отметив моду, медиану и среднее арифметическое на графике. Сделать вывод о характере распределения изучаемого признака.

Задание 3. Для доказательства правила трех сигм построить кривую нормального распределения, отметив на оси абсцисс – нормированные отклонения, на оси ординат – относительные частоты в %.

Задание 4. Рассчитать коэффициенты асимметрии и эксцесса, дав им оценку достоверности.

Задание 5. Определить долю вариант под кривой нормального распределения, используя таблицу 1 (Рокицкий), в пределах: а) от 0 до +2,15σ; б) от -1,09σ до -0,08σ; в) за пределами ±2,31σ; г) от X до 2,34σ; д) между ±1,98σ.

Таблица I

Таблица вероятностей при нормальном распределении.
Доли площади под нормальной кривой в пределах от $-t$ до $+t$

t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6680	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9960	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9981	9986	9990	9993	9995	9997	9998	9999	9999

Задание 6. Рассчитать статистические показатели и ошибку выборочности по экспериментальным данным о длине тела у 85 экземпляров густеры озера Швакшта (в мм):

143 143 128 130 143 127 143 157 120 119 94 145 138 118 134 95
148 144 120 140 140 120 138 142 153 130 138 153 135 124 130 148
150 138 130 137 135 134 135 136 142 124 114 142 139 111 133 165
164 127 126 145 126 145 125 132 134 172 139 137 138 137 137 133
151 139 139 117 141 131 100 107 140 129 132 125 120 142 158 141
124 154 154 139 117

Данные представить в виде рабочей таблицы и произвести расчеты способом условной средней:

Границы классов	Срединные значения классов, x_i	x_i	разноска	f_i	α	$f\alpha$	$f\alpha^2$
						$\sum f\alpha =$	$\sum f\alpha^2 =$

Вопросы для самоконтроля

1 Как осуществить расчет нормального распределения?
2 Как определяется тип кривой при помощи критерия Пирсона?
3 Для каких целей применяются статистические критерии?
4 Дать определение «Ошибка выборочности»?
5 Как определяются основные ошибки статистических показателей?

Лабораторное занятие 7 «Репрезентативность выборочных показателей»

Материал и оборудование: данные статистических величин и расчетов по лабораторным работам тем 5, 6; таблицы; калькуляторы.

Цель: оценить степень достоверности статистических показателей

Пояснения к заданиям:

Точность опыта, или процент ошибки наблюдения – это процент расхождения между генеральной и выборочной средней, который вычисляется по формуле: $p = \frac{100 \cdot m_x}{X}$ или же по формуле $p = \frac{C}{\sqrt{N}}$. Точность опыта показывает, насколько процентов можно ошибиться, если утверждать, что генеральная средняя равна полученной выборочной средней.

Полученный процент ошибки сопостовляется с заданным: если он не больше заданного, точность достаточная, а если больше, то точность результата является неудовлетворительной; значит, следует увеличить число наблюдений.

После вычисления того или иного статистического показателя необходимо проверить степень его надежности или достоверности путем деления величины данного показателя на величину его основной ошибки: $t = \frac{X}{m_x}$, где

X - величина любого статистического показателя;

m_x - величина ошибки любого статистического показателя.

Если частное t получится равным или больше трех, то значение показателя является надежным, достоверным, и им можно пользоваться для разных сопоставлений и выводов. Если же это отношение будет меньше трех, то данный показатель оказывается ненадежным, величина его не достоверна и является лишь в той или иной мере вероятной. Такие показатели нельзя сопоставлять между собой или производить на основе их заключения. Нередко приходится решать вопрос, насколько существенно различие в значениях показателей какого-либо признака, вычисленных для разных совокупностей. С этой целью находится основная ошибка разницы чисел и доказывается ее достоверность по выше описанному принципу. Ошибка разности вычисляется как корень квадратный из суммы квадратов основных ошибок исследуемого показателя, то есть $m_s = \sqrt{m_1^2 + m_2^2}$.

Полученную разность показателей делят на его ошибку. Находится показатель существенности различия средних значений: $t = \frac{X_1 - X_2}{m_s} = \frac{X_1 - X_2}{\sqrt{m_1^2 + m_2^2}}$.

Если этот показатель получится больше трех, то различие существенно, доказано, и данное мероприятие вызвало существенное изменение; обе сравниваемые выборочные совокупности являются представителями качественно разных генеральных совокупностей. Если же он получится меньше трех, то можно утверждать, что расхождение оказалось случайным, недостоверным и во всяком случае целесообразность данного мероприятия осталась недоказанной.

Расчет доверительных интервалов статистических показателей. Среднее значение, основное (квадратическое) отклонение, коэффициент изменчивости, асимметрия, эксцесс дают представление о величине и форме распределений наблюдений. Однако они не дают представления о возможных значениях случайной величины. Оно заключается в вычислении вероятности того, что значение величины будет заключаться в определенных границах. Считается, что границы достоверно определены если вероятность близка к единице, например, 0,99 или 0,999 (99,0% или 99,9%). Соответствующие границы называются доверительными. В зависимости от типа распределения данных доверительный интервал рассчитывается двумя способами. В симметричных распределениях, близких к нормальному, размах отклонений данных от средней арифметической обычно равен приблизительно 3σ в обе стороны от значения средней. (так называемый закон трех сигм).

При расчете доверительного интервала для трех стандартных доверительных уровней: 95%, 99%, 99,9% t выбирается по числу степеней свободы из таблицы.

Значения показателя t (критерия Стьюдента)

Число степеней свободы	Доверительные уровни		
	95%	99%	99,9%
9	2,3	3,2	4,8
10	2,2	3,2	4,6
11-14	2,2	3,0	4,3
15-20	2,1	2,9	3,9
21-30	2,1	2,8	3,7
31-60	2,0	2,7	3,5
61-120	2,0	2,6	3,4
∞	1,96	2,58	3,29

Доверительный интервал статистического показателя, например, средней арифметической строится по формуле:

$$\overline{X} - t \cdot m_{\bar{x}} < \mu < \overline{X} + t \cdot m_{\bar{x}}$$

Подставляя в формулу величины среднего арифметического, коэффициента вариации, коэффициента асимметрии, эксцесса и их ошибок определяются доверительные интервалы для этих показателей.

Ход работы:

Задание 1. Получены данные о количестве хвостовых щитков у змей:

42	58	44	54	41	50	46	46	54	48	43	49
50	48	46	46	45	53	48	48	53	53	48	41
46	40	50	43	49	51	52	46	42	44	48	45
47	46	43	50	47	45	48	40	44	42	48	45
54	50	56	48	45	45	51	42	44	47	46	45

Провести анализ:

- Вычислить статистические характеристики (Mo, Me, X, σ, m_x, P, C_v).
- Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H0; критерий Стьюдента).
- Сделать обоснованный вывод. Ответ.

Задание 2. Провести статистические расчеты при качественной вариации признаков. Из 100 вакцинированных заболело 8 человек. Определить уровень заболеваемости в исследуемой группе людей и сравнить с теоретической 0,12. Провести анализ:

- Составить диаграмму заболеваемости.
- Вычислить статистические характеристики (p, δ_p, m_p, P, C_v).
- Дать оценку достоверности (доверительный интервал при трех уровнях значимости; H_0; критерий Стьюдента).
- Сделать обоснованный вывод. Ответ.

Задание 3. Используя критерий Стьюдента разницы двух средних, ответить на вопрос: отличаются ли по температуре тела самцы и самки тушканчиков.

Самцы:	37,5	35,9	37,5	37,8	37,2	37,9	36,9
Самки:	37,5	35,4	37,1	37,9	37,0	37,7	37,9

Вопросы для самоконтроля

1 Сущность нулевой гипотезы?

2 Как определяется доверительный интервал при нормальном распределении статистических показателей?

3 Что такое уровень значимости?

4 Что такое доверительная вероятность?

Лабораторное занятие 8 «Однофакторный дисперсионный анализ»,

Материалы и оборудование: калькуляторы, таблицы.

Цель. Определение значимости различия признака биологического объекта методом однофакторного дисперсионного анализа.

Пояснения к заданиям:

Дисперсионный или вариансный анализ (analysis of variance) предполагает установление роли отдельных факторов в изменчивости того или иного признака, при котором общая дисперсия как количественных, так и качественных признаков раскладывается на отдельные составляющие**.** У изучаемых признаков в эксперименте имеется не одно, а несколько значений, которые называют *градациями или уровнями фактора А.* Число наблюдений (вариант) в каждой группе обозначается как «n».

Схема обозначения членов вариационного ряда при однофакторном дисперсионном анализе

Число градаций фактора А	Повторности, x_i / x_i^2			n	$\sum x_i$ / $\sum x_i^2$	$\frac{(\sum x_i)^2}{n}$
	1	2	3			
контроль						
Смесь 1						
Смесь 2						
				$\sum n=N$	$\sum\sum x_i / \sum\sum x_i^2$	$\sum=$

Различают 3 типа варьирования:

а) σ_y^2- общее варьирование вариант, независимо от того, в какой группе они находятся, вокруг общей средней $\bar{x}$;

б) σ_x^2- варьирование средних каждого уровня данного изучаемого фактора, вокруг общей средней $\bar{x}$;

в) σ_z^2 - варьирование вариант внутри каждой группы вокруг каждой групповой средней $\bar{x}_i$ (так называемая остаточная).

Между ними существует соотношение:

$\sigma_y^2 = \sigma_x^2 + \sigma_z^2$

Для каждого типа варьирования вычисляются суммы квадратов отклонений по следующим формулам:

Общая сумма квадратов: $C_y = = \Sigma\sum X^2 - \frac{(\Sigma\sum X)^2}{N}$

Сумма квадратов для групповых средних (факторальная):

$$C_x = \sum \langle \Sigma X \rangle^2 / n - \frac{\left(\Sigma \sum X\right)^2}{N}$$

Сумма квадратов для внутригрупповая (случайная):

$$C_z = \Sigma \Sigma x \frac{2}{} - \frac{\left(\Sigma \sum X\right)^2}{N}$$

Где $\sum x_i$ для каждой группы (уровня фактора А);
n_i - число наблюдений в каждой группе;
N - общее число вариант.
r - число уровней (градаций) фактора.

При делении сумм квадратов, обозначаемых, на число степеней свободы получаются ***средние квадраты (вариансы)*** — σ^2 непосредственно измеряющие суммарную вариацию.

Оценка дисперсии каждой из групп связана со степенью свободы, при этом необходимо учитывать большую вариансу (например, если $\sigma^2_x \geq \sigma^2_z$, то за $df_1 = df_x = r-1$. $df_2 = df_z = N-r$. Далее проводится проверка гипотезы $H_0 : \overline{X}_1 = \overline{X}_2 = \ldots = \overline{X}_k$, т.е. утверждения, что все групповые средние не зависят от влияния фактора А. Если верна H_0, то межгрупповая дисперсия (в генеральной совокупности) должна быть равна внутригрупповой, т. е. $H_0 : \sigma^2_M = \sigma^2_b$. При этом вычисленное значение F меньше табличного при уровне значимости а. Следовательно, гипотезу об отсутствии влияния фактора А не отклоняют.

Ход работы

Задание 1. Провести дисперсионный анализ по предложенной схеме. Задача. Изучали процент гемоглобина в крови кур разных пород. Влияет ли породность на % гемоглобина%:

Породы	повторности			
итальянские	87	92	86	91
куропатчатые	91	90	88	89
минорки	85	82	85	86
бентамы	82	82	85	84

Задание 2. Провести дисперсионный анализ по предложенной схеме. Задача. При кормлении тушканчиков получены данные о средних температурах тела. Влияет ли пол на изменчивость температуры тела?

пол	повторности			
Самки	36,9	36,8	37,0	36,6

самцы	36,7 36,7 36,8 36,6

Вопросы для самоконтроля

1 В чем состоит сущность метода дисперсионного анализа?
2 Как проводится оценка варьирования при дисперсионном анализе?

Лабораторное занятие 9 «Допущения дисперсионного анализа»

Материалы и оборудование: данные замеров статистических величин и расчетов; калькуляторы.

Цель. Определение соответствия расчетных и фактических кривых распределений диаметров и высот по критерию Колмогорова – Смирнова и Пирсона.

Пояснения к заданиям:

Критерий согласия Колмогорова-Смирнова. Один из наиболее простых и удобных при сопоставлении эмпирических совокупностей большого объема – критерий, предложенный А.Н. Колмогоровым и Н.В. Смирновым. Этот непараметрический показатель, обозначаемый греческой буквой λ (лямбда), представляет собой максимальную разность (d_{max}) между значениями накопленных частот эмпирического и вычисленного рядов (без учета знаков d), отнесенную к корню квадратному из суммы всех вариант совокупности: $\lambda = \frac{d_{\max}}{\sqrt{n}}$.

Условием применения критерия «лямбда» служит достаточное число (не менее 100) наблюдений.

Предельные значения критерия лямбда, соответствующие трем уровням доверительной вероятности – P_1 = 0,95, P_2 = 0,99 и P_3 = 0,999 – соответственно равны 1,36, 1,63 и 1,95.

Расчет критерия «лямбда» показан на примере распределения высот в 40-летнем сосняке.

Пример расчета критерия Колмогорова-Смирнова

Срединные значения классов (x)	Эмпирии-ческие частоты (p)	Теорет. вычисл. частоты (округл,) (p´)	Накопленные частоты		p-p´=d
			p	p´	

8,8	1	1	1	1	0
9,3	2	2	3	3	0
9,8	3	4	6	7	1
10,3	7	8	13	15	2
10,8	10	12	23	27	4
11,3	16	15	39	42	3
11,8	16	16	55	58	3
12,3	17	15	72	73	1
12,8	12	12	84	85	1
13,3	7	8	91	93	2
13,8	4	4	95	97	2
14,3	3	2	98	99	1
14,8	2	1	100	100	0
Сумма	100	100	-	-	-

Расчет необходимых значений показан в таблице. Максимальное значение разности p - p′ = 4, откуда $\lambda = \frac{4}{\sqrt{100}} = 0{,}4$. Полученная величина значительно меньше предельного значения лямбда (1,36) для P = 0,05. Следовательно, расхождения между эмпирическими и вычисленными частотами симметричного распределения лежат в пределах случайных колебаний, они не достоверны. На этом основании распределение высот в исследованном древостое можно считать нормальным.

Критерий согласия Пирсона. Критерии различия, при помощи которых могут быть сравнены статистические совокупности, разделяются на две группы: параметрические и непараметрические. К первой группе относятся критерии, для применения которых необходимо вычислить среднюю арифметическую, сигму или ошибки параметров (критерий Стьюдента, Фишера).

Непараметрические критерии не требуют для своего применения вычисления названных показателей, что упрощает процесс сравнения совокупностей. Критерий согласия Пирсона (или χ^2), критерий Колмогорова (или лямбда λ) относятся к непараметрическим критериям.

Оценка близости, согласованности в распределении частот, вычисленных для любого типа распределений и полученных по фактическим данным производится при помощи критерия Пирсона. Он может быть также применен как для сравнения двух вариационных рядов, так и для установления правильности выбора теоретического распределения. Он рассчитывается по формуле:

$$\chi^2 = \frac{1}{K-1}\sum\frac{(n-n^1)^2}{n^1}$$, где

χ^2 - критерий Пирсона;

К – количество классов, включая добавленные при проведении расчетов;

n – частота фактическая;

n´ - частота расчетная.

Если критерий согласия равен или больше 2, то расхождение сравниваемых рядов считается существенным, и если меньше 2, - расхождение несущественное. Более точная оценка значимости коэффициента проводится по специальным таблицам.

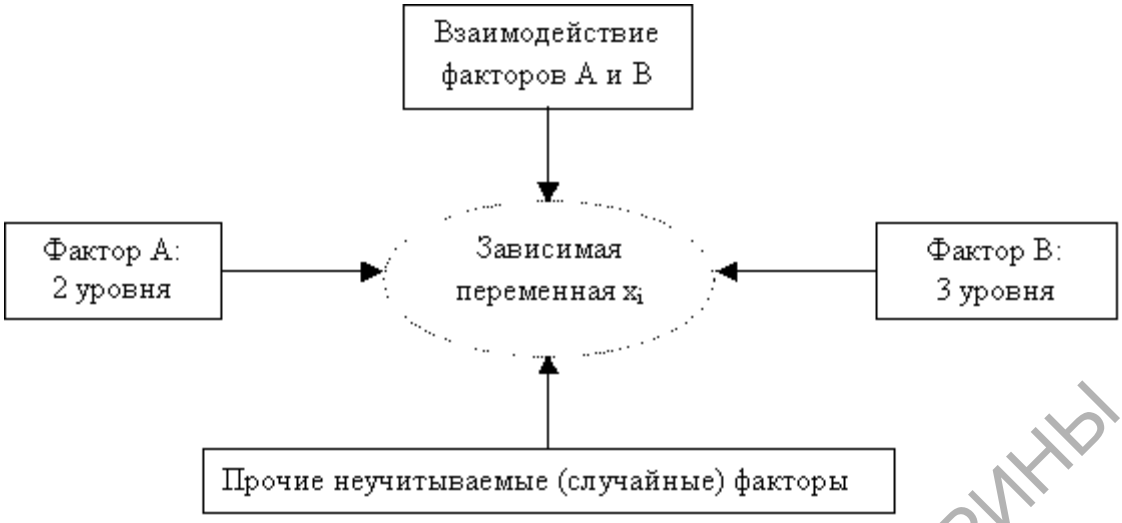
Расчет критерия согласия Пирсона оформляется в виде таблицы.

Пример расчета критерия согласия Пирсона

X	n	n´	$n-n^1$	$(n-n^1)^2$	$(n-n^1)^2$/ n´
63	0	1	-1	1	1
77	4	3	1	1	0,33
91	6	8	-2	4	0,50
105	15	16	-1	1	0,06
119	27	21	6	36	1,71
133	16	19	-3	9	0,47
147	10	12	-2	4	0,33
161	5	4	1	1	0,25
175	0	1	-1	1	1
189	2	1	1	1	1
203	0	1	-1	1	1
	Σ 85	Σ 87			Σ 4,65
χ^2 = (1/ 11-1) x 4,65 = 0,465					

Ход работы

Задание. Найти критерий согласия и провести двуфакторный анализ по схеме:



Расчеты произвести по формулам:

Двухфакторная дисперсионная модель имеет вид:

$$xijk=\mu+Fi+Gj+Iij+\varepsilon ijk,$$

где xijk - значение наблюдения в ячейке ij с номером k;

μ - общая средняя;

Fi - эффект, обусловленный влиянием i-го уровня фактора A;

Gj - эффект, обусловленный влиянием j-го уровня фактора B;

Iij - эффект, обусловленный взаимодействием двух факторов, т.е. отклонение от средней по наблюдениям в ячейке ij от суммы первых трех слагаемых;

εijk - возмущение, обусловленное вариацией переменной внутри отдельной ячейки.

Предполагается, что εijk имеет нормальный закон распределения N(0; c2), а все математические ожидания F*,G*, Ii*, I*j равны нулю.

Групповые средние находятся по формулам:

в ячейке: $\bar{X}_{ij*} = \frac{\sum_{k=1}^{n} X_{ijk}}{n}$ по столбцу: $\bar{X}_{*j*} = \frac{\sum_{i=1}^{m} \bar{X}_{ij*}}{m}$,

по строке: $\bar{X}_{i**} = \frac{\sum_{j=1}^{l} \bar{X}_{ij*}}{l}$, общая средняя: $\bar{X}_{***} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{l} \bar{X}_{ij*}}{ml}$.

Таблица– Базовая таблица дисперсионного анализа

Компоненты	Сумма квадратов	Число	Средние

дисперсии		степеней свободы	квадраты
Межгрупповая (фактор А)	$Q_1 = ln\sum_{i=1}^{m}(\bar{x}_{i**} - \bar{x}_{***})^2$	m-1	$S_1^2 = \frac{Q_1}{m-1}$
Межгрупповая (фактор В)	$Q_2 = mn\sum_{j=1}^{l}(\bar{x}_{*j*} - \bar{x}_{***})^2$	l-1	$S_2^2 = \frac{Q_2}{l-1}$
Взаимодействие	$Q_3 = n\sum_{i=1}^{m}\sum_{j=1}^{l}(\bar{x}_{ij*} - \bar{x}_{i**} - \bar{x}_{*j*} + \bar{x}_{***})^2$	(m-1)(l-1)	$S_3^2 = \frac{Q_3}{(m-1)(l-1)}$
Остаточная	$Q_4 = \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij*})^2$	mln - ml	$S_4^2 = \frac{Q_4}{mln - ml}$
Общая	$Q = \sum_{i=1}^{m}\sum_{j=1}^{l}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{***})^2$	mln - 1	

Задача. Проводились опыты по удобрению карповых прудов негашёной известью (600 кг/га) и суперфосфатом (72,8 кг/га), а также их смесью. Четвертый пруд в каждом блоке не удобрялся. Итак, в четвертом пруду продуктивность составила: 58 84 39; при применении фосфатов 72 72 64; при применении извести 49 55 48; при смешивании 74 74 85. Влияют ли Са, Р и их смеси на продуктивность пруда.

Вопросы для самоконтроля

1 Что такое параметрические и непараметрические критерии и как они вычисляются?

2 Для каких целей применяются статистические критерии?

Лабораторное занятие 10 «Корреляция. Корреляционный анализ»

Материал и оборудование: таблицы; калькуляторы.

Цель: Определение простого коэффициентов корреляции и оценка его достоверности.

Пояснения к заданиям:

Коэффициент корреляции. Отличительной особенностью биологических объектов является многообразие признаков, характеризующих каждый из них. Часто наблюдается связь между вариациями по различным признакам. В простейшем случае связь между двумя переменными величинами строго однозначна. Например, вес образцов, сделанных из одного и того же материала, определяется их объемом. Такого рода зависимость принято называть ***функциональной.*** Для биологических объектов связь обычно бывает менее жесткой: объекты с одинаковым значением одного признака имеют, как правило, разные значения по другим признакам. Такую связь между вариациями разных признаков называют ***корреляцией*** (дословный перевод: соотношение) между признаками. Заполняется рабочая таблица

X	У	X^2	$У^2$	X ·У
Σ	Σ	Σ	Σ	Σ

Затем вычисляется коэффициент корреляции, его ошибка и достоверность.

При малых объемах выборок (до 30-50) коэффициент корреляции вычисляется по формуле:

$$r = \frac{\sum x \cdot y - (\sum x \cdot \sum y)/N}{\sqrt{\left[\sum x^2 - (\sum x)^2/N\right] \cdot \left[\sum y^2 - (\sum y)^2/N\right]}}$$

Достоверность коэффициента корреляции можно оценить по формуле $t = (0{,}5 \cdot \ln(^{1+r}/_{1-r})) \cdot \sqrt{N-3}$, где t – критерий Стьюдента при числе свободы ν =N-2. Можно применить метод Z, определить критические значения r по таблице 7 (Рокицкий).

По результатам расчетов делается вывод о характере связи:

- связь между признаками прямая (r>0) или обратная (r< 0);
- теснота связи близка к функциональной r =1.0;
- r= 0.901-0.999 связь очень высокая;

- r= 0.701-0.900 высокая;
- r= 0.501-0.700 значительная;
- r= 0.301-0.500 слабая;
- r= 0 -0.300 отсутствует.

Затем определяется достоверность вычисленных показателей и сравниваются их значения, полученные по способу смешанных моментов и по приведенной формуле для малой выборки.

Ход работы:

Задание 1. У окуня озера Баторино измерены длина головы *x* и длина грудного плавника *y*

x	66	61	67	73	51	59	48	47	58	44	41	54	52	47	51	45
y	38	31	36	43	29	33	28	25	36	26	21	30	28	27	28	26

Вычислите коэффициент корреляции и определите его достоверность.

Задание 2. У 15 серебристо-черных лисиц (совхоз «Белорусский») были измерены (м см) длина туловища *x* и длина хвоста *y*.

x	70	65	66	65	71	68	64	57	66	65	67	62	67	62	63
y	40	40	40	40	40	42	39	38	41	43	39	45	43	38	40

Вычислите коэффициент корреляции и определите его достоверность.

Вопросы для самоконтроля

1 Что такое коэффициент корреляции и как он вычисляется?
2 Метод Z?
3 В чем преимущество числа z перед коэффициентом корреляции r? Можно ли переводить r в Z и обратно?
4 Какова сфера применения корреляционного анализа в биологических исследованиях?

Лабораторное занятие 11 «Линейный регрессионный анализ»

Материалы и оборудование: таблицы; калькуляторы.

Цель. Определение величины коэффициентов уравнений регрессии методом наименьших квадратов и их статистической оценки.

Пояснения к заданиям:

Расчет линии регрессии. Регрессионный анализ предполагает аналитическое выражение вероятностной связи между признаками уравнениями различного вида.

Регрессионные модели обычно используют для выражения разного рода связей в лесной таксации, лесоводстве и в других лесных дисциплинах. Чаще всего они применяются для нахождения общей зависимости по экспериментальным данным. Выведенное уравнение сглаживает (выравнивает) полученные (экспериментальные) данные. В этом случае сохраняется главная тенденция изменения функции в зависимости от изменения аргументов, и устраняются случайные отклонения.

По форме различают линейную регрессию и не линейную. По направлению связи различают прямую т.е. с увеличением признака х увеличивается признак у и обратную т.е. с увеличением х уменьшается у. Наиболее точная оценка принадлежности к виду связи производится с помощью метода наименьших квадратов (МНК). При МНКmin сумма квадратов отклонений эмпирических значений у от теоретических полученных по выбранному уравнению регрессии стремится к минимуму.

Простейшей теоретической линией регрессии является прямая линия, или парабола первого порядка, которая имеет вид:

$$Y' = a_0 + a_1 \cdot x,$$

где Y' - теоретические значения функции или зависимой переменной; х – аргумент или независимая переменная; a_0, a_1, -коэффициенты уравнения, имеющие различное значение в зависимости от специфики изучаемого явления; У –эмпирические значения зависимой переменной.

Коэффициенты определяются по формулам:

$$a_0 = \overline{Y} - a_1 \cdot \overline{X} \qquad a_1 = \Sigma(x-\overline{X})\cdot(y-\overline{Y}) \Big/ \Sigma(x-\overline{X})^2,$$

где $\overline{X}$ и $\overline{Y}$ средние арифметические рядов аргументов (Х) и функции (У).

Ход вычислений приведен в таблице.

х	у	$x-\overline{X}$	$(x-\overline{X})^2$	$(y-\overline{Y})$	$(x-\overline{X})\cdot(y-\overline{Y})$	$(y-\overline{Y})^2$	У’
Σ	Σ	Σ	Σ		Σ	Σ	

В качестве исходных данных используются вариационные ряды частичной совокупности или средние значения высот и диаметров при вычислении смешанных моментов.

Если известны среднеквадратические отклонения для рядов x и y и найден коэффициент корреляции между ними (см. вычисление смешанных моментов), то величина a_1 вычисляется по формуле: $a_1 = r_{xy} \cdot (\sigma_y / \sigma_x)$, где σ_y , σ_x - дисперсии выборочных (или усредненных) рядов; r_{xy} - мера корреляционной взаимосвязи.

Точки пересечения с осями ординат и абсцисс равны соответственно:

$y = a_0$, $x = -(a_0/a_1)$.

Оценка регрессионных уравнений. Поскольку в определении линий регрессии участвуют несколько параметров, то необходимо оценить пределы изменчивости каждого из них.

Наиболее вероятная область расположения линии прямой регрессии по отношению к оси абсцисс определяется величиной коэффициента a_1 и тангенсом угла, геометрическим смыслом которого является коэффициент корреляции. При отсутствии регрессии r=0, и тогда линия регрессии y по x располагается горизонтально по отношению к оси абсцисс, а линия регрессии x по y - вертикально. Место их пересечения соответствует средним значениям обоих признаков.

Второй коэффициент определяет величину отрезка, отсекаемого на оси y линией регрессии. Величина его определяет границы колеблемости регрессии по ординате, которая расширяется в обе стороны от средней точки ($\bar{x}$, $\bar{y}$).

Поскольку опытные данные всегда имеют определенную величину изменчивости, то и все показатели в том числе и уравнения регрессии определяются с некоторой степенью достоверности.

Определение величины ошибки найденных уравнений и оценка достоверности полученных коэффициентов уравнения прямой проводится по формулам:

$m_{y \cdot x} = \sqrt{\Sigma(y - y')^2 / (N - n)}$, или $\sigma_{y \cdot x} = \sqrt{\dfrac{\Sigma(y_i - y_i')^2}{n - 2}}$ где $\sigma_{y \cdot x}$, $m_{y \cdot x}$ - ошибка уравнения; y - эмпирические значения функции; y', - теоретические значения функции; N - число точек эмпирической линии регрессии, по которым вычислялось уравнение регрессии; n - число коэффициентов уравнения, включая свободный член.

Здесь величина $\sigma_{y \cdot x}$ имеет такое же значение как и σ в вариационном ряду. В пределах одной $\sigma_{y \cdot x}$ отклонения распределяются вверх и вниз от линии регрессии в 68% случаев. В 95% они лежат в пределах $2\sigma_{y \cdot x}$, а в 99,7% случаев отклонения от теоретической линии регрессии составляют величину $3\sigma_{y \cdot x}$. Ошибку

уравнения регрессии можно определить и по формуле: $m_{y\cdot x}=\sigma_y\sqrt{1-r^2}$, где $m_{y\cdot x}$- ошибка теоретических значений функции; σ_y- среднее квадратическое отклонение ряда y; r – коэффициент корреляции между x и y (можно использовать и корреляционное отношение при наличии криволинейности связи между признаками). Эта формула представляет упрощенный вариант вычислений и применяется для больших выборок.

В таблицах по вычислению коэффициентов уравнений в последней колонке рассчитываются теоретические значения функции. Получаем попарно разности $(y-y')$, возводим все разности в квадрат и получаем их сумму: $\Sigma(y-y')^2$. Применив формулу: $m_{y\cdot x}=\sqrt{\Sigma(y-y')^2/N-n}$, к уравнению прямой, параболы определим их ошибку.

Достоверность найденного коэффициента a_1 определяется по формуле: $t=a_1\cdot\sigma_x\cdot\sqrt{N-1}/m_{yx}$, где t – величина критерия Стьюдента, сравниваемая с критической при числе степеней свободы ν = N-2; σ_x - среднее квадратическое отклонение ряда аргументов; m_{yx}- ошибка уравнения; N- объем выборки.

Если вычисленная величина меньше табличной, то связь между x, y и значение a_1 достоверны, а если вычисленная будет больше табличной величины, то связь данных признаков и значение первого коэффициента недостоверны.

Достоверность отличия от нуля коэффициента a_0 можно оценить по формуле:

$t=\dfrac{a_0}{m_{y\cdot x}\sqrt{1/N+1/N-1\cdot\left(\bar{O}/\sigma_x\right)^2}}$, где t – величина критерия Стьюдента, сравниваемая с критической при числе степеней свободы ν = N-2; σ_x - среднее квадратическое отклонение ряда аргументов; m_{yx}- ошибка уравнения; N- объем выборки.

Ход работы

Задание 1. На основе основных положений темы и расчетов по лабораторной работе темы 12 определить взаимообусловленность признаков X и У аналитическими уравнениями.

Задание 2. Определить ошибки регрессионных уравнений и достоверность коэффициентов линейных уравнений.

Вопросы для самоконтроля

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.

Решить зачетную задачу описательной статистики по предложенной схеме:

ВАРИАНТ 1

Изучена плодовитость (число щенков) самок серебристо-черных лисиц:

4	5	3	4	6	7	8	3	1	4	6	4	4	3	2	5	3	4	5	4	5	3
4	5	4	4	4	6	5	7	6	4	5	4	4	4	4	6	2	3	4	5	5	4
4	6	4	4	4	8	7	5	4	9	4	3	4	4	5	4	6	4	4	3	4	4
4	2	4	4	5	5	4	5	3	4	6	7	8	3	1	7	9	4	4	8	2	5
5	3	1	4	5	6	1	5														

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Корреляция. Формула КК (способ прямой через центральные отклонения).

ВАРИАНТ 2

Получены данные о количестве хвостовых щитков у змей:

42	58	44	54	41	50	46	46	54	48	43	49
50	48	46	46	45	53	48	48	53	53	48	41
46	40	50	43	49	51	52	46	42	44	48	45
47	46	43	50	47	45	48	40	44	42	48	45
54	50	56	48	45	45	51	42	44	47	46	45

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Регрессионный анализ. Общее уравнение регрессии.

ВАРИАНТ 3

Имеются данные о весе кроликов (в кг):

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

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Ход дисперсионного анализа (ДА I).

ВАРИАНТ 4

Получены данные о длине листьев садовой земляники:

8,2	9,7	5,6	7,4	8,0	6,4	6,6	6,8	8,4	7,1
9,0	6,0	7,6	8,1	11,8	5,8	9,3	7,3	8,2	7,2
7,2	6,4	7,7	9,0	8,1	7,1	7,1	8,8	7,5	9,2
7,5	6,8	7,0	6,4	7,4	8,2	6,3	7,0	8,1	10,0
7,0	7,1	8,7	6,3	8,6	7,7	7,3	8,0	8,4	9,3
7,3	6,0	7,7	6,1	9,6	7,4	7,2	7,2	8,7	7,5

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Формула для расчета силы влияния учтенного фактора при дисперсионном анализе.

ВАРИАНТ 5

Было подсчитано число лучей в хвостовых плавниках камбалы:

53	51	52	55	56	49	51	51	52	54	56
54	53	52	53	51	55	53	55	53	54	51
51	56	54	54	53	54	54	55	53	51	51
52	55	53	53	56	53	56	53	52	56	52
52	56	55	50	54	49	54	55	54	55	54
52	51	55	52	55	54	51	54	53	54	55

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Теоремы сложения и умножения вероятностей.

ВАРИАНТ 6

Количество птенцов в гнездах береговой ласточки было следующим:

4	5	4	5	5	4	6	3	4	5	5	4	5	4	6	1	6	4	4	4
5	5	5	6	4	4	5	5	3	5	5	4	6	4	6	2	3	4	5	5
5	4	5	5	6	4	4	6	2	2	5	5	3	3	6	6	5	5	4	1
5	4	4	2	4	4	4	4	6	2	6	6	5	4	4	5	5	5	5	4

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, χ, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Вероятность. Математическое выражение вероятности.

ВАРИАНТ 7

Получены данные о длине правого уха (в см) серебристо-черных лисиц:

12	10	14	14	13	12	12	12	15	13	11	12	12	14	12	11	13
12	13	14	11	13	14	12	13	12	12	14	12	14	13	13	12	13
12	13	12	12	13	12	11	11	12	13	14	12	14	12	14	15	13
10	11	10	11	15	11	16	11	16	11	11	11	12	15	14	15	12

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, χ, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Система уравнений при регрессионном анализе.

ВАРИАНТ 8

Измеряли длину хвоста (в мм) у оленьих мышей:

58	57	64	61	56	65	63	58	63	60	59	61	54	58	66	67
63	63	61	60	58	57	65	61	60	68	67	64	63	56	59	64
61	64	57	60	63	58	52	60	59	57	61	54	58	64	62	59
60	63	60	60	64	59	63	63	59	62	63	61	65	61	64	57
59	54	64	63	57	59	59	58	63	62	63	62	62	60	62	57

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, χ, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Коэффициент Фишера.

ВАРИАНТ 9

Изучен живой вес телят при рождении(в кг):

27	32	32	31	32	28	37	35	26	28	32	39	34	30	37
26	27	40	35	37	28	43	26	35	45	26	35	32	32	35
35	28	32	36	32	36	37	33	26	31	36	33	33	28	23
26	34	32	36	27	32	39	30	30	36	38	24	32	30	31

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. События невозможные, достоверные и случайные. Примеры.

ВАРИАНТ 10

Были получены данные о длине коренного зуба млекопитающего (в мм):

3,2	2,8	2,9	3,0	3,1	3,3	2,9	3,1	2,7	3,4	2,9	3,0
2,9	2,8	2,6	3,0	2,8	3,0	3,1	2,9	3,0	3,2	2,7	3,0

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Корреляция. Формула для вычисления КК прямым способом через значения вариант.

ВАРИАНТ 11

Имеются следующие данные о росте (длина тела, в см) взрослых мужчин:

162	151	161	170	167	164	166	164	173	172	165
153	164	169	170	154	163	159	161	167	168	164
170	166	176	177	159	158	160	161	167	155	166
167	173	165	175	165	174	169	168	171	163	165
166	166	166	169	167	166	167	172	169	171	168
162	165	168	171	174	165	168	167	170	170	168

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Частные уравнения регрессии (У от Х и обратная связь).

ВАРИАНТ 12

Были получены данные о длине крыльев самцов (в мм) скворцов:

120	120	121	122	122	126	122	123	125	125	126
123	124	125	125	126	127	127	127	128	128	129
129	122	122	125	127	127	127	128	129	120	122

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. События совместимые и несовместимые. Примеры.

ВАРИАНТ 13

Длина тела (в мм) у плотвы озера Нарочь были следующими:

143	157	148	153	150	142	164	139	139	140	143	120
144	130	138	124	127	137	139	129	128	119	120	138
130	114	126	138	117	132	130	145	140	153	137	142
145	147	141	125	143	138	140	135	135	139	125	137
131	120	127	118	120	124	134	111	132	133	100	132
143	134	138	130	135	133	134	151	107	110	94	92

Провести анализ:

1. Составить ВР и изобразить его графически.
2. Вычислить статистические характеристики (Mo, Me, $\bar{X}$, δ, m_x, P, C_v).
3. Дать оценку достоверности (доверительный интервал при трех уровнях значимости;H_0; критерий Стьюдента).
4. Сделать обоснованный вывод. Ответ.
5. Формулы для определения сумм квадратов при дисперсионном анализе (ДА 1).

Методические рекомендации для преподавателя

Методические рекомендации для преподавателя должны указывать на средства, методы обучения, способы учебной деятельности, применение которых для освоения тех или иных тем или разделов наиболее эффективно. Рекомендации для преподавателей могут идти в русле следующих предписаний:

1. Изучив глубоко содержание учебной дисциплины, целесообразно разработать матрицу наиболее предпочтительных методов обучения и форм самостоятельной работы студентов, адекватных видам лекционных и семинарских занятий.

2. Необходимо предусмотреть развитие форм самостоятельной работы, выводя студентов к завершению изучения учебной дисциплины на её высший уровень.

3. Организуя самостоятельную работу, необходимо постоянно обучать студентов методам такой работы.

4. Пакет заданий для самостоятельной работы следует выдавать в начале семестра, определив предельные сроки их выполнения и сдачи. Задания для самостоятельной работы желательно составлять из обязательной и факультативной частей.

5. Вузовская лекция – главное звено дидактического цикла обучения. Её цель – формирование у студентов ориентировочной основы для последующего усвоения материала методом самостоятельной работы. Содержание лекции должно отвечать следующим дидактическим требованиям:

- изложение материала от простого к сложному, от известного к неизвестному;
- логичность, четкость и ясность в изложении материала;
- возможность проблемного изложения, дискуссии, диалога с целью активизации деятельности студентов;
- опора смысловой части лекции на подлинные факты, события, явления, статистические данные;
- тесная связь теоретических положений и выводов с практикой и будущей профессиональной деятельностью студентов.

Преподаватель, читающий лекционные курсы в вузе, должен знать существующие в педагогической науке и используемые на практике варианты лекций, их дидактические и воспитывающие возможности, а также их методическое место в структуре процесса обучения.

6. При проведении аттестации студентов важно всегда помнить, что систематичность, объективность, аргументированность – главные принципы, на которых основаны контроль и оценка знаний студентов. Проверка, контроль и оценка знаний студента, требуют учета его индивидуального стиля в осуществлении учебной деятельности. Знание критериев оценки знаний обязательно для преподавателя и студента.

Методические указания для студентов

Методические указания для студентов представляют собой комплекс рекомендаций и разъяснений, позволяющих студентам оптимальным образом выстроить работу по изучению дисциплины и создающих условия для успешной самостоятельной работы. Наличие методических рекомендаций особо важно для организации учебного процесса студентов - заочников.

Методические указания студентам должны раскрывать рекомендуемый режим и характер учебной работы по изучению теоретического курса (или его раздела/части), практических и/или семинарских занятий, лабораторных работ (практикумов), и практическому применению изученного материала, по выполнению заданий для самостоятельной работы, по использованию информационных технологий, выполнению курсовых работ, написанию рефератов и т.д.

Методические указания должны мотивировать студента к самостоятельной работе и не подменять учебную литературу.

Программа по организации контролируемой самостоятельной работы студентов

Согласно учебному плану на самостоятельную работу студента по дисциплине «Биометрия» на 2013/14 учебный год выделено 6 часов.

Основной вид реализации самостоятельной работы:

- конспектирование первоисточников и другой учебной литературы;
- решение домашних задач;
- проработка учебного материала (по конспектам лекций учебной и научной литературе);
- поиск и обзор научных публикаций и электронных источников на русском и иностранных языках;
- написание рефератов.

Обеспеченность образовательного процесса по дисциплине специализированным и лабораторным оборудованием

В учебном процессе для освоения дисциплины «Биометрия» используются следующие технические средства:

- Компьютеры;
- Таблицы;
- Призентации;
- Интернет-ресурсы.

Карта обеспеченности литературой

В таблице приведены сведения об обеспеченности студентов, обучающихся по дисциплине «Биометрия», учебной и учебно-методической литературой.

Наименование литературы	**Объем фонда учебной и учебно-методической литературы (количество)**	
	Учебная	**Учебно-методическая**
	Экз.	**Экз.**
1. Лакин Г.Ф. Биометрия. Учебное пособие для биол. спец. вузов. 4-е изд. Пер. и дополн. М. Высшая школа. 1990.	42	
2. Рокицкий П.Ф. Биологическая статистика. Минск. Высшая школа. 1969.	15	
3. Терентьев П.В. Практикум по биометрии. Учебное пособие по биометрии. М. 1977.	2	
4. Демьянов Ю. Э. Литвин Н.Ф. Применение математических методов и ЭВМ в биологии. Под ред. Селькова. М. Изд-во МГУ. 1981.	1	
5. PDF версия на сайте университета текста лекций по курсу «Биометрия» (составитель Кураченко И.В.)		
6. Методические указания к лабораторным занятиям по курсу «Биометрия» (составитель Кураченко И.В.)		15
7. Электронные материалы (наборы видео- и аудио- материалов, компьютерные программы «STATISTICA 6.0», «StatSoft», пакет программ «MS Excel», электронные учебники, электронный словарь статистических терминов, презентации и др.)		

ТЕКУЩАЯ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ СТУДЕНТОВ ПО ДИСЦИПЛИНЕ

Положение о рейтинговом контроле знаний

Курс “Биометрия” состоит из материала теоретического и прикладного характера, который излагается на лекциях, практически осуществляется при проведении лабораторных занятий, а также частично выносится на СУРС.

Курс завершается зачетом, сопровождаемым рейтинговыми баллами. Суммарный рейтинговый балл составляется из баллов, полученных за три промежуточных этапа, оканчивающихся контрольными работами и баллов, полученных на зачете. При вынесении семестровой оценки экзаменатор суммирует баллы трех промежуточных этапов и баллы, полученные при опросе и на основании полученного результата определяет суммарный рейтинговый балл по курсу за семестр и итоговый результат.

Цели и задачи балльно-рейтинговой аттестации студентов, обучающихся по дисциплине

Основными целями введения балльно-рейтинговой аттестации являются:

1. Стимулирование повседневной систематической работы студентов;
2. Снижение роли случайностей при сдаче зачета;
3. Повышение состязательности в учебе;
4. Исключение возможности протежирования не очень прилежных студентов;
5. Создание объективных критериев при определении кандидатов на продолжение обучения (магистратура, аспирантура и т.п.);
6. Повышение мотивации студентов к освоению профессиональных образовательных программ на базе более высокой дифференциации оценки результатов их учебной работы.

Состав и планирование в баллах рейтинговых контрольных мероприятий по дисциплине

Состав и планирование в баллах контрольных рейтинговых мероприятий, ориентировочное распределение баллов по видам отчетности в рамках дисциплины представлены в таблице.

Текущая, промежуточная и итоговая аттестация студентов, обучающихся по дисциплине «Биометрия»

Вид отчетности	1 рейтинговый контроль	2 рейтинговый контроль	3 рейтинговый контроль	Зачет
Текущий (опрос-собеседование, письменный опрос)	6	6	6	60-71
Домашние задания.	6	6	6	
Контрольные работы	9	9	9	
Коллоквиум	-	-	9	
Посещение занятий	3	3	3	
Всего	24	24	33	

График балльно-рейтинговых контрольных мероприятий по дисциплине

По дисциплине «Биометрия» предусмотрены текущая, промежуточная и итоговая формы контроля.

Текущий контроль: оформление рабочей тетради на лабораторных занятиях, включающей решение домашних задач, письменный опрос и собеседование.

Промежуточный контроль: обязательное тестирование (по три контрольные точки в каждом семестре). Для промежуточного контроля студентов выполнено по 200 тестовых заданий. Тесты оформлены в формате АСТ (имеется электронная копия в формате WORD) согласно требованиям к

аттестационным педагогическим измерительным материалам для компьютерного тестирования. Тесты размещены в электронной базе ГГУ.

Учетная документация при рейтинг-контроле по дисциплине

Нормативными документами учета успеваемости студентов, обучающихся по балльно-рейтинговой системе являются:

- ведомость учета текущей успеваемости;
- зачетная ведомость;

Ведомость текущей успеваемости заполняется преподавателем 3 раза в течение семестра.

Порядок и сдача зачета

Для допуска к зачету студент должен набрать в ходе текущего и рубежного контроля не менее 60 баллов. Для допуска к зачету необходимо выполнение всех запланированных по программе лабораторных работ независимо от числа набранных баллов по дисциплине.

Зачетные вопросы

1. Предмет и основные понятия биометрии. История биометрии.
2. Группировка данных, совокупность и вариационный ряд.
3. Совокупность, примеры различных совокупностей. Отличие выборочной совокупности от генеральной совокупности.
4. Принципы группировки данных при качественной дискретной и непрерывной изменчивости.
5. Вариационный ряд. Особенности распределения вариант в вариационном ряду. Графическое изображение вариационного ряда.
6. Статистические показатели для характеристики совокупности.
7. Размах вариационного ряда и лимиты. Мода и медиана.
8. Средняя арифметическая и ее свойства. Формулы для вычисления.
9. Варианса и среднее квадратическое отклонение.
10. Понятие степень свободы.
11. Средняя геометрическая. Формулы для ее вычисления.

12. Коэффициент вариации, его отличие от среднего квадратического отклонения.

13. Закономерности случайной вариации. Вероятность. Формулы для вычисления вероятности.

14. Нормальная вариационная кривая и ее характеристика. Нормированное отклонение.

15. Уровни значимости. Связь между уровнем значимости и вероятностью.

16. Доверительные вероятности или доверительный интервал.

17. Оценка достоверности статистических показателей. Выборочные и генеральные совокупности.

18. Средние ошибки, ошибки выборочности. Формулы вычисления.

19. Критерий Стьюдента, случаи и примеры его использования.

20. Нулевая гипотеза. Сущность нулевой гипотезы.

21. Формулы для определения необходимого объема выборочной совокупности. Охарактеризуйте основные предпосылки выборочного метода.

22. Измерение связи. Корреляция. Понятие о корреляции. Положительная и отрицательная корреляция.

23. Коэффициент корреляции. Формулы для его вычисления.

24. Выборочность коэффициента корреляции. Оценка его достоверности.

25. Понятие о регрессии. Односторонняя и двусторонняя регрессия.

26. Коэффициент регрессии. Ошибка коэффициента регрессии и его достоверность.

27. Статистический анализ вариации по качественным признакам.

28. Альтернативная вариация. Средняя арифметическая и среднее квадратическое отклонение при альтернативной вариации.

29. Средняя ошибка при альтернативной вариации. Доверительные границы для доли.

30. Дисперсионный анализ. Сущность дисперсионного анализа.

31. Общая схема дисперсионного анализа при однофакторном опыте.

32. Установление достоверности влияния изучаемого фактора. Фактические и табличные значения F.

33. Изучение степени соответствия фактических данных теоретически ожидаемым.

34. Критерий соответствия хи-квадрат. Формулы для его вычисления.

35. Закономерности распределения $\chi 2$. Понятие вероятности и значимости в применении $\chi 2$.

36. Фактические данные и нулевая гипотеза. Области отбрасывания нулевой гипотезы.

Тесты по дисциплине «Биометрия».

1. Основы науки, названной биометрикой, в 1899 году разработал:
+: Гальтон;
-: Льюин;
-: Фишер;
-: Госсет.

2. Множество отдельных отличающихся друг от друга и в то же время сходных в некоторых отношениях объектов называется:
-: вариацией;
-: дисперсией;
+: совокупностью;
-: медианой.

3. Объемом совокупности называют:
-: различия в совокупности;
-: вариацию совокупности;
+: число единиц в совокупности;
-: дисперсию совокупности.

4. Синонимом термина «дисперсия» является:
-: количество;
-: совокупность;
-: качество;
+: вариация.

5. Вариация – это:
+: различия между единицами совокупности;
-: сходство между единицами совокупности;
-: число единиц в совокупности;
-: объем совокупности.

6. Варианта – это:
-: объем совокупности;
+: значение единицы совокупности;
-: средняя арифметическая;
-: среднее квадратическое отклонение.

7. Варианты являются числовыми значениями:
-: средней арифметической;
+: случайной переменной;
-: средней геометрической;
-: постоянной переменной.

8. Теоретически бесконечно большую или приближающуюся к бесконечности совокупность называют:
-: выборочной;
-: постоянной;
+: генеральной;
-: варьирующей.

9. Выборочные совокупности по своим размерам являются:
-: теоретически бесконечными;
+: сравнительно небольшими;
-: включающими одну единицу;
-: приближающимися к бесконечности.

10. Совокупность животных характеризуется по масти. Такую вариацию называют:
-: количественной;
-: сходной;
+: качественной;
-: постоянной.

11. На прерывную (дискретную) и непрерывную разделяется:
+: количественная вариация;
-: ограниченная вариация;
-: качественная вариация;
-: случайная вариация.

12. Число детенышей в помете у совокупности серебристо-черных лисиц можно отнести к:
-: случайной вариации;
-: ограниченной вариации;
+: количественная вариация;
-: качественная вариация;

13. Отличие прерывной (дискретной) вариации от непрерывной заключается в следующем:
-: выражается только дробными числами
-: может выражаться как целыми, так и дробными числами;
+: выражается только целыми числами.

14. Частным случаем качественной вариации является:

-: количественная;
-: ограниченная;
-: дисперсная;
+: альтернативная.

15. В совокупности выделяют только две группы. Такая вариация называется:
+: альтернативной;
-: генеральной;
-: случайной;
-: количественной.

16. Количество вариант от 60 до 100 подразделяют на:
-: 5-6 классов;
-: 8-12 классов;
+: 7-10 классов;
-: 10-15 классов.

17. На 10 – 15 классов подразделяется:
-: 100 вариант;
-: 50 вариант;
-: 25 вариант;
+: более 200 вариант.

18. Расположение вариант от меньших величин к большим называется:
+: ранжировкой;
-: группировкой;
-: объединением;
-: слиянием.

19. Ряды, получаемые в ходе распределения вариант по классам называются:
-: переменными;
+: вариационными;
-: случайными;
-: количественными.

20. Класс, обладающий наибольшей частотой получил название:
-: вариационный;
-: запредельный;
+: модальный;
-: лимитный.

21. Модальным называется класс, обладающий:
-: наименьшей частотой;
-: включающий среднюю арифметическую;
+: наибольшей частотой.

22. Лимитами называются значения:
-: модального класса;
-: средней арифметической;
+: крайнего класса;
-: среднего квадратического отклонения.

23. Полигон распределения применятся при:
-: непрерывной вариации;
+: дискретной вариации;
-: случайной вариации;
-: постоянной вариации.

24. Кривая распределения - это:
+: графическое изображение вариационного ряда;
-: распределение вариационного ряда по классам;
-: расчет частоты встречаемости;
-: определение модального класса в вариационной ряду.

25. При построение полигона распределения на ось абсцисс наносятся:
-: частоты;
-: лимиты;
+: классы;
-: медианы.

26. При построение полигона распределения на ось ординат наносятся:
+: частоты;
-: лимиты;
-: классы;
-: медианы.

27. Классы объединяют несколько значений вариант. В этом случае наиболее подходящим является построение:
-: полигона распределения;
-: вариационной кривой;
+: гистограммы распределения;
-: кривой распределения.

28. Полигон распределения получается многовершинным в случае, если обнаруживается:
-: один модальный класс;
-: два лимита;
-: несколько медиан;
+: несколько модальных классов.

29. При изучении графического распределения, в вариационных рядах обычно наблюдается следующее:

-: частота вариант постепенно возрастает к краям вариационного ряда;
+: частота вариант постепенно убывает к краям вариационного ряда;
-: частота вариант остается неизменной.

30. Причиной многовершинности вариационных рядов не является:
-: малый объем выборки;
-: однородность биологического материала;
+: отсутствие модального класса;

31. Значение модального класса называется:
-: лимитом;
-: медианой;
+: модой;
-: пределом.

32. Величина, в биологической статистике обозначаемая *Me* называется:
-: модой;
+: медианой;
-: случайной переменной;
-: модальным классом.

33. Модальным является класс «46-48». В этом случае мода равняется:
-: 46;
+: 47;
-: 48;
-: 94.

34. Значение варианты, находящейся точно в середине ряда называется:
-: лимитом;
-: модой;
-: пределом;
+: медианой.

35. Средняя арифметическая обозначается:
-: σ;
+: $\bar{x}$;
-: x_i;
-: $\sum$.

36. Объем совокупности обозначается:
-: x_i;
+: n;
-: x_g;
-: S.

37. Сумма значений всех вариант, входящих в совокупность, разделенное на общее число вариант, будет выражать:
-: среднюю геометрическую;
-: среднее квадратическое отклонение;
-: среднюю ошибку;
+: среднюю арифметическую.

38. Вариационный ряд включает следующие значения: 31, 36, 37, 43, 48. Средняя арифметическая будет:
+: больше x_3;
-: меньше x_3
-: равна x_3.

39. Средняя арифметическая вычисляется по формуле:
+: $\bar{x} = \sum x_i / n$
-: $\bar{x} = \sum x_i \times n$
-: $\bar{x} = \sum x_i + n$
-: $\bar{x} = \sum x_i - n$

40. Синонимом термина «варианса» является:
-: средняя арифметическая;
-: средняя ошибка средней арифметической;
+: средний квадрат отклонений вариант от средней арифметической;
-: средняя геометрическая.

41. Среднее квадратическое отклонение обозначается как:
-: $\bar{x}$;
-: t;
-: n;
+: σ.

42. Сумма квадратов отклонений отдельных значений данной переменной от средней арифметической, деленной на число вариант называется:
-: медианой;
+: вариансой;
-: модой;
-: средней геометрической.

43. Число степеней свободы обозначается как:
-: $\bar{x}$;
-: S_x;
+: n – 1;
-: σ.

44. Число степеней свободы в выборке включающей 41 вариант равняется:

-: 82;
-: 42;
+: 40;
-: 41.

45.Варианса вычисляется по формуле:
+: $\sigma = \sum (x_i - \bar{x})^2 \Big/ n$
-: $\sigma = \sum (x_i - \bar{x})^2$
-: $\sigma = \left(\sum (x_i - \bar{x})^2\right) \times n$

46. Основным критерием для применения средней геометрической является:
-: возрастание данного признака путем арифметического прибавления к первоначальному значению какой-то величины;
+: возрастание данного признака путем умножения пропорционально степени;
-: убывание данного признака путем вычитания от первоначального значения какой-то величины;
-: убывание данного признака путем деления пропорционально степени.

47. Среднее квадратическое отклонение выражается в тех же единицах, что и:
-: число степеней свободы;
+: средняя арифметическая;
-: объем совокупности.

48. Коэффициент вариации обозначается:
-: σ;
-: σ^2;
+: v;
-: $\sum$.

49.Средняя геометрическая обозначается:
-: $\bar{x}_i$;
+: $\bar{x}_g$;
-: $\bar{x}_n$;
-: $\bar{x}_v$.

50. Процентное соотношение, которое составляет σ от $\bar{x}$ составляет:
+: коэффициент вариации;
-: коэффициент ассиметрии;
-: коэффициент корелляции.
-: коэффициент регрессии.

51 В случае если средняя арифметическая равна 6,8; варианса 0,8, коэффициент вараиции будет равен:

-: (6,8/0,8) х 100%;
+: (0,8/6,8) х 100%;
-: (0,8 х 6,8) х 100%;
-: (6,8 + 0,8) х 100%.

52. Взвешенная средняя арифметическая применяется для анализа:
-: альтернативной совокупности;
+: сложной совокупности, состоящей из нескольких частных;
-: выборочной совокупности;
-: постоянной совокупности.

53. Свойством средней арифметической не является:
-: отражение всей совокупности в целом;
-: обобщение характеристики данного изучаемого признака;
+: отражение минимального значения изучаемой совокупности.

54. Синонимом термина «вероятностный» является:
-: статистический;
-: постоянный;
+: стохастический;
-: определенный.

55. Число степеней свободы, которым характеризуется данная выборка равно 75. Объем выборки в этом случае равен:
-: 70;
-: 150;
-: 74;
+: 76.

56. На каждой из сторон кубика написаны цифры 1,2,3,4,5,6. Вероятность того, что наверху будет цифра 4 равна:
-: $\frac{1}{4}$;
-: 50%;
+: $\frac{1}{6}$;
-: 25%.

57. Каждое отдельное явление, взятое само по себе, представляется случайным. Но взятые в массе они обнаруживают:
-: вероятностные закономерности;
+: статистические закономерности;
-: стохастические закономерности;
-: случайные закономерности.

58. Варианса представляет собой сумму квадратов:

-: средней геометрической;
-: средней арифметической;
+: среднего отклонения от средней арифметической;
-: средней ошибки средней арифметической.

59. В данной породе за несколько последних лет обнаружено 110 комолы телят из общего количества 55000 родившихся. Вероятность рождения рогатого теленка равна:
-: 50%;
-: 0,002;
-: 0,998;
-: 0%.

60. Априорными называются вероятности:
-: известные после проведения опыта;
+: известные до проведения опыта;
-: равные сумме вероятностей до и после проведения опыта.

61. Вероятности, которые становятся известными после проведения эксперимента называются:
-: априорными;
-: стохастическими;
+: апостериорными;
-: случайными.

62. Символом F обозначается:
-: сумма квадратов отклонений;
+: частота встречаемости класса;
-: вариационный ряд;
-: средняя геометрическая.

63. При возрастание данного признака путем умножения пропорционально степени целесообразно применять:
+: среднюю геометрическую;
-: среднюю арифметическую;
-: среднюю ошибку средней арифметической;
-: средний квадрат отклонений.

64. Синонимом термина «средний квадрат отклонений вариант от средней арифметической» является;
-: ковариант;
-: регрессия;
+: варианса;
-: хи-квадрат.

65. Из перечисленных ученых проблемами биостатистики не занимался:

-: Фишер;
-: Госсет;
-: Гальтон;
-: Эйвери.

66. Апостериорными называются вероятности:
+: известные после проведения опыта;
-: известные до проведения опыта;
-: равные сумме вероятностей до и после проведения опыта.

67. Распределение вариант в виде вариационного ряда, частоты в котором соответствуют коэффициентам разложения бинома Ньютона можно наглядно показать с помощью:
-: аппарата Фишера;
+: аппарата Гальтона;
-: аппарата Паусона;
-: аппарата Госсета.

68. Треугольник из цифр, в котором цифры каждого последующего ряда получаются путем сложения двух цифр ряда, расположенного над ним называется:
+: треугольником Паскаля;
-: треугольником Ньютона;
-: треугольником Пуассона;
-: треугольником Фишера.

69. Средняя арифметическая генеральной совокупности обозначается:
-: $\bar{x}$;
+: μ;
-: x_i;
-: σ.

70. Средняя ошибка средней арифметической вычисляется по формуле:
+: $S_{\bar{x}} = \sigma / \sqrt{n}$;
-: $S_{\bar{x}} = \sigma + \sqrt{n}$;
-: $S_{\bar{x}} = \sigma \times \sqrt{n}$;
-: $S_{\bar{x}} = \sigma - \sqrt{n}$;

71. Под псевдонимом Стьюдент работал английский математик:
-: Фишер;
-: Гальтон;
-: Пирсон;
+: Госсет.

72. Нормированное отклонение обозначается:
-: S_x;

-: μ
-: x_i;
+: t.

73. Отношение численности выборочной совокупности (n) к общей численности генеральной совокупности (N) носит название:
-: коэффициент вариации;
-: нормированное отклонение;
+: доля выборки;
-: дисперсия.

74. Погрешность, которую измеряет средняя ошибка называется:
-: ошибкой точности;
+: ошибкой выборочности;
-: ошибкой вариации;
-: ошибкой дисперсии.

75.Закон больших чисел заключается в следующем:
-: чем меньше объем изучаемой выборки, тем больше разница между $\bar{x}$ и μ;
+: чем больше объем изучаемой выборки, тем меньше разница между $\bar{x}$ и μ;
-: $\bar{x}$ и μ во всех случаях одинаковы.

76. Распределение вероятности, полученное Стьюдентом получило название:
-: fx – распределение по Стьюденту;
+: t – распределение по Стьюденту;
-: σ – распределение по Стьюденту;
-: $\bar{x}$ – распределение по Стьюденту;

77.Возможные границы, в пределах которых находится средняя арифметическая генеральной совокупности получили название:
-: выборочных;
-: переменных;
-: стохастических;
+: доверительных.

78. Нулевая гипотеза основывается на следующем утверждении:
-: между данными показателями существуют значительные отличия;
-: между данными показателями существуют незначительные отличия;
+: между данными показателями различий нет.

79. Желаемая точность наблюдений вычисляется по формуле:
-: $\Delta = \bar{x} \times t$;
-: $\Delta = \sigma \times t$;
+: $\Delta = t \times S_x$;
-: $\Delta = n \times \sigma$.

80. Одним из условий правильного отбора выборки является:
-: отбор типичных образцов;
+: отбор вариант для выборки на основе случайности;
-: отбор определенных вариант;
-: отбор вариант с наибольшими значениями.

81. Случайная бесповторная выборка предполагает что:
-: взятые образцы возвращаются обратно в генеральную совокупность;
-: отбираются только типичные образцы;
+: взятые образцы не возвращаются обратно в генеральную совокупность;
-: отбираются только наибольшие и наименьшие варианты.

82. Средняя ошибка коэффициента вариации вычисляется по формуле:
+: $S_v = \upsilon / \sqrt{2n}$;
-: $S_v = \upsilon^2 \times \sigma$;
-: $S_v = \upsilon \times \sqrt{2n}$;
-: $S_v = \upsilon^2 / \sigma$.

83. Полученное среднее арифметическое является верным если:
+: фактическое нормированное отклонение больше табличного;
-: фактическое нормированное отклонение меньше табличного;
-: фактическое нормированное отклонение не отличается от табличного.

84. Правило трех сигм гласит:
+: если разница превышает свою ошибку почти в 3 раза, она достоверна с верностью 0,99;
-: если разница не превышает свою ошибку, она достоверна с верностью 0,33.
-: если разница меньше своей ошибки в 3 раза, она достоверна с верностью 0,99;

85. Функциональные зависимости свидетельствуют о том, что:
-: численному значению одной переменной величины соответствует множество значений другой переменной;
+: каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной;
-: численные значения переменных не зависят друг от друга.

86. Корреляционная связь свидетельствует о том, что:
+: численному значению одной переменной величины соответствует множество значений другой переменной;
-: каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной;
-: численные значения переменных не зависят друг от друга.

87. При положительной корреляции зависимость между признаками следующая:

-: увеличение одного признака соответственно связано с уменьшением другого;
+: увеличение одного признака соответственно связано с увеличением другого признака;
-: признаки не влияют друг на друга.

88. При отрицательной корреляции зависимость между признаками следующая:
+: увеличение одного признака соответственно связано с уменьшением другого;
-: увеличение одного признака соответственно связано с увеличением другого признака;
-: признаки не влияют друг на друга.

89.Чем больше детенышей в помете многоплодных животных тем меньший каждый из них весит. Это является примером:
+: отрицательной корреляции;
-: функциональной зависимости;
-: нулевой гипотезы;
-: положительной корреляции.

90. Нормированное отклонение t представляет собой:
+: отклонение тех или иных вариант от их средней арифметической, выраженной в долях среднего квадратического отклонения;
-: отклонение тех или иных вариант от их вариансы;
-: отклонение тех или иных вариант от их медиан, выраженное в процентном соотношении;
-: сходство тех или иных вариант, выраженное в процентном соотношении.

91. Коэффициент корреляции обозначается
-: t;
-: σ;
+: r;
-: fx.

92. Латинской буквой r в биологической статистики обозначается:
-: коэффициент ассиметрии;
-: коэффициент вариации;
-: коэффициент распределения;
+: коэффициент корреляции.

93. Коэффициент корреляции равен нулю. Это означает что:
-: вариация обоих признаков взаимосвязана;
-: имеет место отрицательная корреляция;
+: вариация обоих признаков происходит независимо;
-: имеет место положительная корреляция.

94. Пределы в которых могут изменятся коэффициенты корреляции варьируют:
+: от 0 до 1 и от 0 до -1;
-: от 0 до 100%;
-: от 0,01 до 0,99;
-: от 1 до ∞.

95.Тесная корреляция возникает когда:
-: $r \geq 0,1$;
-: $r \geq 0,5$;
+: $r \geq 0,7$;
-: $r = 0$.

96. На слабую корреляционную связь указывает значение коэффициента корреляции:
+: ниже 0,5;
-: ниже 0,1;
-: больше 0,1 но меньше 0,3.
-: равное нулю.

97. Ошибка выборочности коэффициента корреляции в больших выборках вычисляется по формуле:
-: $S_r = \sum r^2$;
-: $S_r = \bar{x} / \sqrt{n}$;
+: $S_r = \frac{1-r^2}{\sqrt{n}}$;
-: $S_r = \bar{x} \times r^2$.

98. Уровни значимости, применяемые в биологии следующие:
-: -1 и +1;
+: 0,05 и 0,01;
-: 0 и 1;
-: 1 и 10.

99. Формула Бравэ применяется в случае:
-: прямого вычисления коэффициента вариации;
-: непрямого вычисления коэффициента вариации;
-: прямого вычисления коэффициента корреляции;
+: непрямого вычисления коэффициента корреляции.

100. Увеличение дозы ионизирующего облучения ведет к увеличению числа мутаций. Это является примером:
+: положительной корреляции;
-: функциональной зависимости;

-: отрицательной корреляции;
-: вероятностных событий.

101.Коэффициент корреляции для генеральной совокупности обозначается:
-: μ;
-: σ;
+: ρ;
-: α.

102. Установить возможные границы, в пределах которых находится средняя арифметичекая генеральной совокупности можно по формуле:
-: $\bar{x} - t S_{\bar{x}}$;
+: $\bar{x} - t S_{\bar{x}} \leq \mu \leq \bar{x} + t S_{\bar{x}}$;
-: $\bar{x} + t S_{\bar{x}}$;
-: $\mu = (\bar{x} - t S_{\bar{x}})(\bar{x} + t S_{\bar{x}})$.

103. множественной корреляцией обычно понимают:
-: зависимость изменения величины *y* от одновременного изменения величины *x*;,
-: зависимость изменения величины *x* от одновременного изменения величины *y*;
+: зависимость изменения величины *x* от одновременного изменения величины *y, z* и т.д;
-: независимость величин *x, y, z* между собой.

104. На каждой из сторон кубика написаны цифры 1,2,3,4,5,6. Вероятность того, что наверху будет цифра 3 равна:
-: $\frac{1}{3}$;
-: 50%;
+: $\frac{1}{6}$;
-: 25%.

105. Средняя ошибка разницы между средними арифметическими обозначается:
-: S_t;
-: S_f;
+: S_d;
-: S_σ.

106. Указывает на степень связи в вариации двух переменных величин, но не дает возможности судить о том, как количественно меняется одна величина по мере изменения другой:
-: коэффициент регрессии;
-: коэффициент вариации;

-: коэффициент распределения;
+: коэффициент корреляции.

107. Устанавливает степень связи в вариации двух переменных величин, а также дает возможность судить о том, как количественно меняется одна величина по мере изменения другой:
+: коэффициент регрессии;
-: коэффициент вариации;
-: коэффициент распределения;
-: коэффициент корреляции.

108. Регрессия может быть выражена несколькими способами, одним из которых не является:
-: построение эмпирических линий регрессии;
-: вычисление коэффициента регрессии;
-: составление уравнений регрессии;
+: построение регрессионной решетки.

109. К способам, позволяющим выразить регрессию графически относят:
+: построение эмпирических линий регрессии;
-: вычисление коэффициента регрессии;
+: составление уравнений регрессии;
-: построение регрессионной решетки.

110. Коэффициент регрессии обозначается:
-: r;
-: S_d;
+: R;
-: S_x.

111. Для вычисления коэффициента регрессии используются следующие формулы:
+: $R_{x/y} = r \times \sigma_x/\sigma_y$;
-: $R_{x/y} = r + \sigma_x/\sigma_y$;
+: $R_{y/x} = r \times \sigma_y/\sigma_x$;
-: $R_{y/x} = r + \sigma_y/\sigma_x$.

112. Латинской буквой R обозначается:
-: коэффициент вариации;
-: коэффициент ассиметрии;
+: коэффициент регрессии;
-: коэффициент корреляции.

113. Односторонней регрессией называется случай, когда:
-: значения двух изучаемых признаков являются строго фиксированными;
-: свободно варьируют два изучаемых признака;

-: определенно варьирует один из двух изучаемых признаков;
+: свободно варьирует один из изучаемых признаков, значения же второго признака являются строго фиксированными;

114. Двусторонней регрессией является:
+: возможность изучения изменения *x* по *y*, и изменение *y* по *x*;
-: возможность изучения изменения *x* по изменению коэффициента корреляции;
+: возможность изучения изменения *z* по *y*, и изменение *y* по *z*;
-: возможность изучения изменения *y* по изменению коэффициента корреляции.

115. Коэффициент регрессии может быть вычислен, если известны:
+: сигмы обоих вариационных рядов по признакам *x* и *y*, и коэффициенты корреляции между ними;
-: средние геометрические по признакам *x* и *y*, и коэффициенты корреляции между ними;
-: средние арифметические по признакам *x* и *y*, и коэффициенты корреляции между ними;
-: коэффициенты вариации и корреляции между признаками *x* и *y*.

116. Коэффициент регрессии равен коэффициенту корреляции в случае, если:
-: $\sigma_x + \sigma_y = 1$;
-: $\sigma_x \times \sigma_y = 1$;
+: $\sigma_x/\sigma_y = 1$;
-: $\sigma_x - \sigma_y = 1$.

117. Коэффициент корреляции между живым весом поросят *y* и их возрастом *x* равен 0,5; $\sigma_x = 4{,}0$; $\sigma_y = 2{,}0$. В этом случае коэффициенты регрессии будут равны:
+: 1 и 0,25;
-: 4,0 и 2,0;
-: 0,5 и 2,5;
-: 1 и 0.

118. Ошибка коэффициента регрессии обозначается следующим образом:
+: $S_{Rx/y}$;
-: S_{Rd};
+: $S_{Ry/x}$;
-: S_{Rt}.

119. Оценка достоверности коэффициента регрессии вычисляется по формуле:
-: $t = R - S_R$;
-: $t = R \times S_R$;
-: $t = R + S_R$;

+: $t = R / S_R$;

120. Ковариация – это:
+: связующее звено между корреляционным и регрессионным анализом;
-: связующее звено между регрессионным и дисперсионным анализом;
-: связующее звено между корреляционным и дисперсионным анализом;
-: связующее звено между дисперсионным и вариационным анализом;

121. Регрессия – это:
-: соотношение численности выборочной совокупности к генеральной;
-: погрешность, которую измеряет средняя ошибка;
-: граница, в пределах которой находится генеральная совокупность;
+: метод определения связи между варьирующими признаками;

122. Коэффициент корреляции между изменением давления крови у женщин *y* и их возрастом *x* равен 0,2; $\sigma_x = 3,0$; $\sigma_y = 2,0$. В этом случае коэффициенты регрессии будут равны:
+: 0,3 и 0,13;
-: 1 и 0,5;
-: 0 и 1;
-: 0,8 и 0,7.

123. Двумя значениями выражается:
-: коэффициент вариации;
-: коэффициент ассиметрии;
+: коэффициент регрессии;
-: коэффициент корреляции.

124. Путем ежедневного взятия проб с поля было изучено изменение высоты растений сои *y* с их возрастом *x*. Для установления степени вариации двух переменных величин, а также определения как количественно меняется один признак по мере изменения другого вычисляют:
-: долю выборки;
+: коэффициент регрессии;
-: доверительные границы;
-: промежуточный интервал.

125. Количественно установить изменение одной величины при изменении другой на единицу можно с помощью:
-: вариационного метода анализа;
+: регрессионного метода анализа;
-: корреляционного метода анализа;
-: установления промежуточного интервала.

126. Основателем биометрики является:
+: Гальтон;

-: Фишер;
-: Стьюдент;
-: Рокицкий.

127. Отбрасывание нулевой гипотезы происходит, когда:
+: нет различий между фактическими и теоретически ожидаемыми результатами.
-: степень различий между фактически полученными и исчисленными теоретическими данными ≥ 0,5;
-: степень различий между фактически полученными и исчисленными теоретическими данными ≤ 0,5;
-: различия между фактическими и теоретически ожидаемыми результатами значительны.

128. Бóльшим объемом обладает:
+: генеральная совокупность;
-: выбороная совокупность;
+: теоретически бесконечная совокупность;
-: популяция.

129. Корреляционный и регрессионный коэффициенты можно связать, используя метод:
-: дисперсии;
+: ковариации;
-: хи-квадрата;
-: критерия Стьюдента.

130. Примером положительной корреляции является:
+: увеличение числа хромосомных мутаций при увеличении дозы радиоактивного излучения;
-: потеря веса подопытного животного по причине заболевания неизвестной болезнью;
-: уменьшение массы детенышей, при увеличении их численности в помете;
-: снижение плодовитости самки, связанное с возрастными изменениями.

131. Дисперсионный анализ позволяет:
+: установить роль отдельных факторов в изменчивости того или иного признака;
-: установить промежуточный интервал между классами;
-: вычислить доверительные границы генеральной совокупности;
-: вычислить объем выборочной совокупности.

132. Методы дисперсионного анализа были разработаны английским математиком и биологом:
-: Пирсоном;
-: Госсетом;

-: Стьюдентом;
+: Фишером.

133. Дисперсионный анализ может различаться:
+: по характеру градаций внутри факторов;
-: по доле выборки;
+: по числу анализируемых факторов;
-: по доверительным границам.

134. Нулевая гипотеза предполагает:
-: значительное влияние фактора А на фактор В;
-: незначительное влияние фактора А на фактор В;
+: данный фактор А не влияет на фактор В.

135. Однофакторными, двуфакторными, трехфакторными бывают:
-: метод регрессии;
-: генеральная совокупность.
-: ковариация
+: дисперсионный анализ;

136. Для проведения дисперсионного анализа необходимо вычислить:
-: ковариянту;
+: сумма квадратов отклонений от средней арифметической;
-: среднюю геометрическую;
-: коэффициент регрессии.

137. Число степеней свободы обозначается следующим образом:
-: Sd;
+: df;
-: N;
-: x_i.

138. Градацией фактора называют:
+: несколько значений изучаемого в эксперименте фактора А;
-: изменение фактора А относительно фактора В;
+: несколько значений изучаемого в эксперименте фактора В;
-: изменение фактора В относительно фактора А.

139. Иерархическими моделями называются:
-: расположение уровней одного фактора случайным образом среди уровней другого фактора;
-: отсутствие строгой закономерности при расположении уровней одного фактора, относительно другого;
+: ступенчатое расположение уровней одного фактора, относительно уровней другого фактора.

140. Установить влияют ли данные факторы на изменчивость признака или нет и какие из них имеют больший удельный вес в общей изменчивости позволяет:
-: методы регрессионного анализа;
-: методы ковариационного анализа;
+: методы дисперсионного анализа;
-: методы корреляционного анализа;

141. При проведении дисперсионного анализа, обычно разные уровни принято обозначать буквой *i*, а отдельные варианты:
-: A;
+: j;
-: r;
-: S_x.

142. Разделение общей суммы квадратов на 4 компонента (вариация под влиянием фактора А, вариация под влиянием фактора В, вариация под совместным влиянием А и В, случайные отклонения) применяется при проведении:
-: однофакторного дисперсионного анализа;
+: двухфакторного дисперсионного анализа;
-: трехфакторного дисперсионного анализа.

143. В дисперсионном анализе общая сумма вариант по каждой изучаемой группе обозначается как:
+: T;
-: S;
-: R;
-: F.

144. Принятие данной гипотезы для признания ее правильности возможно в случае если:
-: фактически полученные данные значительно расходятся с теоретически ожидаемыми;
-: степень несоответствия фактических наблюдений с теоретически ожидаемым результатом $\geq 0,5$;
-: степень несоответствия фактических наблюдений с теоретически ожидаемым результатом $\leq 0,5$;
+: фактически полученные данные совпадают с теоретически ожидаемыми;

145. Критерий хи-квадрат оценивает:
+: степень соответствия фактических данных ожидаемым;
-: вариацию фактора А от взаимодействия факторов В и С.
-: степень изменчивости данного признака;
-: долю выборочной совокупности в общей численности генеральной совокупности.

146. С математической точки зрения критерий хи-квадрат означает:
-: отношение суммы значений всех вариант на общее число выборки;
-: отношение сигм обоих вариационных рядов по признакам *x* и *y*, помноженное на коэффициенты корреляции между ними;
+: сумма частных от деления квадратов отклонений фактически полученных чисел от ожидаемых на число ожидаемых.

147. Хи-квадрат обозначается следующим образом:
-: γ^2;
-: σ^2;
+: χ^2;
-: x_g.

148. Фактически полученные и теоретически ожидаемые числа полностью совпадают в том случае, если:
-: $\chi^2 = -1$;
+: $\chi^2 = 0$;
-: $\chi^2 = 1$;
-: $\chi^2 = 100\%$.

149. Значения χ^2 могут быть:
+: только положительными;
-: только отрицательными;
-: как положительными, так и отрицательными;
-: никогда не равны нулю.

150. Нулевая гипотеза в отношении χ^2 обозначает, что:
-: имеются существенные различия между фактически полученными и исчисленными теоретическими данными;
-: степень различий между фактически полученными и исчисленными теоретическими данными $\leq 0,5$;
-: степень различий между фактически полученными и исчисленными теоретическими данными $\geq 0,5$;
+: нет различий между фактически полученными и исчисленными теоретическими данными.

151. Допустимой границей вероятности в биологии является:
-: 0,07;
+: 0,05;
-: 0,03;
-: 0,001.

152. Отбрасывание нулевой гипотезы – это признание того, что:
+: различия между фактическими и теоретически ожидаемыми результатами являются значимыми;

-: степень различий между фактически полученными и исчисленными теоретическими данными ≥ 0,5;
-: степень различий между фактически полученными и исчисленными теоретическими данными ≤ 0,5;
-: различия между фактическими и теоретически ожидаемыми результатами являются незначительными.

153. χ^2 вычисляется по формуле:
-: $\chi^2 = \sum ((O - E)^2 \times E)$;
+: $\chi^2 = \sum ((O - E)^2 / E)$;
-: $\chi^2 = \sum (O - E)^2 + E$;
-: $\chi^2 = \sum (O - E)^2 - E$.

154. Если отбрасывание нулевой гипотезы производится при p = 0,01, то шанс на ошибку равен:
-: 0,01 из 100;
-: 0,1 из 100;
+: 1 из 100;
-: 10 из 100.

155. Бóльшим основанием для отбрасывания нулевой гипотезы является:
-: если фактически полученное значение χ^2 превышает табличное в графе вероятности 0,99;
-: если фактически полученное значение χ^2 превышает табличное в графе вероятности 0,1;
-: если фактически полученное значение χ^2 превышает табличное в графе вероятности 0,05;
+: если фактически полученное значение χ^2 превышает табличное в графе вероятности 0,01;

156. В биологических исследованиях принято отбрасывать нулевую гипотезу (при df = 1) когда χ^2 превышает 3,841, (при df = 2 когда χ^2 превышает 6,000, (при df = 3) когда χ^2 превышает 7,82. Значения же χ^2 превышающего эти величины составляют:
+: область отбрасывания нулевой гипотезы;
-: доверительные границы нулевой гипотезы;
-: промежуточный интервал нулевой гипотезы;
-: полигон распределения нулевой гипотезы.

157. Число степеней свободы при вычислении χ^2 обозначает:
+: общее число величин, по которым вычисляются соответствующие показатели, минус число тех условий, которые связывают эти величины;
-: объем выборочной совокупности минус 1;
-: общее число величин, по которым вычисляются соответствующие показатели, плюс число тех условий, которые связывают эти величины;
-: объем генеральной совокупности минус объем выборочной совокупности.

158. Поправка на непрерывность Йетса применяется при вычислении:
-: коэффициента регрессии;
-: приведении двухфакторного дисперсионного анализа;
+: вычислении χ^2;
-: вычислении коэффициента корреляции.

159. Пуассоново распределение применяется к событиям обладающим:
-: очень большой вероятностью;
-: вероятность равной 0,5;
+: очень малой вероятностью.

160. Таблицами сопряженности называются таблицы в которых должно быть:
+: распределение вариант по 2 признакам, связь между которыми нужно установить;
-: распределение вариант строго в ранжированном виде;
-: распределение вариант по частоте встречаемости;
-: распределение вариант по значению коэффициента корреляции.

161. Наименьшая существенная разность в абсолютных цифрах выражается по формуле:
-: $НСР_{05(01)} = (t_{05(01)} + S_d)$;
+: $НСР_{05(01)} = (t_{05(01)} \times S_d)$;
-: $НСР_{05(01)} = (t_{05(01)} - S_d)$;
-: +: $НСР_{05(01)} = (t_{05(01)} \times S_d) \times 100\%$.

162. Общее число наблюдений вычисляется по формуле:
+ $N = e \times n$;
-: $N = n - 1$;
-: $N = \sigma^2 / \bar{x}$;
-: $N = \sum fx / n$.

163. Корректирующий фактор вычисляется по формуле:
+: $C = (\sum x^2) / N$;
-: $C = (\sum \sigma^2) / N$;
-: $C = (\sum t^2) / N$;
-: $C = (\sum S_x) / N$.

164. Вероятность суммируется по формуле:
-: $\sum p^2 + \sum q^2 = 1$;
-: $p^2 + q^2 = 1$;
+: $p + q = 1$;
-: $p^2 + 2pq + q^2 = 1$.

165. На первом этапе дисперсионного анализа проводится:
-: суммирование всех значений вариант изучаемого признака;

-: определение коэффициента корреляции для каждого изучаемого признака;
+: разложение общей вариации изучаемого признака на варьирование вариантов, повторения и случайные отклонения;
-: вычисление суммы квадратов отклонений для вариантов и распределение на компоненты, соответствующие источником варьирования.

166. На втором этапе дисперсионного анализа проводится:
-: суммирование всех значений вариант изучаемого признака;
-: определение коэффициента корреляции для каждого изучаемого признака;
-: разложение общей вариации изучаемого признака на варьирование вариантов, повторения и случайные отклонения;
+: вычисление суммы квадратов отклонений для вариантов и распределение на компоненты, соответствующие источником варьирования.

167. Двумерное графическое изображение зависимости между двумя или несколькими переменными называется:
-: таблицей сопряженности;
+: кривой распределения;
-: корреляционной решеткой;
-: многопольной таблицей;

168. Переменная, значения которой не определяются экспериментатором называется:
+: независимая;
-: корреляционная;
-: дисперсионная;
-: зависимая.

169. Величину, которую можно измерить, контролировать и изменять в исследованиях называют:
-: ковариант;
-: градация;
-: дисперсия;
+: переменная.

170. Метод нахождения промежуточных значений некоторой величины по известному дискретному набору значений называется:
+: интерполяция;
-: дисперсия;
-: ковариация;
-: экстраполяция.

171. Метод, позволяющий определить приближенное значение функции в точках вне некоторого отрезка, по имеющимся значениям внутри этого отреза, т.е. позволяющий «продлить» функцию, называется:
-: интерполяция;

-: дисперсия;
-: ковариация;
+: экстраполяция.

172. Мера линейной зависимости двух величин называется:
-: интерполяция;
-: дисперсия;
+: ковариация;
-: экстраполяция.

173. Две группы, в одной из который имеется данный признак, а в другой он отсутствует является примером:
-: количественной вариации;
-: полигона распределения;
+: альтернативной вариации;
-: пуассонова распределения.

174. Вероятность вычисляется по формуле:
+: $p = \frac{m}{n}$
-: $p = \sum\sigma^2 / n$;
-: $p = t \times S_{\bar{x}}$;
+: $p = 1 - q$.

175. Метод Ван-дер-Вардена позволяет вычислить одним из способов:
-: объем генеральной совокупности;
-: хи-квадрат;
+: среднюю ошибку доли;
-: регрессию.

176. Расчет необходимой численности выборочной совокупности при альтернативной вариации осуществляется по формуле:
+: $n = t^2 [p(1-p)/\Delta^2]$;
-: $n = 1 + N$;
-: $n = \sum fx / \bar{x}$;
-: $n = (t^2 \times \sigma^2)/ \Delta^2$.

177. Расчет необходимой численности выборочной совокупности при количественной вариации осуществляется по формуле:
-: $n = t^2 [p(1-p)/\Delta^2]$;
-: $n = 1 + N$;
-: $n = \sum fx / \bar{x}$;
+: $n = (t^2 \times \sigma^2)/ \Delta^2$.

178. Синонимом термина «критерий согласия» является:
-: коэффициент корреляции;

+: хи – квадрат;
-: дисперсионный анализ.
-: коэффициент регрессии;

179. В биологической статистике латинской буквой N обозначается:
-: вероятность;
+: объем генеральной совокупности;
-: средняя ошибка;
-: объем выборочной совокупности.

180. Фишером был разработан:
-: метод регрессионного анализа;
-: метод хи-квадрат;
+: метод дисперсионного анализа;
-: критерий соответствия.

181. Вероятность при Пуассоновом распределении вычисляется по формуле:
+: $p = \frac{\lambda}{n}$;
-: p = 1 – q;
-: $p = \frac{m}{n}$;
-: p = λ + n.

182. При дисперсионном анализе к разным типам варьирования не относят:
+: варьирование общих средних $\bar{x}$;
-: варьирование вариант x_{ij} внутри каждой группы вокруг каждой групповой средней $\bar{x}_i$;
-: варьирование групповых средних $\bar{x}_i$;
-: общее варьирование всех вариант x_{ij}, независимо от того, в какой группе они находятся, вокруг общей средней $\bar{x}$.

183. Распределение общей суммы квадратов на группы, включающие: эффект факторов A,B,c; взаимодействие факторов A и B, A и C, B и C, и A,B,C вместе, а также на случайные отклонения применяется при:
-: расчете χ^2;
-: двухфакторном дисперсионном анализе;
-: определении коэффициента регрессии;
+: трехфакторном дисперсионном анализе.

184. Показателем вариационного ряда, которому соответствует доля при количественной вариации является:
-: коэффициент корреляции;
+: среднее арифметическое;
-: коэффициент регрессии;
-: объем выборки.

185. Ошибка для абсолютных численностей групп вычисляется по формуле:
+: $S_p = \sqrt{\frac{p(n-p)}{n}}$;
-: $S_p = \sqrt{p+q}$;
-: $S_p = \sqrt{\sum fx/n}$;
-: $S_p = \sqrt{n-1}$.

186. Возможные пределы, в которых находятся значение доли для генеральной совокупности P определяемые по формуле $p - ts_p < P < p + ts_p$, называются:
-: промежуточными интервалами;
-: областью отбрасывания нулевой гипотезы;
-: экстраполяцией;
+: доверительными границами.

187. Средняя ошибка разницы между средними арифметическими $\bar{x}_1$ и $\bar{x}_2$ вычисляется по формуле:
+: $Sd = \sqrt{S_{x_1}^2 + S_{x2}^2}$
-: $Sd = \sqrt{S_{x_1} + S_{x2}}$
-: $Sd = \sqrt{S_{x_1}^2 - S_{x2}^2}$
-: $Sd = \sqrt{S_{x_1} - S_{x2}}$

188. По мере увеличения разницы между фактическими числами и ожидаемыми величинами χ^2 будет:
-: уменьшаться пропорционально степени;
-: убывать;
-: не изменится;
+: возрастать.

189. По формуле $\sum \frac{(O-E)^2}{E}$ вычисляется:
-: коэффициент корреляции;
-: средняя ошибка средней арифметической;
+: хи-квадрат;
-: ваианса.

190. Из перечисленных величин табличные значения имеют:
+: критерий Стьюдента;
-: коэффициент регрессии;
-: число степеней свободы;

+: хи-квадрат.

191. Среднее квадратическое отклонение выражается символом:
-: p_x;
-: N;
+: σ;
-: S_d.

192. Символами n-1 и df обозначаются:
-: коэффициент ассиметрии;
-: ковариант;
+: число степеней свободы;
-: объем выборки.

193. Вероятность появления события выражается символом:
+: p;
-: q;
-: n;
-: f.

194. Символом *υ* обозначается:
+: коэффициент вариации;
-: коэффициент корреляции;
-: коэффициент регрессии;
-: коэффициент ассиметрии.

195. Вероятность непоявления события выражается символом:
-: p;
+: q;
-: n;
-: f.

196. Средняя арифметическая для подгрупп внутри градаций по А и В при дисперсионном анализе выражается:
+: $\bar{x}_{ij}$;
-: $\bar{x}_g$;
-: $\bar{x}_n$;
-: x_i.

197. Уровень значимости обозначается символом:
-: N;
+: P;
-: T;
-: S.

198. Сумма квадратов отклонений обозначается символом:

-: fx;
-: df;
+: ss;
-: ms.

199. Частота классов обозначается символом:
-: x_i;
+: f;
-: p;
-: S_d.

200. Варианса или средний квадрат при дисперсионном анализе обозначается:
+: ms;
-: fx;
-: df;
-: pq.

Учреждение образования
«Гомельский государственный университет имени Франциска Скорины»

УТВЕРЖДАЮ
Проректор по учебной работе
УО «ГГУ им. Ф. Скорины», профессор

_____ И.В. Семченко
(подпись)
____ _____ 20__ г.
(дата утверждения)
Регистрационный № УД-_____/уч.

БИОМЕТРИЯ

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности
1-31 01 01 - 02 Биология (научно-педагогическая деятельность)

2015

Учебная программа составлена на основе образовательного стандарта Республики Беларусь ОСРБ ОСВО 1-31 01 01-2013 и типовой программы «Биометрия» по специальности 1-31 01 01 Биология, утвержденной Министерством образования Республики Беларусь 30 июня 2010 года, Регистрационный № ТД-G.314/тип.

Составитель:
И.В. Кураченко, старший преподаватель кафедры зоологии, физиологии и генетики УО «Гомельский государственный университет имени Франциска Скорины»

Рекомендована к утверждению: кафедрой зоологии, физиологии и генетики

протокол № 11, 04.05. 2015 г.,

Научно-методическим советом УО «Гомельский государственный университет имени Франциска Скорины»

протокол № 7, 27.05. 2015 г.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Современная биология давно перестала быть исключительно описательной наукой. Сегодня ее существование и развитие невозможно без использования методов и подходов такой области математики как статистика. Статистика позволяет компактно описать данные, понять их структуру, провести классификацию, увидеть закономерности в хаосе случайных явлений. Игнорирование и недооценка статистической обработки и математического анализа полученного исследователем материала может свести на нет результаты многих важных опытов, привести к необоснованным и даже ошибочным выводам. Умелое применение статистических методов позволяет объективно оценивать результаты массовых наблюдений, выявлять скрытые закономерности, правильно трактовать их, что в конечном итоге делает биологию точной наукой. В связи этим дисциплина «Биометрия» является обязательной при подготовке специалистов биологического профиля.

Цель дисциплины – сформировать у студентов целостную систему знаний о современных подходах статистического анализа данных.

В задачи дисциплины входит освоение методов, позволяющих выявлять количественные закономерности в биологических явлениях; ознакомление с принципами построения математических моделей биологических явлений и процессов; формирование навыков и умений компьютерной обработки экспериментальных данных; ознакомление с правилами корректного представления результатов исследований; формирование способности к критическому анализу представляемых в публикациях данных.

В результате изучения дисциплины обучаемый должен:

знать:

– классификацию основных методов статистического анализа биологических данных;

– способы описания центральной тенденции и разброса в совокупностях, подчиняющихся различным законам распределения;

– условия применения параметрических и непараметрических методов анализа данных;

– основные методы сравнения двух и более совокупностей;

– методы выявления связи между биологическими признаками и ограничения по их применению;

– методы анализа частот;

уметь:

– распознавать разные типы биологических данных;

– строить графические изображения вариационных рядов;

– описывать наиболее выраженные свойства анализируемой совокупности по графическому изображению вариационного ряда;

– рассчитывать основные показатели описательной статистики;

– выполнять сравнение двух и более выборок;

– выполнять анализ частот;

– выполнять корреляционный и регрессионный анализы.

В курсе подробно рассматриваются традиционные методы анализа данных. Наряду с этим большое внимание уделяется непараметрическим методам, использование которых в практике биологических исследований постоянно возрастает. На примере кластерного и дискриминантного анализов, а также метода главных компонент слушатели знакомятся с элементами многомерной статистики. Теоретические положения лекционного курса развиваются и закрепляются на лабораторных занятиях, при выполнении которых студенты приобретают навыки и умения статистической обработки данных при помощи персонального компьютера. Студенты изучают методы расчета параметров описательной статистики, построения кривых распределения и гистограмм, выполнения дисперсионного анализа и сравнения двух групп, расчета коэффициентов корреляции, анализа частот, выполнения регрессионного анализа.

Организация самостоятельной работы студентов по курсу основана на размещении в сетевом доступе комплекса учебных и учебно-методических материалов (программа, список рекомендуемой литературы и информационных ресурсов, вопросы для самоконтроля, темы лабораторных занятий и методические и информационные материалы к ним и др.).

Материал дисциплины «Биометрия» основывается на ранее полученных студентами знаний по таким курсам, как «Высшая математика», «Информатика», «Зоология» и др.

Изучение дисциплины «Биометрия» предусмотрено студентами специальности 1 – 31 01 01 02 «Биология (научно-педагогическая деятельность).

Общее количество часов для **студентов 3 курса дневной формы обучения** – 78 (5 семестр), аудиторных – 42: лекционных – 20 часов (из них управляемая самостоятельная работа – 6), лабораторных занятий – 22 часа. Форма отчетности – зачёт.

Общее количество часов для **студентов 3/4 курса заочной формы обучения** – 78, в 6 семестре аудиторных – 10 (из них лекции – 6 часов, лабораторные занятия – 4 часа. Форма отчетности – зачёт (7 семестр).

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

I. ВВЕДЕНИЕ

Биометрия как наука. Значение биометрии в исследовательской работе и профессиональной подготовке специалистов-биологов. Роль работ У. Петти, Дж. Гранта, П.-С. де Лапласа, П. Пуассона, П. Л. Чебышева, А. Кетле, К. Ф. Гаусса, Ф. Гальтона, К. Пирсона, У. Госсета, Р. Фишера и других ученых в развитии биометрии.

Понятие о наименьшей выборочной единице (единице наблюдения) и данных в биологии. Переменные (признаки). Генеральная совокупность и выборка. Количественные переменные: дискретные и непрерывные. Качественные переменные. Ранговая шкала измерений. Производные переменные: пропорции, индексы, интенсивности протекания процессов.

II. ЭЛЕМЕНТЫ ТЕОРИИ ПЛАНИРОВАНИЯ ИССЛЕДОВАНИЙ

Сплошное и выборочное обследование совокупностей. Важность случайного (рандомизированного) отбора единиц наблюдения при формировании выборок. Понятие о репрезентативной и смещенной выборках. Полностью случайный отбор и его реализация при помощи таблиц случайных чисел. Стратифицированный отбор. Систематический отбор.

III. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Группировка данных в вариационный ряд. Способы графического изображения вариационного ряда: полигон (кривая) распределения, гистограмма. Теоретические распределения случайных величин и их свойства: биномиальное распределение, распределение Пуассона, нормальное распределение. Коэффициенты асимметрии и эксцесса.

Средние величины: средняя арифметическая, взвешенная средняя, геометрическая средняя. Меры разброса единиц совокупности: дисперсия и стандартное отклонение. Коэффициент вариации.

Мода. Медиана и процентили. 25-й и 75-й процентили (квартили).

Расчет параметров описательной статистики при качественной изменчивости.

Оценка репрезентативности выборочных показателей при помощи стандартной ошибки. Центральная предельная теорема. Закон больших чисел. Определение достаточного объема выборки. Доверительные интервалы для средней арифметической и для доли.

Способы представления средних величин, мер разброса, стандартных ошибок и доверительных интервалов в научных публикациях.

IV. СТАТИСТИЧЕСКАЯ ГИПОТЕЗА

Понятие о статистической гипотезе. Нулевая и альтернативная гипотезы. Статистические критерии (тесты). Вероятность справедливости нулевой гипотезы (уровень значимости). Статистические ошибки I и II типа. Мощность критерия (теста). Понятие о параметрических и непараметрических критериях (тестах). Способы трансформации данных для приведения их к нормальному распределению: логарифмирование, извлечение квадратного корня, преобразование Бокса-Кокса, угловое преобразование.

V. ОСНОВЫ ДИСПЕРСИОННОГО АНАЛИЗА

Назначение дисперсионного анализа (ANOVA). Нулевая гипотеза при дисперсионном анализе. Расчет внутри- и межгрупповой дисперсий при однофакторном анализе с равномерным дисперсионным комплексом. *F*-критерий Фишера. Определение внутри- и межгруппового числа степеней свободы. Однофакторный дисперсионный анализ повторных измерений. Понятие о многофакторном дисперсионном анализе.

Допущения дисперсионного анализа. Проверка нормальности распределения данных: визуальный анализ гистограммы распределения, использование нормальной вероятностной бумаги, тесты Колмогорова-Смирнова и Шапиро-Уилка. Проверка равенства групповых дисперсий: тесты Бартлетта, Левене, Кохрана, *F*-тест Хартли.

Эффект множественных сравнений. Апостериорный (*post-hoc*) анализ и его методы: тесты Тюки, Нюмена-Кейлса, Шеффе, Даннета.

Непараметрические аналоги однофакторного дисперсионного анализа: *H*-тест Крускала-Уоллиса и тест Фридмана.

Сравнение двух групп. Тест Стьюдента как частный случай дисперсионного анализа. *t*-распределение. Тест Стьюдента для парных измерений. Использование доверительных интервалов для проверки гипотезы о равенстве двух средних. Введение поправки Бонферрони для *t*-критерия при проведении множественных сравнений средних. Непараметрические аналоги критерия Стьюдента: *U*-тест Манна-Уитни, тест Уилкоксона, тест Уэлча.

VI. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Понятие о функциональной и корреляционной зависимостях. Степень и направление корреляционной зависимости. Коэффициент корреляции Пирсона и оценка его статистической значимости. Коэффициент ранговой корреляции Спирмена.

VII. РЕГРЕССИОННЫЙ АНАЛИЗ

Назначение регрессионного анализа. Общий вид регрессионного уравнения. Связь коэффициента регрессии с коэффициентом корреляции. Оценка параметров регрессионного уравнения по выборке с помощью метода наименьших квадратов. Статистическая значимость регрессии. Проверка нулевой гипотезы о равенстве коэффициента регрессии нулю. Стандартные ошибки параметров регрессионного уравнения. Коэффициент детерминации. Анализ остатков. Оценка величины остаточной дисперсии с помощью F-критерия. Нахождение доверительной области для линии регрессии. Понятие о нелинейной и множественной регрессионной зависимости.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Рекомендуемый перечень лабораторных работ

Лабораторная работа 1 «Первичная и вторичная группировка экспериментальных данных»

Лабораторная работа 2 «Совокупность и вариационный ряд»

Лабораторная работа 3 «Закономерности распределений»

Лабораторная работа 4, 5 «Средние величины»

Лабораторная работа 6 «Показатели вариации»

Лабораторная работа 7 «Репрезентативность выборочных показателей»

Лабораторная работа 8 «Однофакторный дисперсионный анализ»

Лабораторная работа 9 «Корреляция, корреляционный анализ»

Лабораторная работа 10, 11 «Линейный регрессионный анализ»

Рекомендуемые формы контроля знаний

1. Контрольные работы
2. Тест-контрольная

Рекомендуемая литература

Основная

1. *Боровиков В. П., Боровиков И. П.* STATISTICA: Статистический анализ и обработка данных в среде Windows / В.П. Боровиков, И.П. Боровиков. М.: Информационно-издательский дом "Филинъ", 1998.
2. *Гланц С.* Медико-биологическая статистика. Пер. с англ. / С. Гланц. М.: Практика, 1999.
3. *Ивантер Э. В., Коросов А. В.* Основы биометрии: Введение в статистический анализ биологических явлений и процессов: Учеб. пособие / Э.В. Ивантер, А.В. Коросов. Петрозаводск: Изд-во Петрозаводского гос. ун-та, 1992.
4. *Лакин Г.Ф.* Биометрия: Учеб. пособие для биол. спец. вузов /Г.Ф. Лакин. 4-е изд., перераб. и доп. – М.: Высшая школа, 1990.

Дополнительная

5. *Плохинский Н. А.* Биометрия / Н.А. Плохинский. М.: Изд-во Моск. ун-та, 1970.
6. *Рокицкий П. Ф.* Биологическая статистика / П.Ф. Рокицкий. М.: Высшая школа, 1973.
7. *Sokal R. R., Rohlf J. F.* Biometry: the principles and practice of statistics in biological research (3rd ed.) / R. R. Sokal, J. F. Rohlf. New-York, W. H. Freeman and Company, 2001.
8. *Zar J. H.* Biostatistical analysis (2nd ed.) / J. H. Zar. Prentice-Hall, Englewood Cliffs, N. J., 1984.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ ДИСЦИПЛИНЫ «БИОМЕТРИЯ» С ДРУГИМИ ДИСЦИПЛИНАМИ СПЕЦИАЛЬНОСТИ 1-31 01 01 - 02 Биология (научно-педагогическая деятельность)

Название дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы по изучаемой учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Зоология	Зоологии, физиологии и генетики	Содержание учебной программы одобрить	Рекомендовать к утверждению учебную программу в представленном варианте Протокол №__ от “__”_____2015 г.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ

на _____/_____ учебный год

№№ пп	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры (протокол № ____ от _____ 201_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан биологического факультета
УО «ГГУ им. Ф. Скорины», д.б.н. _____ В.С. Аверин

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА ДИСЦИПЛИНЫ

дневной формы получения высшего образования

Номер раздела, темы, занятия	Название раздела, темы, занятия; перечень изучаемых вопросов	Всего часов	Количество аудиторных часов				Материальное обеспечение занятия (наглядные, методические пособия и др.)	Литература	Формы контроля знаний
			лекции	практические (семинарские) занятия	лабораторные занятия	управляемая (контролируемая) самостоятельная работа студента			
1	2	3	4	5	6	7	8	9	
1	**ВВЕДЕНИЕ.** 1.Биометрия как наука. 2. Роль работ П.-С. де Лапласа, П. Пуассона, П. Л. Чебышева, К. Ф. Гаусса, Ф. Гальтона, К. Пирсона, У. Госсета, Р. Фишера и других ученых в развитии биометрии. 3.Значение биометрии в исследовательской работе и профессиональной подготовке специалистов-биологов.	**2**	**2**				Таблицы презентация	[1], [2], [3], [4],	

1	2	3	4	5	6	7	8	9	10
2	**ЭЛЕМЕНТЫ ТЕОРИИ ПЛАНИРОВАНИЯ ИССЛЕДОВАНИЙ** 1. Сплошное и выборочное обследование совокупностей. 2. Понятие о репрезентативной и смещенной выборках. Виды отбора. 3. Группировка данных в вариационный ряд. 4.Способы графического изображения вариационного ряда: полигон (кривая) распределения, гистограмма.	**4**	**2**		**2**		Таблицы схемы	[1], [2], [3], [4], [6], [7],	Защита отчетов по лабораторным работам
	ОПИСАТЕЛЬНАЯ СТАТИСТИКА								
3	**Тема Закономерности распределений** 1. Вероятность и ее свойства. 2. Теоретические распределения случайных величин и их свойства: биномиальное распределение, распределение Пуассона, нормальное распределение. 3. Коэффициенты асимметрии и эксцесса. 4.Использование закономерностей распределения в биологии.	**4**			**2**	**2**	Таблицы	[1], [2], [3], [4], [8],	Защита отчетов по лабораторным работам

1	2	3	4	5	6	7	8	9	10
4	**Тема Средние величины** 1. Общие свойства средних величин. 2. Средняя арифметическая простая и взвешенная. Другие средние величины. 3. Использование средних величин при групповой характеристике развития признаков-приростов выходов продукции, показателей продуктивности.	**6**	**2**		**4**		Таблицы схемы	[1], [2], [3], [4], [7], [8],	Защита отчетов по лабораторным работам
5	**Тема Показатели вариации** 1 Изменчивость и разнообразие биологических объектов. 2 Показатели разнообразия: лимиты и размах. 3 Среднее квадратическое отклонение (сигма), коэффициент вариации. 4 Использование показателей разнообразия в биологических исследованиях и в практической работе с биологическими объектами.	**4**			**2**	**2**	Таблицы схемы	[2], [3], [4], [7], [8],	Защита отчетов по лабораторным работам
6	**Тема Статистический анализ вариации по качественным признакам** 1. Группировка вариант, отличающихся качественными признаками. 2. Альтернативная вариация. 3. Средние величины и их оценка при качественной вариации признака.	**4**			**2**	**2**	Таблицы схемы	[1], [2], [3], [4], [7], [8],	Защита отчетов по лабораторным работам

1	2	3	4	5	6	7	8	9	10
7	**СТАТИСТИЧЕСКАЯ ГИПОТЕЗА** 1. Метод оценки генеральных параметров по выборочным показателям. 2. Доверительные границы и доверительные интервалы выборочных параметров. 3. Критерий достоверности разницы средних, разностей между выборочной и генеральной долями.	**4**	**2**		**2**		Таблицы	[1], [2], [3], [4], [5], [7],	Защита отчетов по лабораторным работам
8	**ОСНОВЫ ДИСПЕРСИОННОГО АНАЛИЗА** 1 Теоретические основы изучения влияния различных факторов на формирование свойств у биологических объектов. 2 Основные элементы дисперсионного анализа. 3. Дисперсионный анализ на основе однофакторных комплексов.	**4**	**2**		**2**		Таблицы схемы	[1], [2], [3], [4], [5],	Защита отчетов по лабораторным работам
9	**КОРРЕЛЯЦИОННЫЙ АНАЛИЗ** 1. Понятие о функциональной и корреляционной зависимостях. 2. Степень и направление корреляционной зависимости. 3. Коэффициент корреляции Пирсона и оценка его статистической значимости.	**4**	**2**		**2**		Таблицы схемы	[1], [2], [3], [4],	Защита отчетов по лабораторным работам

1	2	3	4	5	6	7	8	9	10
	РЕГРЕССИОННЫЙ АНАЛИЗ								
9	**Тема Основы регрессионного анализа** 1. Назначение регрессионного анализа. 2. Общий вид регрессионного уравнения. 3. Связь коэффициента регрессии с коэффициентом корреляции. 4. Статистическая значимость регрессии. Выравнивание эмпирических рядов. Способы регрессионного анализа.	**6**	**2**		**4**		Таблицы схемы	[1], [2], [3], [4], [5],	Защита отчетов по лабораторным работам
Итого часов		**42**	**14**		**22**	**6**			**зачет**

Старший преподаватель И.В. Кураченко

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА ДИСЦИПЛИНЫ

заочной формы получения высшего образования

Номер раздела, темы, занятия	Название раздела, темы, занятия; перечень изучаемых вопросов	Всего часов	Количество аудиторных часов				Материальное обеспечение занятия (наглядные, методические пособия и др.)	Литература	Формы контроля знаний
			лекции	практические (семинарские) занятия	лабораторные занятия	управляемая (контролируемая) самостоятельная работа студента			
1	2	3	4	5	6	7	8	9	
1	**ВВЕДЕНИЕ.** 1.Биометрия как наука. 2. Роль работ П.-С. де Лапласа, П. Пуассона, П. Л. Чебышева, К. Ф. Гаусса, Ф. Гальтона, К. Пирсона, У. Госсета, Р. Фишера и других ученых в развитии биометрии. 3.Значение биометрии в исследовательской работе и профессиональной подготовке специалистов-биологов.	**2**	**2**				Таблицы презентация	[1], [2], [3], [4],	

1	2	3	4	5	6	7	8	9	10
2	**ЭЛЕМЕНТЫ ТЕОРИИ ПЛАНИРОВАНИЯ ИССЛЕДОВАНИЙ** 1. Сплошное и выборочное обследование совокупностей. 2. Понятие о репрезентативной и смещенной выборках. Виды отбора. 3. Группировка данных в вариационный ряд. 4.Способы графического изображения вариационного ряда: полигон (кривая) распределения, гистограмма.	Самостоятельное изучение					Таблицы схемы	[1], [2], [3], [4], [6], [7],	Защита отчетов по лабораторным работам
	ОПИСАТЕЛЬНАЯ СТАТИСТИКА								
3	**Тема Закономерности распределений** 1 Вероятность и ее свойства. 2 Теоретические распределения случайных величин и их свойства: биномиальное распределение, распределение Пуассона, нормальное распределение. 3 Коэффициенты асимметрии и эксцесса. 4 Использование закономерностей распределения в биологии.	Самостоятельное изучение					Таблицы	[1], [2], [3], [4], [8],	

4	**Тема Средние величины** 1. Общие свойства средних величин. 2. Средняя арифметическая простая и взвешенная. Другие средние величины. 3. Использование средних величин при групповой характеристике развития признаков-приростов выходов продукции, показателей продуктивности.	**4**	**2**		**2**		Таблицы схемы	[1], [2], [3], [4], [7], [8],	
5	**Тема Показатели вариации** 1 Изменчивость и разнообразие биологических объектов. 2 Показатели разнообразия: лимиты и размах. 3 Среднее квадратическое отклонение (сигма), коэффициент вариации. 4 Использование показателей разнообразия в биологических исследованиях и в практической работе с биологическими объектами.	Самостоятельное изучение					Таблицы схемы	[2], [3], [4], [7], [8],	
6	**Тема Статистический анализ вариации по качественным признакам** 1. Группировка вариант, отличающихся качественными признаками. 2. Альтернативная вариация. 3. Средние величины и их оценка при качественной вариации признака.	Самостоятельное изучение					Таблицы схемы	[1], [2], [3], [4], [7], [8],	

1	2	3	4	5	6	7	8	9	10
7	**СТАТИСТИЧЕСКАЯ ГИПОТЕЗА** 1. Метод оценки генеральных параметров по выборочным показателям. 2. Доверительные границы и доверительные интервалы выборочных параметров. 3. Критерий достоверности разницы средних, разностей между выборочной и генеральной долями.	**4**	**2**		**2**		Таблицы	[1], [2], [3], [4], [5], [7],	
	ОСНОВЫ ДИСПЕРСИОННОГО АНАЛИЗА 1 Теоретические основы изучения влияния различных факторов на формирование свойств у биологических объектов. 2 Основные элементы дисперсионного анализа. **3** 3. Дисперсионный анализ на основе однофакторных комплексов.	Самостоятельное изучение					Таблицы схемы	[1], [2], [3], [4], [5],	
8	**КОРРЕЛЯЦИОННЫЙ АНАЛИЗ** 1. Понятие о функциональной и корреляционной зависимостях. 2. Степень и направление корреляционной зависимости. 3. Коэффициент корреляции Пирсона и оценка его статистической значимости.	Самостоятельное изучение					Таблицы схемы	[1], [2], [3], [4],	

	РЕГРЕССИОННЫЙ АНАЛИЗ 1. Назначение регрессионного анализа. 2. Общий вид регрессионного уравнения. 3. Связь коэффициента регрессии с коэффициентом корреляции. 4. Статистическая значимость регрессии. Выравнивание эмпирических рядов. Способы регрессионного анализа.	Самостоятельное изучение					Таблицы схемы	[1], [2], [3], [4], [5],	
Итого часов		**14**	**6**		**4**				**зачет**

Старший преподаватель И.В. Кураченко

ПЕРЕЧЕНЬ РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Литература

Основная

1. *Рокицкий П. Ф.* Биологическая статистика / П.Ф. Рокицкий. М.: Высшая школа, 1973.− 327с.

2. *Гланц С.* Медико-биологическая статистика. Пер. с англ. / С. Гланц. М.: Практика, 1999.

3. *Ивантер Э. В., Коросов А. В.* Основы биометрии: Введение в статистический анализ биологических явлений и процессов: Учеб. пособие / Э.В. Ивантер, А.В. Коросов. Петрозаводск: Изд-во Петрозаводского гос. ун-та, 1992.

4. *Лакин Г.Ф.* Биометрия: Учеб. пособие для биол. спец. вузов /Г.Ф. Лакин. 4-е изд., перераб. и доп. – М.: Высшая школа, 1990.

Дополнительная

1. *Плохинский Н. А.* Биометрия / Н.А. Плохинский. М.: Изд-во Моск. ун-та, 1970. − 364с.

2. *Терентьев П.В*. Практикум по биометрии./ П.В. Терентьев, Н.С. Ростова Л., 1977. – 152с.

3. *Sokal R. R., Rohlf J. F.* Biometry: the principles and practice of statistics in biological research (3rd ed.) / R. R. Sokal, J. F. Rohlf. New-York, W. H. Freeman and Company, 2001.

4. *Zar J. H.* Biostatistical analysis (2nd ed.) / J. H. Zar. Prentice-Hall, Englewood Cliffs, N. J., 1984.

5. *Боровиков В. П., Боровиков И. П.* STATISTICA: Статистический анализ и обработка данных в среде Windows / В.П. Боровиков, И.П. Боровиков. М.: Информационно-издательский дом "Филинъ", 1998.

6. *Гашев С.Н.* Статистический анализ для биологов (Пакет программ «STATAN – 1996»./ С.Н. Гашев Тюмень: ТюмГУ, 1998. – 51 с.

7. *Гельман В.Я.* Медицинская информатика. / В.Я. Гельман СПб: Питер, 2002. – 480 с.