

А. Н. Голубев Л. Н. Грибина Т. С. Дьяченко



**Ситуационные задачи
по медицинской статистике
с примерами решений
в программе Microsoft Excel**

Волгоград, 2014

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Волгоградский государственный медицинский университет

А.Н. Голубев, Л.Н. Грибина, Т.С. Дьяченко

Ситуационные задачи по медицинской статистике с примерами решений в программе Microsoft Excel

*Учебно-методическое пособие к практическим занятиям
по дисциплине «Медицинская информатика»
Под общей редакцией профессора В.И. Сабанова*

**Для студентов медицинских ВУЗов, обучающихся по специальностям:
060101 65 Лечебное дело, 060103 65 Педиатрия, 060105 65 Медико-
профилактическое дело, 060201 65 Стоматология**

Волгоград
Издательство
ВолгГМУ
2014

УДК 681:614.1 (07)
ББК 73+5

Авторы:

директор медкафедрального центра специалистов ВолгГМУ, ассистент кафедры общественного здоровья и здравоохранения ВолгГМУ,
к.м.н. *А.Н. Голубев*
доцент кафедры общественного здоровья и здравоохранения ВолгГМУ,
к.м.н. *Л.Н. Грибина*
доцент кафедры общественного здоровья и здравоохранения ВолгГМУ,
к.м.н. *Т.С. Дьяченко*

Рецензенты:

Заведующий кафедрой математики и информатики ВолгГМУ,
к. ф.-м.н., доцент *З.А. Филимонова*
Заведующий кафедрой биотехнических систем и технологий ВолгГМУ,
к.т.н. *С.А. Безбородов*

Печатается по решениям ЦМС ВолгГМУ протокол № 2/20 от 25.12.2013

Голубев А.Н. , Грибина Л.Н., Дьяченко Т.С. Ситуационные задачи по медицинской статистике с примерами решений в программе Microsoft Excel. Учебно-методическое пособие к практическим занятиям по дисциплине «Медицинская информатика». Под общей редакцией профессора В.И. Сабанова. – Волгоград: Изд-во ВолгГМУ, 2014. – 188с.: ил.

Учебно-методическое пособие к практическим занятиям подготовлено на кафедре общественного здоровья и здравоохранения Волгоградского государственного медицинского университета. Издание имеет целью оптимизацию изучения студентами важнейшего раздела дисциплины «Медицинская информатика» - «Медицинская статистика». В нем приведены примеры типовых ситуационных задач, встречающихся в медицинской практике и медико-биологических исследованиях, их решение, а также задания для самостоятельной работы. Пособие предназначено для студентов медицинских вузов по направлениям подготовки 060101 65 Лечебное дело, 060103 65 Педиатрия, 060201 65 Стоматология, 060105 65 Медико-профилактическое дело.

Содержание

Введение	4
I. Основы применения программы Excel.....	5
II. Базовые понятия теории вероятностей	15
III. Организация и этапы статистического исследования.....	30
IV. Статистические таблицы.....	37
V. Относительные величины, динамические ряды.....	41
VI. Вариационные ряды, средние величины, вариабельность признака.....	55
VII. Проверка статистических гипотез, критерий Стьюдента.....	76
VIII. Критерий Хи-квадрат.....	88
IX. Метод стандартизации	94
X. Дисперсионный анализ	101
XI. Метод корреляции.....	112
XII. Метод регрессии	125
Тестовые задания	131
Приложение 1	188
Приложение 2	189
Приложение 3	191
Список сокращений	192
Основная литература	193
Дополнительная литература	193

Введение

В современных условиях медицинские работники постоянно встречаются с множеством статистических данных. В основе решения вопроса об эффективности применения нового способа лечения или профилактики заболеваний лежит статистическая обработка данных и проверка статистических гипотез. Для понимания сущности изучаемых явлений врачу необходимо ориентироваться в фундаментальных понятиях и методах статистики, знать терминологию, уметь правильно оценивать статистические критерии и показатели.

Развитие средств вычислительной техники и внедрение информационных систем в здравоохранение значительно расширили возможности статистической обработки материалов научно-практических работ. Большинство врачей получили возможность применять удобные программные средства не только для автоматизации своей основной деятельности, но и для статистической обработки данных. Владение методологией применения способов статистического анализа с помощью компьютера является основой для получения обоснованных выводов социально-гигиенических и медико-биологических исследований.

В настоящем пособии рассматриваются примеры решения типовых задач, встречающихся в различных сферах системы здравоохранения. Оно рассчитано на студентов и практических врачей, изучающих медицинскую статистику и осваивающих методы статистической обработки данных с применением компьютерной техники. Пособие может использоваться для проведения занятий со студентами медицинских вузов в рамках учебной программы «Медицинская информатика», а также при изучении дисциплины «Общественное здоровье и здравоохранение».

I. Основы применения программы Excel

Широкое распространение компьютерной техники, а также достижения информатики позволяют использовать персональные компьютеры, оснащенные прикладным программным обеспечением с целью подготовки, хранения и обработки данных медико-биологических исследований. Для решения таких задач разработаны специализированные программы статистического анализа экспериментальных данных, однако большинство из них сложны в применении либо недоступны медицинским работникам. В то же время, современные пакеты офисных программ включают основные модули статистической обработки данных, что позволяет применять их для целей сбора и анализа медицинской информации. Такими возможностями обладают программы, относящиеся к группе «Электронные таблицы» или «Табличные редакторы». Они предназначены для работы с таблицами данных и выполнения относительно несложных расчетов, например, бухгалтерских, а также изготовления рисунков деловой графики (диаграмм).

В России наибольшее распространение получил табличный редактор Microsoft Excel. Он имеет ряд существенных преимуществ, одним из которых является исключительно богатые графические возможности. Благодаря тому, что Excel входит в Microsoft Office, значительно облегчается оформление деловой документации, так как таблицы и рисунки диаграмм могут быть перенесены в текстовый редактор Microsoft Word с сохранением форматирования.

Запуск программы Excel выполняется с помощью кнопки «Пуск» из раздела «Все программы» выбором в списке меню «Microsoft Office» команды «Microsoft Office Excel». Программа открывается в стандартном окне системы Windows, и содержит в верхней части кнопки панелей инструментов, а также типовые элементы управления окном: заголовок, кнопки действий с окном, полосы прокрутки, основное меню программы и др. Рис. 1.

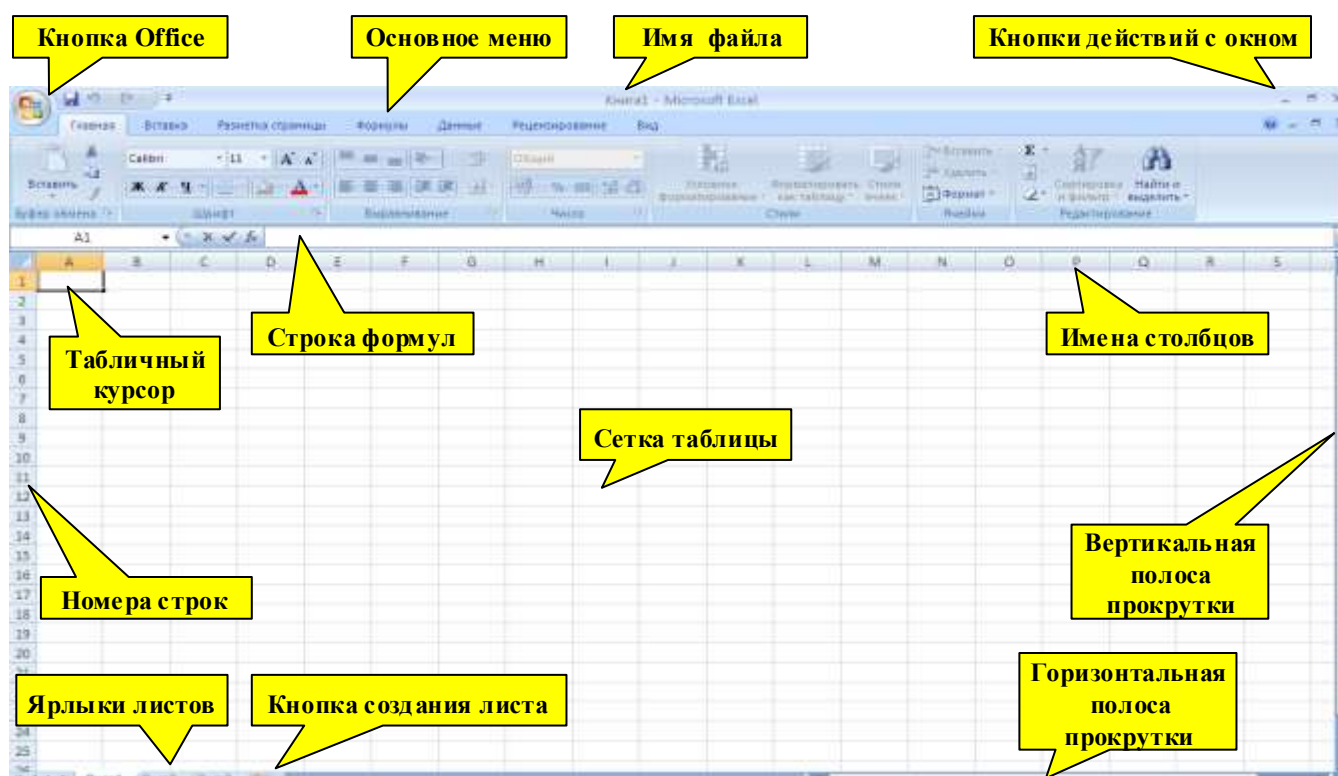


Рис. 1. Окно программы Excel.

Работа в программе Excel предъявляет высокие требования к внимательности и аккуратности пользователя. Необходимо строго соблюдать всю последовательность действий выполняемой команды, не пропуская и не добавляя лишних элементов и действий, например пробелов или дополнительных нажатий на клавиши клавиатуры и кнопки мыши.

Данные, вводимые пользователем в программе Excel, сохраняются в файле, который называется «Табличный документ». При создании нового файла этого типа именем документа, которое по умолчанию присваивается таблице и отображается в заголовке окна, является «Книга1». Такие файлы обозначаются расширением «xls» или «xlsx». Файл табличного документа может быть защищен от изменений, как в целом, так и по отдельным частям. С целью защиты части рабочей книги применяется режим «Рецензирование». Защита табличного документа с назначением пароля на его открытие выполняется методом шифрования выбором в меню кнопки «Office», режима «Подготовить», командой «Зашифровать документ».

Рабочая область окна программы Excel, в которой размещаются данные пользователя, называется «ЛИСТ». Он состоит из сетки таблицы разделенной на отдельные ячейки, которые предназначены для хранения данных. Максимальный размер каждого листа зависит от версии программы. Например, в Excel 2003 он составляет 256 столбцов и 64 тысячи строк, а для версий 2007 и 2010 – около 16 тысяч столбцов и более 1-го миллиона строк.

Файл, созданный этой программой, может содержать несколько листов. Их названия отображаются в закладках (ярлычках) расположенных в нижней части окна. Переключение к следующему листу производится щелчком левой клавиши мыши на ярлычке листа. Изменение названия листа выполняется двойным щелчком левой клавиши мыши на его корешке. Для создания нового листа служит кнопка «Вставить лист», расположенная справа от списка листов в нижней части окна.


Совокупность листов Excel составляет рабочую книгу, в которой может быть составлено несколько статистических таблиц. При этом текстовая, числовая и другие виды информация должны размещаться в отдельных ячейках.

На рабочем листе размещаются основные структурные элементы таблицы: «Столбцы», «Строки», «Ячейки».



СТОЛБЦЫ – это вертикальные составляющие экранной сетки Excel. Они обозначены буквами латинского алфавита (A, B, C, AA, AB и т.д.) в верхней части листа.

СТРОКИ – горизонтальные составляющие экранной сетки Excel. Каждая строка обозначена цифрой (1, 2, 3 и т.д.) и располагаются слева.

ЯЧЕЙКИ – места пересечения столбцов и строк в таблице. Каждая ячейка обозначена адресом, состоящим из латинской буквы столбца, и цифры номера строки. Ячейка является основной рабочей единицей таблицы Excel.

Ввод данных осуществляется в активную ячейку, обозначенную табличным курсором – «жирным» прямоугольником по ее границам. Для завершения ввода требуется нажать клавишу Enter или кнопку  «Ввод» в строке формул программы. Смена активной ячейки производится с помощью клавиш клавиатуры управления курсором (стрелок) или щелчком левой клавиши мыши.

Если длина текстовой строки или размерность числа не помещаются в установленные размеры ячейки, ширина столбцов или высота строк сетки таблицы может быть изменена. Это действие выполняется с помощью протаскивания границ между именами столбцов в верхней части окна и номерами строк слева. В каждой ячейке действуют стандартные команды изменения шрифта, выравнивания отступа и др.

Ячейки таблицы, в которых содержатся только числовые данные, могут использоваться в формулах. Для отделения целой и дробной части десятичного числа применяется знак запятой. При выполнении вычислений дробная часть числа может включать до 9 знаков после запятой. С целью уменьшения разрядности таких чисел с их округлением до требуемого количества знаков после запятой может применяться кнопка  «Уменьшить разрядность», для обратного преобразования используется кнопка  «Увеличить разрядность».

Когда при вводе числа или выводе результата вычислений после нажатия клавиши Enter в ячейке отображаются знаки решетки (####), это указывает на недостаточную ширину ячейки для отображения значения числа. Исправления такой ошибки для дробных чисел производится командой уменьшения разрядности. А также может быть выполнено увеличение ширины столбца перемещением границы между именами столбцов или уменьшение размера шрифта в ячейке.

Ввод в ячейку знаков алфавита служит для обозначения наименований и пояснений к таблицам, Адрес такой ячейки не применяется в вычислениях. Текстовая строка запоминается в одной ячейке, но может продолжаться, накрывая следующие.

Часть ячеек таблицы может быть выделена с целью их одновременного форматирования, копирования, перемещения или других групповых манипуляций. Выделение ячеек выполняется с помощью курсора мыши помещенного в центр ячейки, нажатия левой клавиши и, при ее удержании, перемещения указателя в следующую ячейку таблицы. Выделенная область ячеек обозначается серым или светло-голубым цветом и называется «Диапазоном». Он обозначается начальным и конечным адресами ячеек, разделенных двоеточием. Например, диапазон адресов B2:D6 включает 15 ячеек.

Данные из отдельной ячейки таблицы или выделенного диапазона могут быть перемещены в другую часть листа. Основным способом перемещения данных является выполнение команд «Вырезать» и «Вставить». Кроме того перемещение данных может выполняться с помощью протаскивания ячейки или выделенного диапазона курсором мыши в форме стрелочки, которая появляется при указании на контур выделения. Однако если протаскивание или вставка выполняется в ячейку содержащую данные, появляется предупреждение о замене информации.

Одной из важных команд программы Excel является способ многократного повторения данных в соседних ячейках таблицы. Такое действие называется «Автозаполнение» ячеек. Оно выполняется протаскиванием курсора мыши в форме черного крестика, который появляется при указании в правый нижний угол контура выделения ячейки (на маркер заполнения). Применение автозаполнения ячеек, включающих формулы вычислений, позволяет выполнить однотипные расчеты в сроках или столбцах таблицы без повторного набора формул с клавиатуры.

Полная информация о данных, хранящихся в активной ячейке, отображается в специальной панели инструментов программы Excel, которая называется «Строка формул».

СТРОКА ФОРМУЛ предназначена для редактирования текстовых данных, чисел и формул в активной ячейке таблицы. Она расположена в верхней части окна и показывает истинное значение числа (без округления) или формулу, помещенную в ячейку, обозначенную табличным курсором.

Одной из важнейших функций электронных таблиц является создание формул, позволяющих выполнять вычисления в таблице. Результат вычисления демонстрируется в ячейке таблицы, а формула отображается в строке формул программы. Ввод и корректировка формул может выполняться с помощью строки формул или непосредственно в ячейке таблицы. При этом формула всегда должна начинаться со знака «=» (равно). Она может включать числовые данные, адреса числовых ячеек, знаки математических операций, а также названия встроенных функций. При вводе формул выполняющих вычисления недопустимо использовать текстовые данные и несуществующие адреса ячеек.

Адрес ячейки, используемый в формуле, обозначается только латинскими буквами, без пробелов. Он может быть записан в относительном и абсолютном виде с добавлением знака денежной единицы «\$». Например, адрес ячейки D5 является относительным, а \$D\$5 – абсолютным. Преобразование относительного адреса ячейки в абсолютный выполняется вводом с клавиатуры знака \$ или клавишей F4. Важно помнить, что при выполнении команды автозаполнения относительный адрес изменяется на единицу в каждой соседней ячейке. Если адрес ячейки записан в абсолютном виде, он не изменяется при выполнении указанной команды.

Одним из способов ввода формул является возможность применения встроенных функций. Функцией в программе Excel называется алгоритм вычислений, который обозначен именем и содержит аргументы, заключенные в круглые скобки. В качестве аргументов функций могут применяться числа, адреса и диапазоны числовых ячеек. Например, для суммирования данных в выделенном диапазоне числовых ячеек таблицы применяется функция =СУММ(диапазон). С целью получения максимального значения числа в диапазоне ячеек используется функция =МАКС(диапазон), а минимального значения =МИН(диапазон). Ввод требуемой функции в ячейку таблицы выполняется с помощью указания ее имени с клавиатуры или выбором из списка предлагаемого «Мастером функций». Он позволяет получить информацию о назначении и правилах использования каждой стандартной функции.

Если формула содержит синтаксическую ошибку, программа Excel не выполнит вычислений и покажет сообщение в ячейке таблицы, где пользователь применил ошибочную формулу. Наиболее распространенными ошибками записи формул являются:

- ошибка #ЗНАЧ! указывает на использование недопустимого типа данных в формуле, например, адреса ячейки, содержащей текстовые данные. Для исправления этой ошибки требуется проверить содержимое ячеек, адреса которых записаны в формуле, и убедиться, что в них внесены числовые данные. А затем исправить формулу с применением адресов только числовых ячеек или ввести числовые данные в ячейки, адрес которых используется в формуле.

- если при вводе формулы в ячейке отображается ошибка #ИМЯ?, это указывает на использование несуществующего или неправильно записанного адреса ячейки. Например, лишний пробел в адресе ячейки (B2 – правильно, B 2 – ошибка) или знак русского алфавита (буква В английская - правильно, буква В русская - ошибка). Для исправления такой ошибки необходимо проверить правильность записи адресов ячеек в формуле и откорректировать ошибочные.
- появление ошибки #ДЕЛ/0 происходит при выполнении вычислений формулой, которая производит деление на ноль. Для исправления ошибки необходимо проверить правильность написания формулы и данные в адресах ячеек, указанных в знаменателе.

Кроме широких вычислительных возможностей программа Excel включает команды построения рисунков деловой графики, которые называются «**Диаграммами**». Диаграмма (от греч. *diagramma* - изображение, рисунок, чертеж) является средством наглядного представления данных с целью визуального сопоставления явлений, выявления закономерностей и тенденций. Диаграммы существенно информативнее числовых данных, так как они обеспечивают наглядное представление информации.

Диаграммы создаются на основе данных в сохраненных ячейках таблицы Excel. Для изготовления диаграммы выделяется область диапазон данных (или несколько диапазонов, возможно несмежные), на основе которых будет создан рисунок. Необходимо выделять все ячейки, которые хранят данные, отображаемые в диаграмме. Если выделяется несколько отдельных диапазонов, то число ячеек в них должны быть равными, то есть необходимо соблюдать симметричность выделенных областей. В случае использования нескольких диапазонов, изготовление диаграммы можно упростить путем копирования требуемых данных на отдельный лист, используя операции выделения, копирования и вставки.

Рисунок диаграммы включает следующие основные элементы: «**Ряды данных**», «**Категории**», «**Легенда**» (рис. 2).

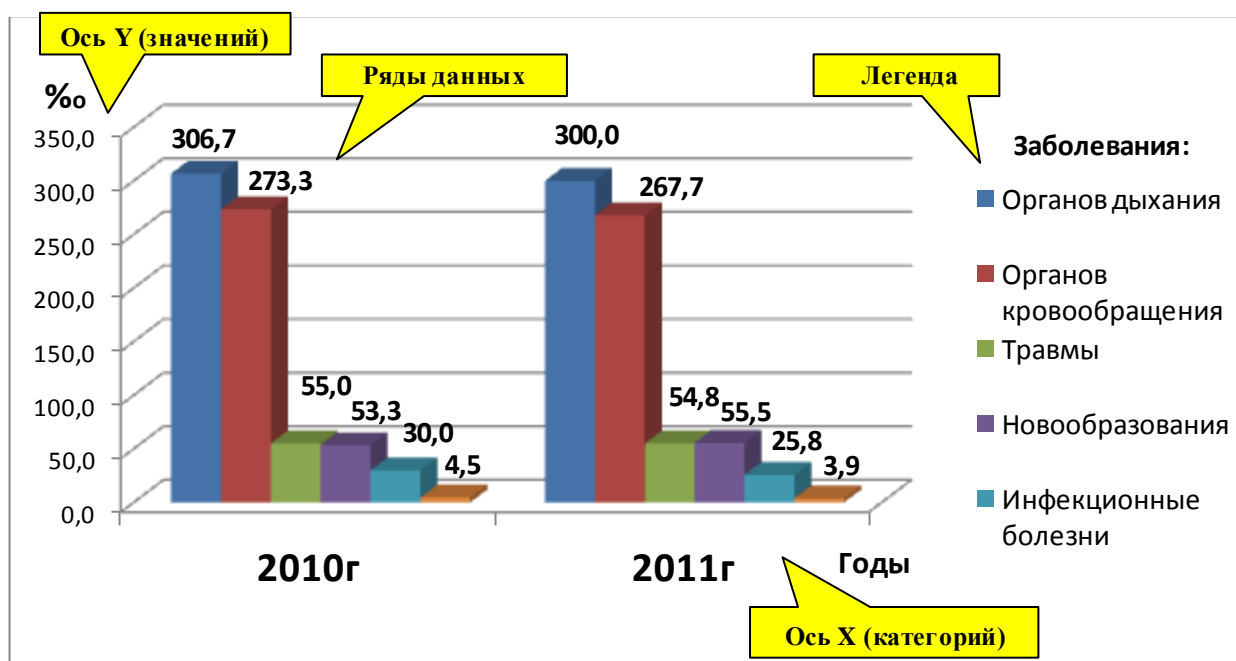


Рис. 2. Основные элементы диаграммы.

РЯД ДАННЫХ - это столбец или строка числовых значений, отображаемых на оси Y диаграммы. Каждому ряду данных соответствует отдельный цвет или способ обозначения, указанный в легенде диаграммы. Диаграммы всех типов, кроме круговой, могут содержать несколько рядов данных.

КАТЕГОРИИ - (аргументы функции на оси X) служат для упорядочения значений в рядах данных, которые изображены на диаграмме.

ЛЕГЕНДА - это условные обозначения рядов данных на диаграмме. Она содержит текстовые пояснения о цветовых или узорных обозначениях, используемых в рисунке.

Команды создания диаграммы расположены в режиме «Вставка» и требуют выбора вида рисунка. Основными видами диаграмм, которые могут создаваться пользователем в программе Excel, являются: «Гистограмма», «Линейчатая», «График», «Круговая», «С областями» и «Точечная».

ГИСТОГРАММА (столбиковая диаграмма) представляет собой набор вертикальных столбиков, высота которых определяется значениями данных. Она является одним из самых распространенных типов диаграмм. Применяется для сравнения значений числовых данных в одном или нескольких рядах, а также представления изменения данных во времени. Одним из вариантов представлений гистограммы является внутрестолбиковая диаграмма, которая демонстрирует высотой столбика уровень явления в целом, в его части показывают доли этого явления.

ЛИНЕЙЧАТАЯ диаграмма – это гистограмма, столбики которой располагаются не вертикально, а горизонтально.

ГРАФИК отображает данные в форме сглаженной или ломаной линии, соединяющей точки, соответствующие значениям данных.

КРУГОВАЯ (секторная) диаграмма является одним из самых простых видов диаграмм. Она строится по одному ряду числовых данных и показывает долю каждого числового значения в сумме значений всего явления в целом. Как правило, каждый сектор круга обозначается подписью величины доли в процентах.

ДИАГРАММА С ОБЛАСТЯМИ иллюстрирует величину изменений явления в зависимости от времени подобно графику, но с накоплением уровня явления изображаемого цветом под линией. Например, график прибыли предприятия по годам, отображается в диаграмме с областями, чтобы обратить внимание на величину общей прибыли.

ТОЧЕЧНАЯ диаграмма отображает взаимосвязь между числовыми значениями в нескольких рядах в координатах X и Y. Точечная диаграмма имеет две оси значений, при этом один набор значений выводится вдоль горизонтальной оси

(оси X), а другой — вдоль вертикальной оси (оси Y). На точечной диаграмме эти значения объединяются в одну точку данных и отображаются с неравными интервалами, или кластерами. Точечные диаграммы обычно используются для представления и сравнения научных или статистических данных.

Дополнительными видами диаграмм являются: «Биржевая», «Поверхность», «Кольцевая», «Пузырьковая», «Лепестковая». Они применяются в дополнение к стандартным изображениям лишь в отдельных случаях. Например, лепестковая диаграмма используется для наглядного изображения повторяющихся циклических процессов. Благодаря внешнему виду она, также, называемая «Паутиной» или «Звездой». Лепестковая диаграмма содержит значения каждой категории вдоль радиальной оси, которая начинается в центре круга и заканчивается на внешнем кольце.

Важнейшей функциональной возможностью программы Excel является ее использование для статистического анализа данных. Хотя Excel уступает специализированным пакетам статистической обработки данных, тем не менее, программа содержит ряд методов, включающий наиболее часто используемые процедуры: средства описательной статистики, критерии проверки статистических гипотез, корреляционные и регрессионные методы и др., что позволяет выполнять статистический анализ медицинских данных.

Часть этих методов реализованы стандартными функциями программы. Например функция =СРЗНАЧ(диапазон) служит для вычисления средней арифметической величины. Большинство статистических методов сгруппированы в специальной надстройке к программе, которая называется «Пакет анализа». Он может быть вызван командой «Данные» - «Анализ данных».

Если в разделе «Данные» отсутствует указанный раздел, необходимо выполнить установку этой надстройки. Для этого требуется открыть основное меню кнопкой «Office», выбрать «Параметры Excel» -> «Надстройки» -> «Перейти» -> «Пакет анализа» -> «ОК». Модуль анализа данных в программе Excel, как правило, не подключается при типовой установке пакета программ Microsoft Office. В связи с этим перед применением программы Excel для статистической обработки данных требуется однократно выполнить установку указанного пакета.

Пример создания табличного электронного документа «Протокол работы студента» для изучения основ медицинской статистики

Задача: создать табличный документ, который должен содержать результаты освоения студентом методики статистического исследования и средств обработки данных в программе Microsoft Excel. Документ должен быть защищен паролем и сохранен в общей папке локальной сети для совместного доступа к нему преподавателя и студентов.

Решение: необходимо запустить программу Excel, на листе1 внести заголовок и название документа, ФИО студента(ов), которые будут работать с ним, срок обучения, цели и основные задачи. Выполнить оформление заголовка полужирным шрифтом и разместить данные в удобном для обзора положении. Пример заполнения ячеек таблицы этими данными показан на рисунке 3.

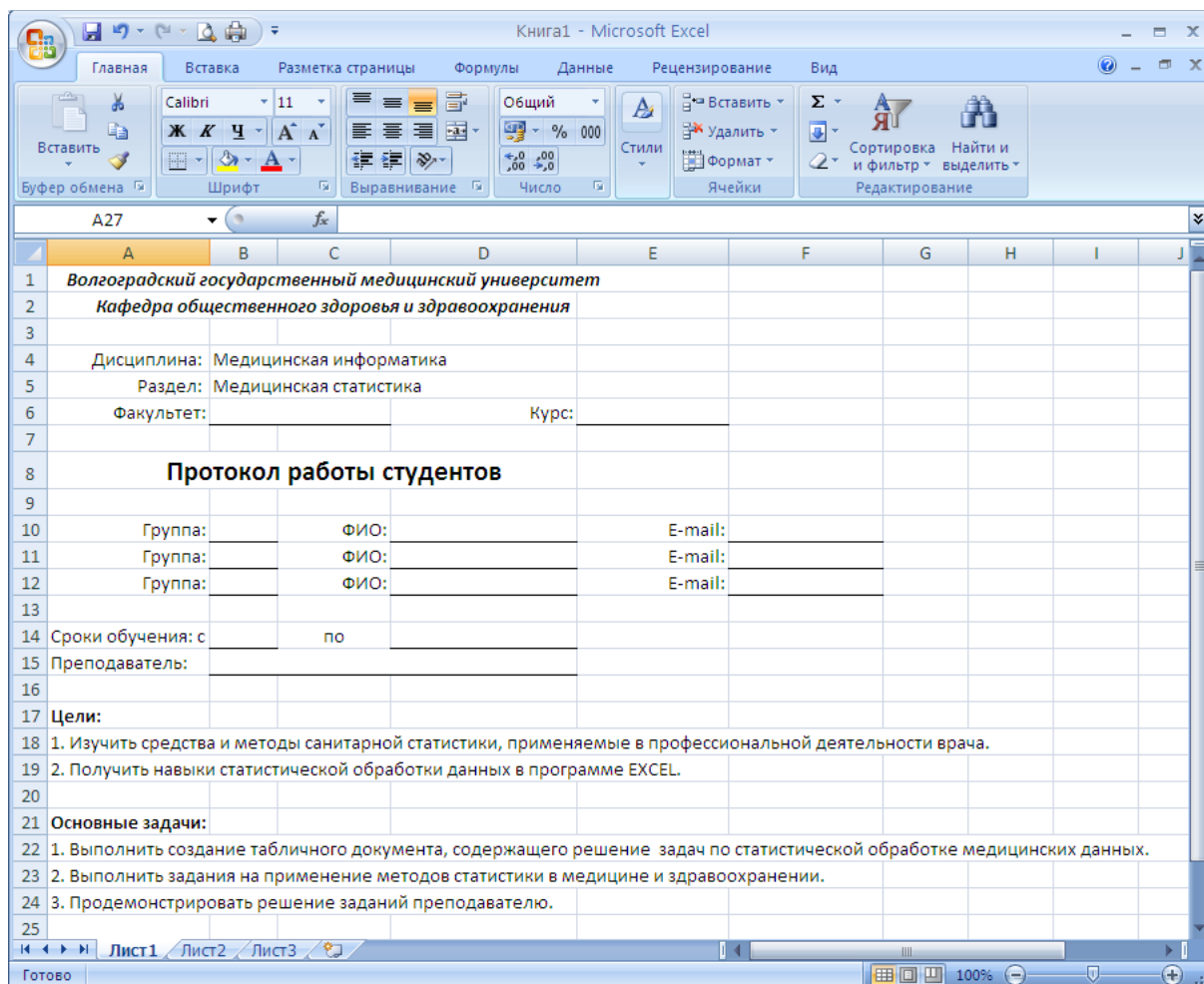



Рис. 3. Пример заполнения табличного документа протокола работы студента.


С целью предотвращения несанкционированного доступа к документу используется защита паролем. Установка пароля на открытие файла производится с помощью кнопки «Office»  в разделе «Подготовить» - > «Зашифровать документ», где дважды необходимо ввести устанавливаемый пароль.

ВНИМАНИЕ! Необходимо обязательно запомнить или записать установленный пароль. Если он будет утерян, открытие файла станет невозможным.

ЗАДАНИЕ

Создайте и заполните табличный документ протокола работы студента. Файл протокола будет использоваться для примеров и решений задач, изложенных ниже.

а) Запустите программу Excel и создайте на рабочем столе компьютера файл для выполнения задания. Имя файла должно состоять из фразы: «Статистика-Фамилии студентов». На листе1 введите текст титульного листа, как показано на рис.3. Заголовок документа «Протокол работы студента(ов)» указывается в центре листа. Ниже заголовка укажите ФИО студента(ов), которые будут работать с этим файлом и период обучения.

ВНИМАНИЕ! Установите защиту на открытие файла с назначением пароля, для этого выполните нажатие кнопки «Office»  и в основном меню программы выберите раздел «Подготовить» -> «Зашифровать документ», введите свой пароль. Запишите пароль в рабочей тетради, если он будет потерян, преподаватель не сможет оценить работу студента. Сохраните файл и закройте программу Excel.

б) Откройте табличный документ с именем «Статистика–Фамилии студентов». Файл может быть открыт только в случае правильного ввода пароля. Измените название «Лист1» на «Титул». В нижней части этого листа введите цели и задачи работы, как показано на рис.3. Сохраните файл.

в) Измените имя «Лист2» и укажите его название «**Формулы**». На этом листе в ячейку В1 внесите текст «Частота пульса студентов на занятии». Заполните диапазон ячеек от А3 до А14 номерами от 1 до 12. Подсчитайте частоту пульса студентов своей учебной группы и внесите полученные числовые данные в ячейки от В3 до В14. В ячейке В15 выполните вычисление средней частоты пульса функцией =СРЗНАЧ(В3:В14). В ячейке В16 определите минимальное значение пульса функцией =МИН(В3:В14). В ячейке В17 определите максимальное значение пульса функцией =МАКС(В3:В14). В ячейке В17 вычислите ранг числа 75 в числовом ряду значений частоты пульса функцией =РАНГ(72; В3:В14).

в) Откройте надстройку «Анализ данных» и выполните запуск модуля «Описательная статистика». Укажите входной интервал ячеек \$B\$3:\$B\$14, а выходной интервал \$D\$3. Установите отметки для вывода данных «Итоговая статистика», «Уровень надежности 95%», «К-й наименьший», «К-й наибольший», нажмите кнопку «ОК».

ВНИМАНИЕ! После решения задач сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки преподавателем правильности выполнения заданий удалите файл рабочего стола своего компьютера.

Вопросы для самоконтроля

1. Назначение программы Excel?
2. Какое название получает новый файл электронного документа, созданный в программе Excel?
3. Из каких элементов состоит рабочая область программы Excel?
4. Каким действием выполняется открытие следующего листа?
5. Как обозначаются заголовки столбцов и строк в программе Excel?
6. Из каких элементов состоит адрес ячейки?
7. Чем отличается относительный и абсолютный адрес ячейки?
8. Какие ячейки могут использоваться в вычислениях?
9. Как обозначается диапазон адресов?
10. Для чего применяется строка формул?
11. Назначение стандартных функций?
12. Основные ошибки, связанные с применением формул в программе Excel?
13. Основные и дополнительные виды диаграмм.
14. Как установить пакет анализа для программы Excel?

Рекомендуемая литература:

1. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
2. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.
3. Симонович С.В. Информатика. Базовый курс: Учебник для вузов. 3-е изд. Стандарт третьего поколения. – СПб.: Питер, 2011. – 640 с.: ил.

II. Базовые понятия теории вероятностей

О теории вероятностей

ТЕОРИЯ ВЕРОЯТНОСТЕЙ — это раздел математической науки, изучающей закономерности случайных явлений (событий).

Даже само высказывание "вычислить вероятность" содержит парадокс. Ведь вероятность, в противоположность достоверности, есть то, чего не знают. Как же можно вычислять то, о чем нет никаких знаний? Пользуясь классическим определением, можно сказать, что вероятность - это отношение числа случаев, благоприятствующих изучаемому событию, к полному числу возможных случаев.

Термины и определения теории вероятностей

ОПЫТ (ИСПЫТАНИЕ), УСЛОВИЯ

Под опытом (испытанием) в теории вероятностей принято понимать наблюдение какого-либо явления при соблюдении определенного комплекса условий, который должен каждый раз строго выполняться при повторении данного опыта. Если то же самое явление наблюдается при другом комплексе условий, то это уже другой опыт (испытание). В теории вероятностей принимается допущение, что испытание может быть повторено бесконечное число раз.

Например, выполнение штрафного броска в баскетболе есть испытание, при этом его благоприятным результатом является попадание мяча в кольцо (благоприятный исход события).

СОБЫТИЕМ называется любой результат опыта (испытания), произошедший в действительности. При проведении испытания требуемый результат (событие) может быть получен с различной степенью возможности под влиянием большого числа известных и неизвестных факторов. В связи с этим исход испытания, как правило, является случайным.

СЛУЧАЙНЫМ СОБЫТИЕМ называется любой факт действительности, который может произойти или не произойти в результате опыта.

В некоторых случаях можно сказать, что одно событие произойдет практически наверняка (достоверное событие), другое практически никогда (невозможное событие).

Примеры: наступление дня после ночи – событие достоверное. Превращение воды в лёд при температуре $+30^{\circ}\text{C}$ – событие невозможное. Вечером пойдёт дождь – событие случайное.

События принято обозначать большими латинскими буквами $A, B, C \dots$. При этом достоверное событие обозначаются буквой W , а невозможное событие символом \emptyset (пустое множество).

ДОСТОВЕРНЫМ СОБЫТИЕМ (W) называется событие, которое наверняка произойдет в результате опыта. Событие называется невозможным (\emptyset), если оно никогда не произойдет в результате опыта.

В отношении друг друга события также имеют особенности, т.е. в одном случае событие A может произойти совместно с событием B , в другом – нет.

Например, если из коробки, содержащей только красные и зеленые шары, наугад вынимают один шар, то появление среди вынутых шаров белого –

невозможное событие (\emptyset). При этом появление красного и появление зеленого шаров образуют полную группу событий.

НЕСОВМЕСТНЫМИ называются события, если появление одного из них исключает появление других. $A \bullet B = \emptyset$

Классическим примером несовместных событий является результат подбрасывания монеты – выпадение лицевой стороны монеты исключает выпадение обратной стороны в одном и том же опыте.

ПРОТИВОПОЛОЖНЫМ событию A , называется событие \bar{A} , состоящее в ненаступлении события A . Очевидно, что события A и \bar{A} несовместны.

Определение вероятности

ВЕРОЯТНОСТЬЮ $P(A)$ события называется отношение числа m исходов опыта, благоприятствующих событию A , к общему числу n возможных исходов опыта, образующих полную группу равновероятных попарно несовместных событий:

$$P(A) = \frac{m}{n}.$$

Аксиомы теории вероятностей

- АКСИОМА 1 – ВЕЛИЧИНА ВЕРОЯТНОСТИ. Каждому случайному событию A соответствует определенное число $P(A)$, называемое его вероятностью и удовлетворяющее условию $0 \leq P(A) \leq 1$.

- АКСИОМА 2 – ВЕРОЯТНОСТЬ ДОСТОВЕРНОГО СОБЫТИЯ. Вероятность достоверного события всегда равна единице. Действительно для достоверного события число благоприятных исходов события $m=n$,

тогда: $P(W) = \frac{m}{n} = \frac{n}{n} = 1$.

В то же время вероятность невозможного события всегда равна нулю, действительно число благоприятных исходов события $m=0$, тогда: $P(\emptyset) = \frac{0}{n} = 0$.

- АКСИОМА 3 - СЛОЖЕНИЯ ВЕРОЯТНОСТЕЙ. Пусть A и B – несовместные события. Тогда вероятность того, что произойдет хотя бы одно из этих двух событий, равна сумме их вероятностей: $P(A+B)=P(A)+P(B)$.

Теоремы теории вероятностей

- ТЕОРЕМА 1 – ВЕРОЯТНОСТЬ ПРОТИВОПОЛОЖНОГО СОБЫТИЯ. Для любого события A вероятность противоположного события \bar{A} выражается равенством $P(\bar{A}) = 1 - P(A)$.

- ТЕОРЕМА 2 - ВЕРОЯТНОСТЬ НЕВОЗМОЖНОГО СОБЫТИЯ. Вероятность невозможного события всегда равна нулю.

- ТЕОРЕМА БЕРНУЛИ – ЗАКОН БОЛЬШИХ ЧИСЕЛ. Относительная частота появления события приближенно равна вероятности этого события и тем точнее, чем больше число испытаний.

Действия над событиями

В математике принято записывать действия над событиями с помощью знаков математических вычислений. В теории вероятностей рассматриваются два основных действия: сложение и умножение событий.

СЛОЖЕНИЕМ (ИЛИ ОБЪЕДИНЕНИЕМ) ДВУХ СОБЫТИЙ A и B называется событие C , заключающееся в том, что произойдет по крайней мере одно из событий A или B . Это действие обозначается так: $C=A+B$.

Задача. В урне 2 зеленых, 7 красных, 5 коричневых и 10 белых шаров. Какова вероятность появления цветного шара?

Решение. Находим соответственно вероятности появления зеленого, красного и коричневого шаров:

$$P(\text{зел.})=2/24; P(\text{кр.})=7/24; P(\text{кор.})=5/24.$$

Так как рассматриваемые события, очевидно, несовместны, то, применяя аксиому сложения, найдем вероятность появления цветного шара:

$$P(\text{цв.}) = P(\text{зел.}) + P(\text{кр.}) + P(\text{кор.}) = \frac{2}{24} + \frac{7}{24} + \frac{5}{24} = \frac{7}{12} = 0,58.$$

УМНОЖЕНИЕМ (ИЛИ СОВМЕЩЕНИЕМ) ДВУХ СОБЫТИЙ A и B называется событие, состоящее в совместном наступлении как события A , так и события B . Это событие будем обозначать $C=AB$.

Задача: Из урны, содержащей 3 белых и 7 черных шаров, вынимают два шара. Какова вероятность того, что оба шара окажутся белыми?

Решение: Находим соответственно вероятности появления первого белого шара $P(\text{бел1})=3/10$, находим вероятность появления второго белого шара $P(\text{бел2})=2/9$. Вероятность появления 2 х белых шаров $P(2 \text{ белых}) = P(\text{бел1}) * P(\text{бел2}) = 3/10 * 2/9 = 2/30 = 0,067$.

Как показано в приведенных примерах, для вычисления вероятности событий необходимо правильно определить общее число возможных исходов таких событий (n), а также возможное число благоприятных исходов (m). Для этого используются формулы комбинаторики.

Элементы комбинаторики

Определение числа возможных перестановок или комбинаций k элементов из их общего числа n выполняется на основе нескольких комбинаторных конфигураций. Основными комбинаторными конфигурациями являются:

- Правило произведения
- Перестановка
- Выборка с возвращением
- Размещение
- Сочетание

ПРАВИЛО ПРОИЗВЕДЕНИЯ применяется, когда необходимо определить число способов (m), которыми могут быть выполнены одно за другим k действий. Если первое действие можно выполнить n_1 способами, второе n_2 способами и так до n_k действия, то возможное число способов равно:

$$m = n_1 * n_2 * n_k.$$

Задача: в гардеробе имеется 5 различных галстуков, 8 рубашек и 3 заколки для галстука. Требуется вычислить возможное число разных вариантов по их использованию в одном комплекте

Решение: число разных вариантов использования в одном комплекте 5 галстуков, 8 рубашек и 3 заколок согласно правилу произведения будет равно:

$$m = 5 * 8 * 3 = 120.$$

В программе Excel такие вычисления выполняются формулой: =5*8*3.

ПЕРЕСТАНОВКА – это такие выборки элементов, которые отличаются только порядком расположения элементов. Число перестановок из n элементов (например чисел $1, 2, \dots, n$) называется всякий упорядоченный набор из этих элементов. Возможное число перестановок из n элементов определяется по формуле:

$$A_n = 1 * 2 * 3 * \dots * (n-2) * (n-1) * n = n!,$$

где: A_n - число перестановок.

Обозначение $n!$ означает факториал числа, например: $4! = 1*2*3*4 = 24$.

В программе Excel для вычисления факториала применяется функция =ФАКТР(n), поэтому формула в ячейке Excel = ФАКТР(4) даст результат 24.

Задача: за круглый стол садятся 7 человек. Требуется вычислить число возможных вариантов их пересаживаний (перестановок) за столом.

Решение: число возможных вариантов размещения 7 человек за столом равно

$$A_7 = 1 * 2 * 3 * 4 * 5 * 6 * 7 = 7! = \text{ФАКТР}(7) = 5040.$$

ВЫБОРКА С ВОЗВРАЩЕНИЕМ – это число способов, которыми можно создать комбинацию из n элементов k действиями с возвращением элементов в исходное множество. Определение возможного числа способов (m), которыми выполняется k действий, для множества (n) с неизменным числом элементов (с возвратом выбранных элементов в множество) равно:

$$m = n^k.$$

Задача: какое количество слогов по две буквы можно получить из четырех букв: А, Б, В, Г?

Решение: из 4-х букв (n), включающих 2 буквы (k) можно составить количество слогов (m) равное:

$$m = n^k = 4^2 = 16.$$

В программе Excel такие вычисления выполняются формулой =4^2.

Действительно, возможные слоги выглядят так: 1-й «АА», 2-й «АБ», 3-й «АВ», 4-й «АГ», 5-й «БА», 6-й «ББ», 7-й «БВ», 8-й «БГ», 9-й «ВА», 10-й «ВБ», 11-й «ВВ», 12-й «ВГ», 13-й «ГА», 14-й «ГБ», 15-й «ГВ», 16-й «ГГ».

Условная вероятность события может быть вычислена в случае, когда оно может произойти при наступлении другого события, если известны вероятности их появления с помощью формулы Байеса.

ТЕОРЕМА БАЙЕСА (ФОРМУЛА БАЙЕСА) – это одна из основных теорем теории вероятностей, которая позволяет определить вероятность того, что произошло какое-либо событие (гипотеза) при наличии лишь косвенных тому подтверждений (данных), которые могут быть неточны (вероятностны):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

где:

$P(A|B)$ – условная вероятность наступления события A при наступлении события B ;

$P(B|A)$ – апостериорная вероятность наступления события B (лат. a'posteriori – «из последующего»: знание, полученное после опыта) – это вероятность истинности гипотезы B при наступлении события A определяемая экспериментально;

$P(A)$ – априорная вероятность события A (лат. a'priori – «из предшествующего»: знание, полученное до опыта) – это классическая вероятность события, вычисленная заранее;

$P(B)$ – априорная вероятность наступления события B .

Формула названа в честь ее автора - Томаса Байеса. Его работа «An Essay towards solving a Problem in the Doctrine of Chances» была опубликована в 1763 году, через 2 года после смерти автора. В настоящее время применяется в экспертных и диагностических информационных системах.

Однако психологические эксперименты показали, что люди при оценках вероятности игнорируют различие априорных вероятностей (ошибка базовой оценки), и потому правильные результаты, получаемые по теореме Байеса, могут существенно отличаться от ожидаемых.

Следствие теоремы Байеса является формула полной вероятности:

$$P(B) = \sum_{i=1}^N P(A_i)P(B|A_i)$$

Она служит для вычисления вероятности наступления события B , зависящего от ряда гипотез A_i , если известны степени достоверности (вероятности) этих гипотез (например, измерены экспериментально). События, которые представляют собой гипотезы, обозначаются знаком « H_i ».

Задача: в аптеку поставляется лекарство, изготовленное на трёх заводах, причем 40% упаковок поступает с завода №1, 50 % – с завода №2 и 10% – с завода №3. На каждом заводе возможно изготовление бракованных упаковок, при этом доля брака в упаковках, изготовленных на заводе №1=1%, на заводе №2=2%, на

заводе №3=4%. Взятая упаковка лекарства оказалась качественной. Определить вероятность того, что она изготовлена на заводе №1.

Решение: пусть событие A – взятая качественная упаковка. При этом возможны три гипотезы H_1, H_2, H_3 – лекарство изготовлено, соответственно на 1-м, 2-м и 3-м заводе. Априорные вероятности каждой из этих гипотез по условию задачи составляют:

$$P(H_1) = 0,4; P(H_2) = 0,5; P(H_3) = 0,1.$$

Соответствующие вероятности события A , если упаковки лежат отдельно:

$$P(A/H_1) = 1 - 0,01 = 0,99;$$

$$P(A/H_2) = 1 - 0,02 = 0,98;$$

$$P(A/H_3) = 1 - 0,04 = 0,96.$$

Тогда полная вероятность события A (взятая качественная упаковка):

$$P(A) = 0,4 \cdot 0,99 + 0,5 \cdot 0,98 + 0,1 \cdot 0,96 = 0,982.$$

Условная вероятность гипотезы H_1 (качественная упаковка взята на заводе №1) равна:

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0,4 \cdot 0,99}{0,982} = 0,403.$$

Задача: два стрелка подбрасывают монетку и выбирают, кто из них стреляет по мишени одной пулей. Первый стрелок попадает по мишени с вероятностью 1 (100%), второй стрелок – с вероятностью 0,1 (10%). После выстрела пуля попала в мишень, необходимо определить какой стрелок выстрелил.

Решение: можно сделать два предположения (гипотезы) об эксперименте: H_1 - выстрелил 1-й стрелок, H_2 - выстрелил 2-й стрелок. Априорные («до опыта») вероятности этих гипотез по условию задачи определяются классическим способом (подбрасывание монеты) и одинаковы: $P(H_1) = P(H_2) = 1/2$. Рассмотрим событие A – пуля попала в мишень. При этом по условию задачи известно, что апостериорная вероятность этого события для первой гипотезы составляет: $P(A|H_1) = 100\% = 1$, а для второй $P(A|H_2) = 10\% = 0,1$.

Полная вероятность попадания в мишень равна:

$$\begin{aligned} P(A) &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \\ &= 1/2 * 1 + 1/2 * 0,1 = 0,55. \end{aligned}$$

Тогда условная вероятность, что выстрелил первый стрелок составляет:

$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A)} = \frac{1 * 0,5}{0,55} = 0,91.$$

Примеры решения задач теории вероятностей в программе Excel

Задача 1. На швейной фабрике изготавливают медицинские халаты. В среднем на 180 халатов приходится восемь со скрытыми дефектами. Найдите вероятность того, что купленный в магазине халат этой фабрики окажется качественным.

Решение. В программе Excel в отдельные ячейки таблицы необходимо ввести текст и данные условия задачи. Вычисление числа качественных халатов производится формулой $=B4-B5$ ($180-8 = 172$). Вероятность покупки качественного халата $P(A)$ определяется с помощью формулы $=B6/B4$ ($172/180 = 0,95555\dots$). Результат вычисления округляется функцией $=ОКРУГЛ()$ до 2-х знаков после запятой, получаем **0,96**.

	A	B	C	D	E	F
1	Задача 1.					
2	На швейной фабрике изготавливают медицинские халаты. В среднем на 180 халатов приходится восемь со скрытыми дефектами. Найдите вероятность того, что купленный в магазине халат этой фабрики					
3	окажется качественным.					
4	Решение.					
4	Число партии халатов	180				
5	Число халатов с дефектами	8				
6	Число качественных халатов	172				
7	Вероятность покупки качественного халата	0,96				
8						

Задача 2. В соревнованиях участвуют 9 студентов - спортсменов из ВолГУ, 3 спортсмена из ВГТУ, 8 спортсменов из ВАГС и 5 из ВолгГМУ. Порядок, в котором выступают спортсмены, определяется жребием. Найдите вероятность того, что спортсмен, который выступает первым, окажется из ВолгГМУ.

Решение. Необходимо запустить программу Excel и в ячейки таблицы ввести текст условия задачи, а затем данные. Вычисление числа участников соревнований выполняется в ячейке B7 формулой $=СУММ(B3:B6)$ т.е. $9+3+8+5=25$. Так как число спортсменов из ВолгГМУ 5 человек, то вероятность $P(A)$ того, что первым участником станет спортсмен из этого вуза, вычисляется формулой $=B6/B7$ ($5/25 = 0,2$).

	A	B	C	D	E	F	G	H
1	Задача 2. В соревнованиях участвуют 9 студентов-спортсменов из ВолГУ, 3 спортсмена из ВГТУ, 8 спортсменов из ВАГС и 5 из ВолгГМУ. Порядок, в котором выступают спортсмены, определяется жребием. Найдите вероятность того, что спортсмен, который выступает							
2	первым, окажется из ВолгГМУ.							
3	Решение.							
3	Спортсменов из ВолГУ	9						
4	Спортсменов из ВГТУ	3						
5	Спортсменов из ВАГС	8						
6	Спортсменов из ВолгГМУ	5						
7	Всего участников	25						
8	Вероятность выбора первого спортсмена из студентов ВолгГМУ	0,2						

Задача 3. В цветочном магазине в одной вазе стоят 3 красных цветка (гвоздика, роза и лилия) и 4 белых цветка (гвоздика,

	A	B	C	D
1	Задача 3. В цветочном магазине в одной вазе стоят 3 красных цветка (гвоздика, роза и лилия) и 4 белых цветка (гвоздика, роза, пион и астра). Сколько вариантов составления букета из 3-х цветов можно создать? Букет составляет продавец, не различающий цвет, какова вероятность появления			
2	букета, состоящего из 1-го красного и 2-х белых цветков?			
3	Решение.			
3	Число красных цветов	3		
4	Число белых цветов	4		
5	Всего цветов (n)	7		
6	Число белых цветов в букете (k)	3		
7	Число возможных сочетаний C _n ^k	35		

роза, пион и астра). Сколько вариантов составления букета из 3-х цветов можно создать? Букет составляет продавец, не различающий цвет, какова вероятность появления букета, состоящего из 1-го красного и 2-х белых цветков?

Решение. В программе Excel в отдельные ячейки необходимо ввести текст и данные условия задачи. Для определения числа возможных букетов из 3-х цветов (k), выбираемых из 7-и цветов (n), выполняются вычисления по формуле сочетаний:

$$C_n^k = n! / k! * (n - k)!,$$

где:

C_n^k – (ячейка В7) возможное число сочетаний по k (число цветов в букете) из n (общее число цветов) равно $C_n^k = \frac{7!}{3! * (7-3)!} = \frac{5040}{6 * 24} = 35$;

n – (ячейка В5) количество всех цветов равно числу белых гвоздик + красных гвоздик (В2 + В3) = 7;

k – (ячейка В6)требуемое количество цветов в букете = 3.

Для вычисления C_n^k в ячейке В7 может использоваться формула =ФАКТР(В5)/(ФАКТР(В6)*ФАКТР(В5-В6)) или функция =ЧИСЛОМБ(В5;В6).

Определение вероятности появления букета, состоящего из 1-го красного и 2-х белых цветков, если они извлекаются из вазы случайно, используются правила сложения и умножения вероятностей. Обозначим знаком A событие «Появление красного цветка», его вероятность $P(A)$

вычисляется в ячейке В9 формулой =3/7, или =В3/В7 и равна 0,43.

Событие «Появление первого белого цветка» обозначим $H1$, тогда его вероятность $P(H1)$ вычисляется в ячейке В10 формулами =4/6, или =В4/В7 и равна 0,67. Событие $H2$ «Появление второго белого цветка» характеризуется вероятностью $P(H2)$. Поскольку один белый цветок ранее был взят из корзины, вероятность извлечения другого белого цветка из оставшихся вычисляется в ячейке В11 формулами = (4-1)/(6-1), или = (В4-1)/(В7-1) и равна 0,6. Вероятность появления букета из одного красного и двух белых цветков по теоремам сложения и умножения вероятностей $P = P(A)+P(H1)* P(H2) = 0,83$ и вычисляется формулами = 0,43+0,67*0,6 или = В9+В10*В11.

	A	B	C	D
Задача 1. В цветочном магазине в одной вазе стоят 3 красных цветка (гвоздика, роза и лилия) и 4 белых цветка (гвоздика, роза, пион и астра). Сколько вариантов составления букета из 3-х цветов можно создать? Букет составляет продавец, не различающий цвет, какова вероятность появления букета, состоящего из 1-го красного и 2-х белых цветков?				
Решение				
Число красных цветов		3		
Число белых цветов		4		
Всего цветов (n)		7		
Число цветов в букете (k)		3		
Число возможных сочетаний C _n ^k		35		
Вероятность события A - появление красного цветка P(A)=3/7			0,43	
Вероятность события H1 - появление появления первого белого цветка.			0,67	
Вероятность события H2 - появление появления второго белого цветка.			0,60	
Вероятность букета из 1-го красного и 2-х белых цветков			0,83	

ЗАДАНИЯ

С помощью программы Excel вычислите вероятность событий.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «ТеорияВероятности» и выполните задачи варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку

своей группы. **Сообщите преподавателю о выполнении задания.** После проверки решения задач преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

Задача 1: Бросают игральную кость один раз. Какова вероятность выпадения 2 или 3?

Задача 2: Первый стрелок попадает в цель с вероятностью 0,9, второй стрелок - с вероятностью 0,5. Какова вероятность совместного попадания двух стрелков?

Задача 3: В гаражном кооперативе для кодирования замков боксов используются трехзначные номера из цифр от 0 до 5. Найти число комбинаций различных кодов. Какова вероятность открыть замок кодом 232?

Задача 4: Ребенок, не знающий алфавит, играет карточками с буквами А, В, Р, Ч. Какова вероятность, что ребенок сложит слово «ВРАЧ»?

Задача 5: Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 для второго. Найти вероятность того, что при аварии сработает только один сигнализатор.

Задача 6: Вероятность хотя бы одного попадания в цель при четырех выстрелах равна 0,9984. Найти вероятность попадания в цель при одном выстреле.

Задача 7: Молодой человек забыл номер своего приятеля, но помнит из него первые 4 цифры. В телефонном номере 7 цифр. Какова вероятность, что молодой человек дозвонится до своего приятеля, если наберёт номер случайным образом?

Задача 8: Монета брошена три раза. Найти вероятность того, что хотя бы один раз появится изображение герба.

Задача 9: В ящике находятся 5 красных и 8 синих шаров. Наугад вынимаются 2 шара. Какова вероятность, что они:

- а) оба синие;
- б) одного цвета;
- в) разных цветов?

Вариант 2

Задача 1: Бросают игральную кость. Какова вероятность выпадения 1 и 6.

Задача 2: Первый стрелок попадает в цель с вероятностью 0,8, второй стрелок - с вероятностью 0,4. Какова вероятность совместного попадания двух стрелков?

Задача 3: В гаражном кооперативе для кодирования замков боксов используются трехзначные номера из цифр от 0 до 9. Найти число комбинаций различных кодов. Какова вероятность открыть замок кодом 323?

Задача 4: Ребенок, не знающий алфавит, играет карточками с буквами А, В, Г, Л, О. Какова вероятность, что ребенок сложит слово «ВОЛГА»?

Задача 5: В квартире установлены две независимо работающих сигнализации. Вероятность того, что при ограблении первая сигнализация сработает, равна 0,85, а вторая 0,7. Найти вероятность того, что при аварии сработает только одна сигнализация.

Задача 6: Вероятность достать из ящика хотя бы один красный шар при четырех попытках равна 0,788. Найти вероятность достать красный шар при одной попытке.

Задача 7: Молодой человек забыл номер своего приятеля, но помнит из него первые 3 цифры. В телефонном номере 5 цифр. Какова вероятность, что молодой человек НЕ дозвонится до своего приятеля, если наберёт номер случайным образом?

Задача 8: Монета брошена три раза. Найти вероятность того, что хотя бы один раз появится изображение герба.

Задача 9: В ящике находятся 3 красных и 7 синих шаров. Наугад вынимаются 2 шара. Какова вероятность, что они:

- а) оба красные;
- б) одного цвета;
- в) разных цветов?

Вариант 3

Задача 1: Бросают игральную кость один раз. Какова вероятность выпадения 1 или 6.

Задача 2: Первый стрелок попадает в цель с вероятностью 0,8, второй стрелок - с вероятностью 0,4. Какова вероятность совместного попадания двух стрелков?

Задача 3: В гаражном кооперативе для кодирования замков боксов используются трехзначные номера из цифр от 0 до 9. Найти число комбинаций различных кодов. Какова вероятность открыть замок кодом 323?

Задача 4: Ребенок, не знающий алфавит, играет карточками с буквами Е, К, М, О, П, Р, Т, Ю. Какова вероятность, что ребенок сложит слово «КОМПЬЮТЕР»?

Задача 5: В квартире установлены две независимо работающих сигнализации. Вероятность того, что при ограблении первая сигнализация сработает, равна 0,85, а вторая 0,7. Найти вероятность того, что при аварии сработает только одна сигнализация.

Задача 6: Вероятность достать из ящика хотя бы один красный шар при четырех попытках равна 0,788. Найти вероятность достать красный шар при одной попытке.

Задача 7: Молодой человек забыл номер своего приятеля, но помнит из него первые 3 цифры. В телефонном номере 5 цифр. Какова вероятность, что молодой человек НЕ дозвонится до своего приятеля, если наберёт номер случайным образом?

Задача 8: Монета брошена три раза. Найти вероятность того, что хотя бы один раз появится изображение герба.

Задача 9: В ящике находятся 3 красных и 7 синих шаров. Наугад вынимаются 2 шара. Какова вероятность, что они:

- а) оба красные;
- б) одного цвета;
- в) разных цветов?

Вариант 4

Задача 1: Бросают игральную кость один раз. Какова вероятность выпадения 2 или 3.

Задача 2: Первый стрелок попадает в цель с вероятностью 0,9, второй стрелок - с вероятностью 0,5. Какова вероятность совместного попадания в цель двух стрелков?

Задача 3: В гаражном кооперативе для кодирования замков боксов используются трехзначные номера из цифр от 0 до 5. Найти число комбинаций различных кодов. Какова вероятность открыть замок кодом 232?

Задача 4: Ребенок, не знающий алфавит, играет карточками с буквами А, Е, И, Н, П, Т, Ц. Какова вероятность, что ребенок сложит слово «ПАЦИЕНТ»?

Задача 5: Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,9 для второго. Найти вероятность того, что при аварии сработает только один сигнализатор.

Задача 6: Вероятность хотя бы одного попадания в цель при четырех выстрелах равна 0,9984. Найти вероятность попадания в цель при одном выстреле.

Задача 7: Молодой человек забыл номер своего приятеля, но помнит из него первые 4 цифры. В телефонном номере 7 цифр. Какова вероятность, что молодой человек дозвонится до своего приятеля, если наберёт номер случайным образом?

Задача 8: Монета брошена три раза. Найти вероятность того, что хотя бы один раз появится изображение герба.

Задача 9: В ящике находятся 5 красных и 8 синих шаров. Наугад вынимаются 2 шара. Какова вероятность, что они:

- а) оба синие;
- б) одного цвета;
- в) разных цветов?

Вопросы для самоконтроля

1. Теория вероятностей как раздел математической науки.
2. Дайте определение понятиям «Испытание» и «Событие».
3. Что называется «Случайным событием»?
4. Как принято обозначать события в теории вероятностей?
5. Какие события называются несовместными?
6. Какие обозначения используются для достоверных, недостоверных и противоположных событий?
7. Какая формула применяется для вычисления классической вероятности события?
8. Назовите аксиомы теории вероятностей.
9. Какие теоремы применяются в теории вероятностей.
10. Как вычисляется вероятность объединения (сложения) двух событий?
11. Приведите пример вычисления вероятности совмещения (умножения) двух событий?
12. Какие конфигурации применяются в комбинаторике?
13. Объясните на примере правило произведения в комбинаторике.
14. В каких ситуациях и как применяется вычисление числа перестановок в комбинаторике?
15. Как в программе Excel выполняется вычисление факториала?
16. Как вычисляют число способов в выборке с возвращением?
17. Какая формула применяется для вычисления числа возможных сочетаний?
18. Какую функцию можно применить в программе Excel для вычисления числа сочетаний?
19. Что называют «Размещением» в комбинаторике и какой функцией выполняются вычисления в программе Excel?
20. Сформулируйте теорему Байеса.
21. Запишите и объясните формулу полной вероятности.

Рекомендуемая литература:

1. Гмурман В.Е. Теория вероятности и математическая статистика: Учеб.пособие для вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2003. – 479 с.

2. Данилин Г. А. и др. Элементы теории вероятностей с EXCEL: практикум для студентов всех специальностей МГУЛа / Г. А. Данилин, В. М. Курзина, П. А. Курзин, О. М. Полещук. М.: МГУЛ, 2004. - 87 с.

III. Организация и этапы статистического исследования

В любом медико-биологическом и социально-гигиеническом исследовании ведущая роль отводится проведению статистического наблюдения. Его организация определяет особенности последующей обработки полученных данных, выбор методов статистики и обоснованность выводов. В связи с этим методика проведения статистического наблюдения является одним из важнейших разделов статистики.

Статистика - это общественная наука, которая изучает количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной.

Санитарная (медицинская) статистика - это раздел статистики, изучающий состояние здоровья населения (показатели общественного здоровья) и деятельность лечебно-профилактических учреждений (ЛПУ), то есть состоит из статистики здоровья и статистики здравоохранения.

Предметами изучения медицинской статистики являются: здоровье населения в целом и отдельных возрастно-половых групп; выявление и установление зависимостей между уровнем здоровья и факторами окружающей среды; данные о сети, деятельности, кадрах учреждений здравоохранения; оценка статистической достоверности результатов медико-биологических, клинических и экспериментальных исследований.

Целью исследования в медицине и здравоохранении является выявление закономерностей изучаемого явления на основе проверки статистических (рабочих) гипотез, сформулированных в начале исследования. В зависимости от цели конкретизируются частные задачи и составляется план проведения всей работы, а также детальные программы сбора, разработки и статистического анализа собранного материала.

Рабочей гипотезой называется эмпирически не проверенное предположение, предсказывающее существование некоторой зависимости между переменными или объектами. Рабочие гипотезы используются для выработки предварительного плана научного исследования.

Статистическое исследование начинается с определения объекта наблюдения, единицы наблюдения и учетных признаков.

Объектом статистического наблюдения является статистическая совокупность, состоящая из единиц, о которых должны быть собраны статистические сведения, взятая в определенных границах пространства и времени. Например, при сборе данных о заболеваемости объектом наблюдения является совокупность всех заболеваний или всех больных среди населения на данной территории за определенный период времени.

Единица статистического наблюдения – это составная часть объекта наблюдения, подлежащая изучению и регистрации в соответствии с программой исследования. Например, отдельный случай заболевания или больное лицо, входящее в состав объекта наблюдения.

Каждая единица наблюдения характеризуется рядом признаков, величины которых исследователь регистрирует в процессе сбора статистического материала. Такие признаки принято называть учетными.

Учетные признаки - это медико-биологические характеристики, регистрируемые у единицы наблюдения в соответствии с целями и задачами исследования. Такими признаками могут быть: пол, возраст, место жительства, диагноз, дата заболевания, длительность болезни, ее исход и т.д.

В медицинской статистике используются следующие виды учетных признаков:

- сходства, по которым единицы наблюдения объединяются в статистическую совокупность;
- различия, которые отличают единицы наблюдения между собой;
- факторные, которые влияют на изучаемое явление;
- результативные, которые изменяются под влиянием факторных признаков.

В зависимости от способа регистрации встречаются следующие типы учетных признаков:

- качественные (атрибутивные), которые могут быть выражены словесно, описательно;
- количественные, фиксирующие числовые значения признака.

Важнейшим понятием медицинской статистики является «Статистическая совокупность».

Статистическая совокупность - это группа, состоящая из большого числа относительно однородных элементов (единиц наблюдения), взятых вместе в известных границах времени и пространства. Необходимо различать два основных вида статистических совокупностей - генеральную и выборочную.

Генеральная совокупность состоит из всех возможных единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования. Генеральная совокупность в медико-биологических исследованиях используется редко. Большинство исследований выполняется с использованием выборочной совокупности.

Выборочная совокупность – это часть генеральной совокупности, отобранная специальным методом и предназначенная для характеристики генеральной совокупности.

Указанные совокупности могут быть сформированы двумя основными методами. Методами статистического наблюдения являются:

- сплошное исследование, при котором изучаются все доступные единицы наблюдения генеральной совокупности;
- выборочное, при котором изучается определенная часть единиц наблюдения, наиболее полно характеризующие статистическую совокупность в целом.

В зависимости от продолжительности исследования применяются следующие виды статистического наблюдения: единовременное и текущее.

Единовременное наблюдение – это исследование, при котором статистические данные собираются на определенный (критический) момент времени. Например, перепись населения.

Текущее наблюдение – это непрерывное, повседневное исследование, производимое в течение определенного периода: месяца, полугода, года. Например, изучение уровня заболеваемости населения города или региона.

Перед проведением исследования требуется сформулировать рабочую гипотезу и определить: что является объектом исследования, какие единицы наблюдения будут изучаться, какой перечень признаков (параметров) необходимо учитывать и какой вид исследования будет выполняться. Затем исследователь собирает статистический материал, выполняет его обработку и формулирует заключение, которое подтверждает или опровергает рабочую гипотезу.

Таким образом, научное исследование состоит из нескольких этапов:

- 1-й: составление программы и плана исследования;
- 2-й: статистическое наблюдение;
- 3-й: группировка и разработка статистического материала;
- 4-й: анализ результатов исследования.

На 1 этапе статистического исследования основными видами работ являются:

- составление плана исследования;
- подготовка программы исследования.

Программа исследования включает вопросы, что и в каком направлении изучать, с обозначением объекта и единиц наблюдения, учетных признаков, методов сбора (анкеты, бланки, первичные документы), разработки и анализа материала (макеты таблиц, выбранные статистические критерии).

План исследования отвечает на вопросы: где, сколько, когда, кто и как выполняет исследование, и включает:

- выбор места проведения исследования;
- пути формирования объекта наблюдения (объем выборки, время, способы сбора материала);
- определение единицы наблюдения;

- способы разработки материала;
- сроки работ по этапам;
- выбор исполнителей;
- финансирование исследования;
- инструкции исполнителям, организационное и методическое сопровождение.

Программа сбора материала представляет собой первичный учетный документ (бланк, карта, анкета), в который включены учетные признаки, соответствующие цели исследования и подлежащие регистрации.

При составлении учетного документа необходимо соблюдать следующие правила:

- 1) документ должен иметь четкое заглавие, в котором сформулирована единица наблюдения;
- 2) учетные признаки должны быть указаны краткими названиями и соответствовать цели исследования;
- 3) на каждый вопрос исследования следует предусмотреть варианты ответов в соответствии с выделенными группами единиц наблюдения.

Программа разработки материала предусматривает определение критериев группировки единиц наблюдения и составление макетов статистических таблиц.

На 2-м этапе проводится сбор статистического материала.

Статистическим материалом в каждом данном случае являются первичные учетные документы, официально существующие или специально разработанные (талоны, карты, анкеты и т.п.). Сбор материала проводят в соответствии с составленной ранее программой и планом статистического исследования.

На 3-м этапе осуществляется обработка собранного статистического материала. Он включает следующие последовательно выполняемые действия: контроль, шифровка, группировка, сводка в статистические таблицы, вычисление статистических показателей и их графическое изображение.

Контроль - это проверка собранного материала с целью отбора учетных документов, имеющих дефекты, для их последующего исправления, дополнения или исключения из исследования.

Шифровка - это применение условных обозначений изучаемых признаков с последующим назначением шифра каждой единице наблюдения. При ручной обработке материала шифры могут быть цифровые, буквенные, знаковые; при машинной, как правило, цифровые.

Группировка - это распределение единиц наблюдения на однородные группы по характеру или величине изучаемых признаков.

Виды группировок:

- типологическая - это группировка по атрибутивным (качественным) признакам. Например: пол, профессия.

- вариационная - это группировка на основе признаков, имеющих числовое выражение. Например: возраст, стаж.

Сводка данных - занесение полученных после подсчета цифровых данных в таблицы.

Для дальнейшего анализа материала необходимо произвести расчеты статистических показателей и средних величин в соответствии с программой исследования и выполнить графическое изображение.

На 4-м этапе выполняется обобщение и анализ полученных данных. Он включает:

- интерпретацию полученных различных статистических величин и графических изображений на основе сопоставления с нормативами, со средними уровнями аналогичных величин, со стандартами, с данными по другим учреждениям и территориям, литературными данными, в динамике;
- обобщение результатов исследования;
- литературное оформление работы;
- выявление закономерностей;
- выводы;
- предложения для внедрения в практику;
- прогноз, рекомендации.

На каждом из указанных этапов исследования может применяться компьютерная техника, оснащенная распространенным пакетом программ Microsoft Office или аналогичными программами. Их применение ускоряет выполнение статистических вычислений, упрощает подготовку статистических таблиц, наглядное представление данных и публикацию научных работ. Использование табличных редакторов, в том числе программы Excel, позволяет выполнять статистическую обработку медицинских данных с использованием большинства основных методов прикладной статистики и сохранять их в файлах электронных документов.

Пример программы статистического исследования в табличном документе «Протокол работы студента»

Задача: необходимо заполнить табличный документ данными о плане проведения статистического исследования для изучения распространенности курения среди студентов медицинского вуза.

Решение: необходимо запустить программу Excel и создать новый лист. В ячейках этого листа сформулировать и ввести текст, который должен включать описание планируемых мероприятий и программу предполагаемого статистического исследования. Выполнить форматирование ячеек таблицы для удобного прочтения данных.

ЗАДАНИЕ

Заполните табличный документ данными о плане и программе статистического исследования. Подготовьте макет анкеты для проведения опроса респондентов.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Выполните следующие действия:

а) На рабочем столе компьютера найдите и откройте свой файл «Статистика ...», создайте новый лист и укажите его название «План». В ячейку А1 внесите текст «План исследования», в ячейку А2 - «Изучение распространенности курения среди студентов медицинского вуза». В ячейках с А3 по А7 введите текстовые данные:

- Объект исследования: студенты медицинского вуза.
 - Объем статистической совокупности: 30% студентов 1-го - 6-го курса медицинского ВУЗа.
 - Сроки проведения исследования: с _____ по _____.
 - Методы сбора материала: анкетирование, выкопировка из медицинских документов студенческой поликлиники.
 - Лица, выполняющие сбор и обработку материала: ФИО студентов
- Сохраните файл.

б) Создайте новый лист «Анкета» и внесите опросный лист для сбора данных о распространенности курения среди студентов медицинских вузов. В ячейки с А9 по А22 внесите текст анкеты:

9	Анкета для изучения распространенности курения среди студентов медицинских вузов												
10	1. Ф.И.О студента _____ (вписать полностью)												
11	2. Курс (подчеркнуть): I, II, III, IV, V, VI (подчеркнуть)												
12	3. Факультет (подчеркнуть): лечебный, медико-профилактический, фармацевтический, факультет военного обучения												
13	4. Возраст (подчеркнуть): до 20 лет, 20, 21, 22, 23, 24, 25 и более												
14	5. Пол(подчеркнуть): муж/жен												
15	6. Курите ли Вы? Да, нет, бросил												
16	7. Признаете ли Вы, что курение вредит здоровью? Да, нет, не знаю												
17	8. Кто из членов вашей семьи курит: отец, мать, брат, сестра, муж, жена, товарищ, никто не курит												
18	Если Вы курите:												
19	9. Возраст, в котором выкурили первую сигарету: до 15 лет, 16-18 лет, старше 18 лет												
20	10. Какое количество сигарет Вы выкуриваете в день? 5-10, 11-20, более 20												
21	11. Что побудило Вас впервые закурить: пример родителей, пример преподавателей, влияние товарищей, желание казаться взрослым, желание похудеть, любопытство, желание не отстать от моды?												
22	Благодарим за искренние ответы!												

ВНИМАНИЕ! Сохраните файл и закройте программу. Затем скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки файла преподавателем удалите его с рабочего стола своего компьютера.

Вопросы для самоконтроля

1. Определение статистики.
2. Определение санитарной (медицинской) статистики.
3. Что такое «Рабочая гипотеза»?
4. Содержание этапов статистического исследования.
5. Перечислите основные элементы первого этапа статистического исследования?
6. Какие работы выполняются на втором этапе статистического исследования?
7. Что включает третий этап статистического исследования?
8. Какие действия выполняются на четвертом этапе статистического исследования?
9. Что такое единица наблюдения?
10. Определение понятия "объект наблюдения".
11. Дайте определение статистической совокупности?
12. Понятие о генеральной и выборочной совокупностях.
13. Назовите виды статистического наблюдения?
14. Что такое группировка (определение)?
15. На основе каких признаков выполняется типологическая группировка?
16. Что такое вариационная группировка?

Рекомендуемая литература:

1. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>
2. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.
3. Симонович С.В. Информатика. Базовый курс: Учебник для вузов. 3-е изд. Стандарт третьего поколения. – СПб.: Питер, 2011. – 640 с.: ил.

IV. Статистические таблицы

Каждое медико-биологическое или социально-гигиеническое исследование начинается с этапа планирования эксперимента. На этом этапе необходимо разработать макеты статистических таблиц, которые должны стать основой для последующей обработки и анализа данных. Знание методики построения статистических таблиц позволяет подготовить проведение исследования в соответствии с рабочей гипотезой.

СТАТИСТИЧЕСКАЯ ТАБЛИЦА - это форма записи изучаемой статистической совокупности, разделенной на группы в соответствии с изучаемыми признаками.

В таблице, как и в грамматическом предложении, выделяют табличное подлежащее и табличное сказуемое. ТАБЛИЧНОЕ ПОДЛЕЖАЩЕЕ – это то, о чем говорится в таблице, основной признак или признаки, которые, как правило, обозначены в строках таблицы. ТАБЛИЧНОЕ СКАЗУЕМОЕ или несколько сказуемых – это количественные признаки, характеризующие подлежащее. Они, как правило, расположены в столбцах (графах) таблицы.

Каждая таблица должна иметь заголовок, отражающий ее содержание, шапку с указанием показателей или признаков и единиц измерений, а также столбец «Всего» и строку «Итого». В таблице не должно быть пустых ячеек. Если в документе встречается несколько таблиц, каждая из них обязательно нумеруется. Номер таблицы указывают над ней в правой части страницы.

В зависимости от размера и содержания таблицы она может относиться к одному из видов:

- ПРОСТАЯ (таблица, в которой подлежащее характеризуется лишь одним признаком);
- ГРУППОВАЯ (таблица, в которой подлежащее характеризуется двумя связанными между собой признаками);
- КОМБИНАЦИОННАЯ (таблица, в которой подлежащее характеризуется тремя и более связанными между собой признаками). Комбинационные таблицы, содержащие более 4-х признаков, обычно не применяются из-за громоздкости и сложности анализа. Исследователю лучше составить несколько таблиц, объединяющих по 3-4 связанных признака.

Пример создания макетов статистических таблиц

Задача: составить макеты простой, групповой и комбинационной таблиц для внесения данных распределения заболевших жителей района N в 2010 году по социальному статусу (безработные, рабочие, служащие), классам заболеваний (болезни органов дыхания, инфекционные заболевания, травмы, прочие болезни) и возрасту (до 19 лет, 20-39, 40-59, 60 лет и старше).

Решение: необходимо запустить программу Excel и создать новый лист. На этом листе требуется ввести заголовок каждой таблицы, текст «Шапки», строку «Итого» и столбец «Всего», выполнить форматирование макета таблицы приемами

объединения ячеек и переноса текста по словам, как показано ниже, назначить границы между ячейками и сохранить файл.

а) макет простой таблицы включает один признак, только табличное подлежащее (таблица 1).

Таблица 1

Распределение абсолютного числа больных по классам заболеваний

Класс заболеваний	Число больных
Болезни органов дыхания	
Инфекционные заболевания	
Травмы	
Прочие болезни	
Итого:	

б) макет групповой таблицы включает два связанных между собой признака, один - табличное подлежащее, другой – сказуемое (таблица 2).

Таблица 2

Распределение абсолютного числа больных по классам заболеваний с учетом социальных групп среди жителей района

Класс заболеваний	Число больных в группах			Всего в группах
	безработные	рабочие	служащие	
Болезни органов дыхания				
Инфекционные заболевания				
Травмы				
Прочие болезни				
Итого:				

в) макет комбинационной таблицы содержит три и более связанных между собой признака, один - табличное подлежащее, другие – сказуемые (таблица 3).

Таблица 3

Распределение абсолютного числа больных по классам заболеваний с учетом возрастных и социальных групп среди жителей района

Класс заболеваний	Число больных															Всего в группах				
	безработные					рабочие					служащие									
	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего
Болезни органов дыхания																				
Инфекционные заболевания																				
Травмы																				
Прочие болезни																				
Итого:																				

ЗАДАНИЯ

В программе Excel изготовьте макеты статистических таблиц.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «Таблицы» и выполните задачи варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки решения задач преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

Для изучения состава инвалидов Н-ского района требуется собрать данные о больных, прошедших медико-социальную экспертную комиссию (МСЭК), по группам инвалидности (I, II, III), причинам (общее заболевание, профессиональное, производственная травма, прочие причины) и занятости (не работает, продолжает работать). Составьте макеты простой, групповой и комбинационной таблиц.

Вариант 2

Изучается распределение среднего медицинского персонала города Волгограда по специальностям (фельдшеры, акушерки, медицинские сестры, лаборанты, рентгено техники, помощники санитарных врачей и пр.), стажу (до 5 лет, от 5 до 10 лет, свыше 10 лет), возрасту (до 19 лет, 20-29, 30-39, 40-49, 50-59, 60 лет и старше). Составьте макеты простой, групповой и комбинационной таблиц.

Вариант 3

Проводится сбор данных о распределении рабочих Н-ской фабрики, имевших временную нетрудоспособность в течение года, по ее видам (заболевание, травма, карантин, уход за больным, беременность и роды), цехам (ткацкий, прядильный, красильный), возрасту (до 19 лет, 20-29, 30-39, 40-49, 50-59, 60 лет и старше) и полу. Составьте макеты простой, групповой и комбинационной таблиц.

Вариант 4

Изучается распределение работающих Н-ского завода по цехам (механический, литейный, модельный, кузнечный и др.), стажу работы (до 5 лет, от 5 до 10 лет, более 10 лет), профессиям (слесари, токари, инструментальщики и другие). Составьте макеты простой, групповой и комбинационной таблиц.

Вариант 5

Производится сбор данных о распределении больных язвенной болезнью желудка и двенадцатиперстной кишки, находящихся на диспансерном наблюдении в Н-ской поликлинике, по длительности заболевания (до 1 года, от 1 года до 3 лет, свыше 3 лет), методам лечения (хирургический, консервативный), возрасту (до 19 лет, 20-39, 40-59, 60 лет и старше). Составьте макеты простой, групповой и комбинационной таблиц.

Вариант 6

Изучается число больных в терапевтическом отделении стационара, госпитализированных по поводу крупозной пневмонии, в зависимости от сроков госпитализации (в 1-й, 2-й, 3-й день болезни и позднее), длительности лечения (до 15 дней, свыше 15 дней), наличия осложнений (с осложнениями, без осложнений). Составьте макеты простой, групповой и комбинационной таблиц.

Вопросы для самоконтроля

1. Статистическая таблица (определение).
2. Из каких элементов состоят статистические таблицы?
3. Назовите типы статистических таблиц?
4. Перечислите правила создания макетов статистических таблиц.

Рекомендуемая литература:

1. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>
3. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.
4. Симонович С.В. Информатика. Базовый курс: Учебник для вузов. 3-е изд. Стандарт третьего поколения. – СПб.: Питер, 2011. – 640 с.: ил.

V. Относительные величины, динамические ряды

Для характеристики изучаемой статистической совокупности используются относительные величины, расчет которых проводится на третьем этапе статистического исследования.

Относительные величины (показатели, коэффициенты) весьма распространены и постоянно, наряду с абсолютными величинами, применяются в медицине и здравоохранении, прежде всего для сопоставления одной совокупности с другой. Вычисление относительных величин выполняется на основе абсолютных значений учетных признаков (абсолютных величин).

Абсолютные величины, полученные непосредственно при измерении учетных признаков каждой единицы наблюдения, сами по себе несут важную информацию о размере того или иного явления (количество родившихся, умерших; число коек в каждой больнице города, число дней болезни каждого больного и др.). Они служат основой для вычисления относительных величин, поскольку без преобразования имеют ограниченное познавательное значение.

Относительные величины вычисляются путем отношения (деления) одной абсолютной величины на другую с последующим умножением полученной дроби на требуемое основание (100, 1000, 10 000, 100 000 и т.д.). Смысл расчета относительных величин состоит в нахождении общей меры, приведении к общему знаменателю (основанию). При этом подбор того или иного множителя связан с тем, что относительные величины целесообразно представлять в целых числах, легко воспринимающихся при анализе. Соответственно, относительные показатели могут быть выражены в процентах (%), промилле (‰), продецимилле (‱), просантимилле (‱‱). Для их условного обозначения применяется знак “Р”.

Наиболее часто используются следующие ВИДЫ относительных показателей: интенсивные, экстенсивные, соотношения и наглядности.

ИНТЕНСИВНЫЕ ПОКАЗАТЕЛИ отображают распространенность (частоту, уровень) развития явления в своей среде, т.е. в среде, которая продуцирует это явление. Применяются они, чаще всего, в медицине и демографии, отвечая на вопрос: как часто явление встречается в известной среде? Интенсивные величины отражают ЧАСТОТУ (встречаемость) явления в СРЕДЕ, ПОРОДИВШЕЙ это явление и, как правило, вычисляются на основании 1000 (в промилле, ‰). Если полученное значение меньше единицы, целесообразно использование множителей 10 000 (‱), 100 000 (‱‱).

$$P_{\text{Интенс}} = \frac{\text{Абсолютный размер явления} * 1000}{\text{Абсолютный размер среды}}$$

ЭКСТЕНСИВНЫЕ ПОКАЗАТЕЛИ отражают структурный состав изучаемой совокупности. Они характеризуют отношение части статистической совокупности к совокупности в целом (долю, удельный вес, часть от целого), т.е. отношение отдельного элемента к итогу. Эти показатели нельзя использовать для определения уровня изучаемого явления, они демонстрируют только соотношение его частей и чаще выражаются в %.

$$P_{\text{Экт}} = \frac{\text{Абсолютный размер части явления} * 100}{\text{Абсолютный размер явления в целом}}$$

ПОКАЗАТЕЛИ СООТНОШЕНИЯ применяются, когда необходимо определить взаимоотношение не связанных между собой совокупностей. Например, обеспеченность населения больничными койками или врачами, соотношение средних медработников и врачей и др. Они вычисляются как отношение величины одного явления к другому явлению и выражаются, в основном, в проецимилле, реже - в процентах, промилле и др.

$$P_{\text{Соотн}} = \frac{\text{Абсолютный размер одного явления} * 10000}{\text{Абсолютный размер второго явления}}$$

Для анализа изменения изучаемого явления во времени вычисляются ДИНАМИЧЕСКИЕ ПОКАЗАТЕЛИ на основе динамического ряда.

Динамическим называется ряд чисел, состоящий из однородных сопоставимых величин, характеризующих изменения какого-либо явления за определенные отрезки времени.

Основными ВИДАМИ динамических рядов являются:

- простой (состоит из абсолютных величин);
- сложный (из относительных и средних);
- моментный (состоит из величин, характеризующих размеры явлений на определенные даты);
- интервальный (состоит из величин, характеризующих размеры явления за определенный интервал времени).

Числа динамического ряда называются УРОВНЕМ.

К ПОКАЗАТЕЛЯМ ДИНАМИЧЕСКОГО ряда относятся:

- АБСОЛЮТНЫЙ ПРИРОСТ - разность уровней изучаемого явления в данном и предыдущем периоде, может быть положительным, нулевым или отрицательным (убыль);
- ПОКАЗАТЕЛЬ НАГЛЯДНОСТИ - отношение каждого последующего уровня к исходному, принятому за 100%;
- ПОКАЗАТЕЛЬ РОСТА, или темп роста - отношение каждого последующего уровня к предыдущему уровню, принятому за 100%;
- ПОКАЗАТЕЛЬ ПРИРОСТА, или темп прироста - отношение абсолютного прироста к предыдущему уровню, принятому за 100%. Как и абсолютный прирост этот показатель может выражаться положительным (прирост), нулевым (стабильность) или отрицательным (убыль) значением.

При оценке показателей абсолютного прироста и темпа прироста имеет значение на знак полученной величины. Если показатель положительный, это указывает на прирост изучаемого явления. При его отрицательном значении говорят об убыли. В случае нулевого значения фиксируется стабильное состояние изучаемого явления.

Для анализа полученных относительных и динамических показателей необходимо не только уметь их рассчитывать, но и правильно выбирать графическое изображение с целью наглядного представления результатов исследования. Применение графического метода в статистическом исследовании делает изучаемые величины более доступными для понимания. При применении

графического метода важно знать, что содержание каждого показателя должно строго соответствовать виду графического изображения.

ИНТЕНСИВНЫЕ показатели, а также показатели СООТНОШЕНИЯ могут быть наглядно представлены в виде 4-х основных типов диаграмм: столбиковые, линейные, картограммы и картодиаграммы. Картограмма - это географическая карта с различной штриховкой. Картодиаграмма - это географическая карта с нанесенными на нее диаграммами.

ЭКСТЕНСИВНЫЕ показатели графически могут быть изображены секторной или внутрестолбиковой диаграммой.

Для графического изображения динамических показателей наиболее часто применяются линейная и столбиковая диаграммы. Радиальная (лепестковая) диаграмма является частным видом линейной диаграммы, построенной на полярных координатах. Она используется при необходимости изобразить графически динамику явления за замкнутый цикл времени.

Каждый рисунок диаграммы сопровождается названием, помещаемым под рисунком, порядковым номером и количественными характеристиками представленных явлений.

Пример вычисления относительных величин, заполнения статистических таблиц и графического отображения данных

Условие задачи: получены данные статистического наблюдения - абсолютные величины заболеваемости в р-не N за 2010 год.

1. Число детей, проживающих в районе N в 2010 году: 3500 человек.
Из них: мальчики - 1700 чел.;
 девочки - 1800 чел.
2. Число болевших детей на протяжении 2010 года: 2900 человек.
Из них: мальчиков - 1400 чел.;
 девочек - 1500 чел.
3. Общее число зарегистрированных заболеваний у детей за 2010 год: 3820 случаев.
Из них: мальчики - 1900 случаев;
 девочки - 1920 случаев.
4. Число заболеваний скарлатиной у детей за 2010 год: 66 случаев.
Из них: мальчики - 35 случаев;
 девочки - 31 случай.

Задание: а) вычислить показатели числа больных лиц, частоту случаев всех заболеваний и случаев скарлатины (интенсивные величины);

б) вычислить показатели доли (удельного веса) мальчиков и девочек среди детей района и долю заболеваний скарлатиной во всех заболеваниях детей, а также среди мальчиков и девочек (экстенсивные);

в) построить статистическую таблицу, содержащую интенсивные показатели;

г) изобразить экстенсивные показатели секторной и внутрестолбиковой диаграммами.

Решение: необходимо запустить программу Excel и создать новый лист. На этом листе ввести условия задачи и решение, как показано на рис. 4, сохранить файл.

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Получены данные статистического наблюдения - абсолютные величины заболеваемости в р-не N за 2010 год.												
3	1. Число детей, проживающих в районе N в 2010 году:						3 500	человек.					
4	Из них:	мальчики	-	1 700	чел.;								
5		девочки	-	1 800	чел.								
6	2. Число болевших детей на протяжении 2010 года:						2 900	человек.					
7	Из них:	мальчиков	-	1 400	чел.;								
8		девочек	-	1 500	чел.								
9	3. Общее число зарегистрированных заболеваний у детей за 2010 год:						3 820	случаев.					
10	Из них:	мальчики	-	1 900	случаев;								
11		девочки	-	1 920	случаев.								
12	4. Число заболеваний скарлатиной у детей за 2010 год:						66	случаев.					
13	Из них:	мальчики	-	35	случаев;								
14		девочки	-	31	случай.								
15	Задание: а) вычислить показатели числа больных лиц, частоту случаев всех заболеваний и случаев скарлатины (интенсивные величины); б) вычислить показатели доли (удельного веса) мальчиков и девочек среди детей района и долю заболеваний скарлатиной во всех заболеваниях детей, а также среди мальчиков и девочек (экстенсивные); в) построить статистическую таблицу, содержащую интенсивные показатели; г) изобразить экстенсивные показатели секторной и внутристолбиковой диаграммами.												
16													
17	Решение:												
18	а) вычисление показателей числа больных лиц, частоты случаев всех заболеваний и случаев скарлатины (интенсивные величины)												
19	$P_{\text{бол.лиц всего}} = (2900/3500)*1000 =$				828,6	‰		$P_{\text{случ.}} = (3820/3500)*1000 =$				1091,4	‰
20	$P_{\text{бол.мальчики}} = (1400/1700)*1000 =$				823,5	‰		$P_{\text{случ.мальчики}} = (1900/1700)*1000 =$				1117,6	‰
21	$P_{\text{бол.девочки}} = (1500/1800)*1000 =$				833,3	‰		$P_{\text{случ.девочки}} = (1920/1800)*1000 =$				1066,7	‰
22													
23	$P_{\text{случ.скарл.}} = (66/3500)*1000 =$				18,9	‰							
24	$P_{\text{случ.у мальчиков}} = (35/1700)*1000 =$				20,6	‰							
25	$P_{\text{случ.у девочек}} = (31/1800)*1000 =$				17,2	‰							

Рис. 4. Пример вычисления относительных величин, а также их наглядного изображения в программе Excel.

Вычисления выполняются с помощью формул, в которых используются адреса ячеек, содержащие абсолютные величины.

а) вычисление интенсивных показателей числа больных лиц, случаев всех заболеваний и заболеваний скарлатиной. Они отражают ЧАСТОТУ (встречаемость) явления в среде, ПОРОДИВШЕЙ это явление. Числитель – явление, знаменатель – среда, основание 1000.

$$P_{\text{Интенс}} = \frac{\text{Абсолютный размер явления} * 1000}{\text{Абсолютный размер среды}}$$

В программе Excel вычисления производятся с помощью формул, которые могут включать абсолютные значения или адреса ячеек, в которые введены данные условия задачи. Формулы расчета интенсивных показателей:

$$P_{\text{бол.лиц всего}} = (2900/3500)*1000 = 828,6\text{‰}$$

$$P_{\text{бол.мальчики}} = (1400/1700)*1000 = 823,5\text{‰}$$

$$P_{\text{бол.девочки}} = (1500/1800)*1000 = 833,3\text{‰}$$

$$P_{\text{случ.}} = (3820/3500)*1000 = 1091,4\text{‰}$$

$$P_{\text{случ.мальчики}} = (1900/1700)*1000 = 1117,6\text{‰}$$

$$P_{\text{случ.девочки}} = (1920/1800)*1000 = 1066,7\text{‰}$$

$$P_{\text{случ.скарл.}} = (66/3500)*1000 = 18,9\text{‰}$$

$$P_{\text{случ.у мальчиков}} = (35/1700)*1000 = 20,6\text{‰}$$

$$P_{\text{случ.у девочек}} = (31/1800)*1000 = 17,2\text{‰}$$

б) вычисление экстенсивных показателей отражающих долю (УДЕЛЬНЫЙ ВЕС, часть явления в этом же явлении) числа мальчиков и девочек, проживающих в изучаемом районе, долю заболеваний скарлатиной в общей заболеваемости детей, а также среди мальчиков и девочек соответственно. Числитель – часть явления, знаменатель – всё явление, основание 100 (выражается в %).

$$P_{\text{Экт}} = \frac{\text{Абсолютный размер части явления} * 100}{\text{Абсолютный размер явления в целом}}$$

Формулы расчета интенсивных показателей:

$$P_{\text{доля мальчиков}} = (1700/3500) * 100 = 48,6\%$$

$$P_{\text{уд.вес случаев скарлатины}} = (66/3820) * 100 = 1,7\%$$

$$P_{\text{доля девочек}} = (1800/3500) * 100 = 51,4\%$$

$$P_{\text{уд.вес скарлатины среди мальчиков}} = (35/1900) * 100 = 1,8\%$$

$$P_{\text{уд.вес скарлатины среди девочек}} = (31/1920) * 100 = 1,6\%$$

в) заполнение статистической таблицы интенсивных относительных величин (таблица 4).

Таблица 4

Распространенность заболеваний среди детей в 2010 году

Пол	Число болевших на 1000 детей (‰)	Число случаев заболеваний на 1000 детей (‰)	Число случаев скарлатины на 1000 детей (‰)
Мальчики	823,5	1117,6	17,2
Девочки	833,3	1066,7	20,6
Итого:	828,6	1091,4	18,9

г) создание секторной (рис. 5) и внутрисклонной (рис. 6) диаграмм, наглядно демонстрирующих экстенсивные показатели. В программе Excel изготовление таких рисунков производится командой «Вставка» с предварительным выделением диапазона ячеек содержащих данные для изображения. Значение изображаемых величин могут быть взяты из ячеек, в которых вычислены доли мальчиков и девочек, проживающих в изучаемом районе.

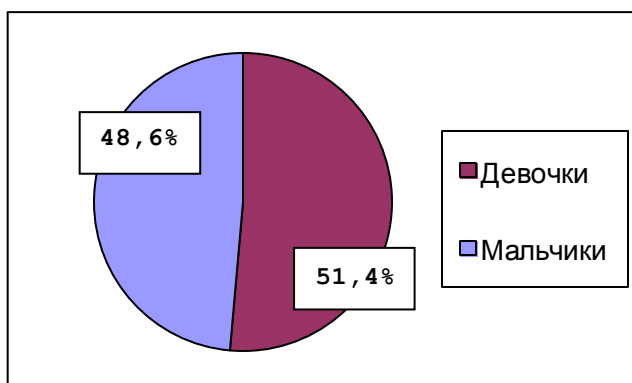


Рис. 5. Доля мальчиков и девочек, проживающих в районе N. Секторная диаграмма.

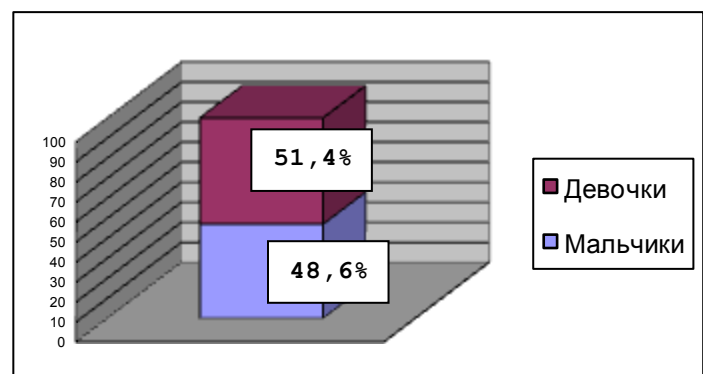


Рис. 6. Доля мальчиков и девочек, проживающих в районе N. Внутрисклонная диаграмма.

Пример расчета и анализа показателей первичной заболеваемости населения и структуры заболеваемости

Условие задачи: получены данные статистического наблюдения: среднегодовая численность жителей города N в 2010 году составляла 300 тысяч человек. За год зарегистрировано 216 850 случаев заболеваний, в том числе по группам: заболевания органов дыхания – 92 000 случаев, болезни органов кровообращения – 82 000 случаев, травмы – 16 500 случаев, новообразования – 16 000 случаев, инфекционные болезни - 9000 случаев, прочие заболевания – 1350.

Задание: составьте статистическую таблицу, вычислите интенсивные показатели частоты и экстенсивные показатели структуры заболеваемости жителей района N за 2010 год. Отобразите эти показатели соответствующими диаграммами. Сделайте вывод.

Решение: необходимо запустить программу Excel, ввести условие задачи, создать макет статистической таблицы, ввести данные и формулы вычислений, как показано на рис.7, сохранить файл. С целью наглядного изображения и анализа относительных величин требуется создать рисунки диаграмм соответствующие типу показателей.

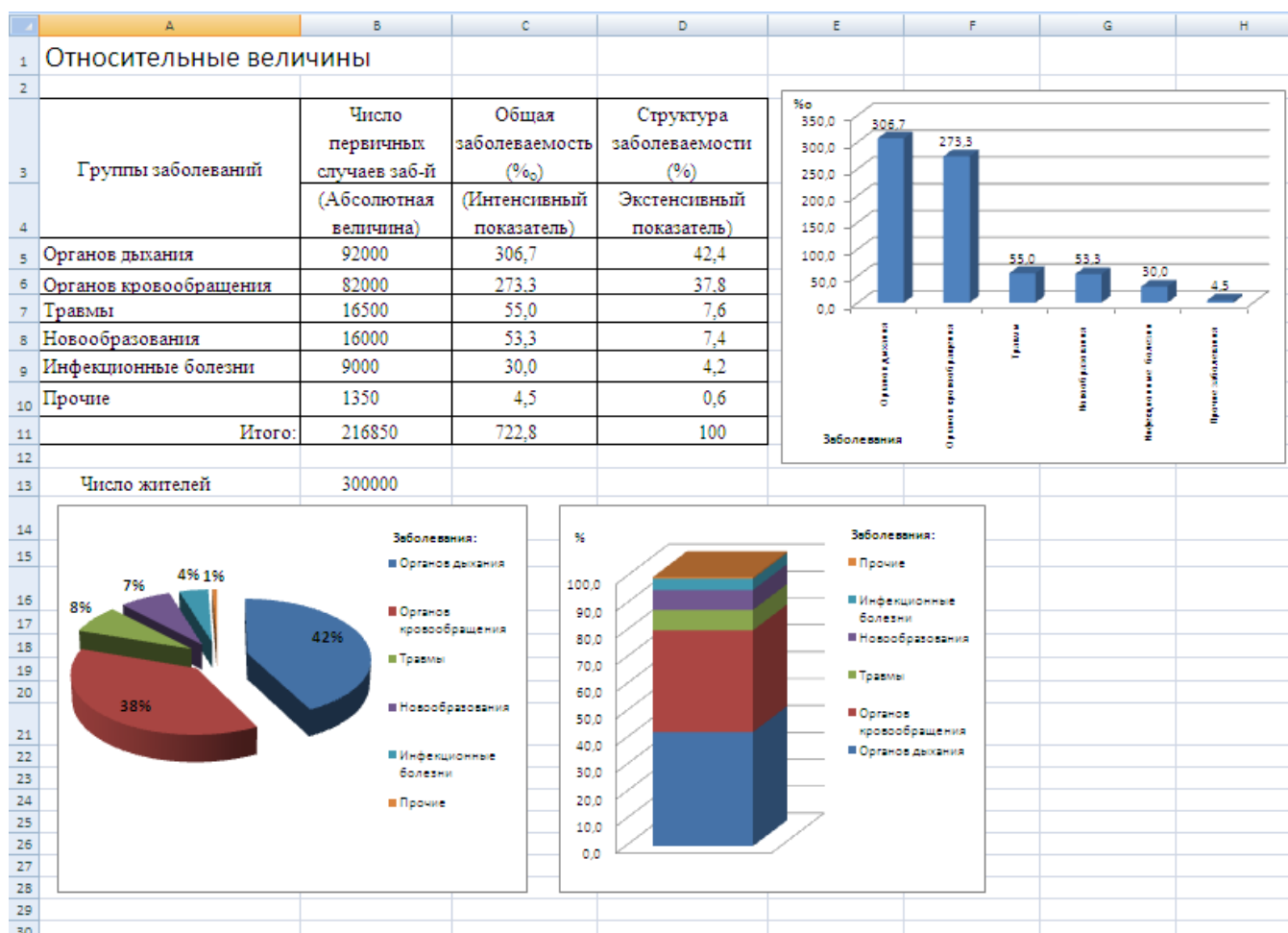


Рис.7. Вычисление относительных величин в программе Excel.

Для вычисления относительных величин необходимо занести абсолютные данные из условия задачи в отдельные ячейки таблицы, а затем, в соответствующих ячейках поместить расчетные формулы. Правила и результаты вычислений интенсивных и экстенсивных показателей заболеваемости представлен в таблице 5, а их графическое изображение на рис. 8 и рис. 9.

Таблица 5

Показатели заболеваемости населения города N за 2010 год

Группы заболеваний	Общая заболеваемость	Структура заболеваемости
	(‰) интенсивный показатель	(%) экстенсивный показатель
Органов дыхания	306,7 =92000/300000*1000	42,4 =92000/216850*100
Органов кровообращения	273,3 =82000/300000*1000	37,8 =82000/216850*100
Травмы	55,0 =16500/300000*1000	7,6 =16500/216850*100
Новообразования	53,3 =16000/300000*1000	7,4 =16000/216850*100
Инфекционные болезни	30,0 =9000/300000*1000	4,2 =9000/216850*100
Прочие	4,5 =1350/300000*1000	0,6 =1350/216850*100
Итого:	722,8 =216850/300000*1000	100

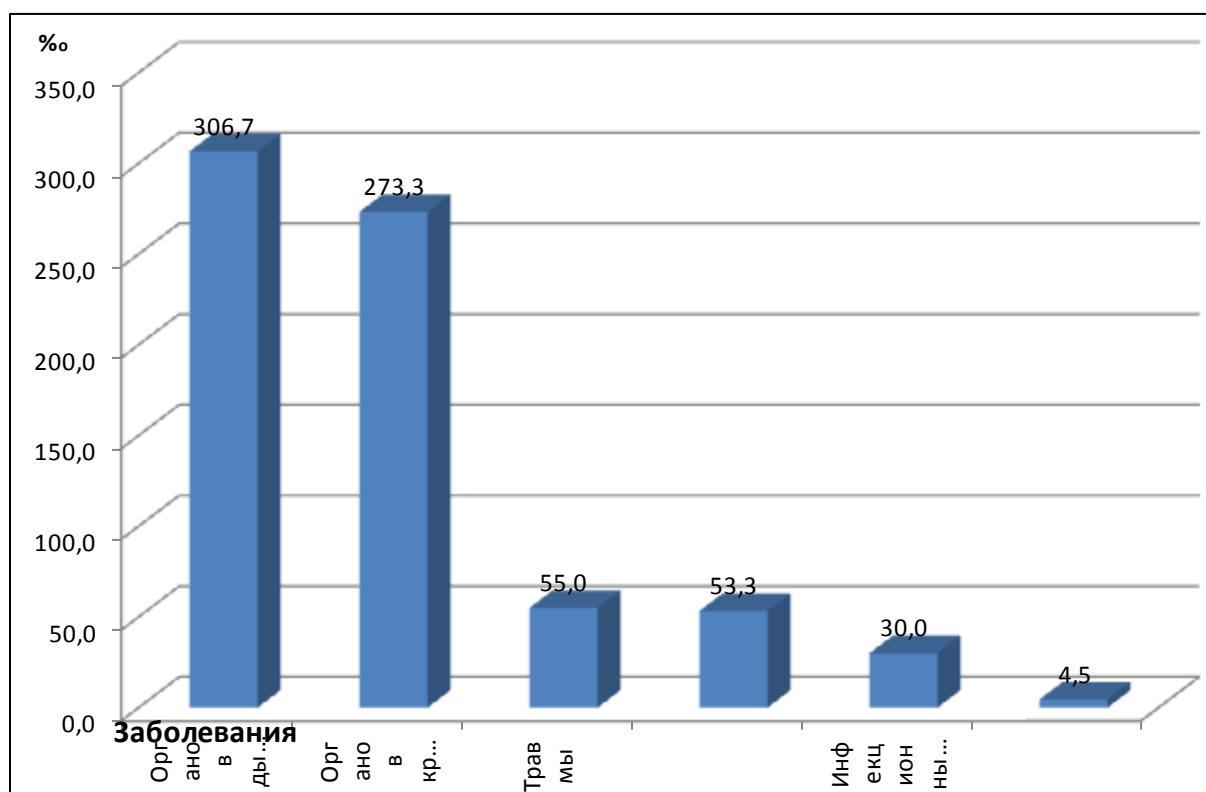


Рис. 8. Уровень заболеваемости по группам болезней жителей города N в 2010 году (интенсивные показатели, ‰).

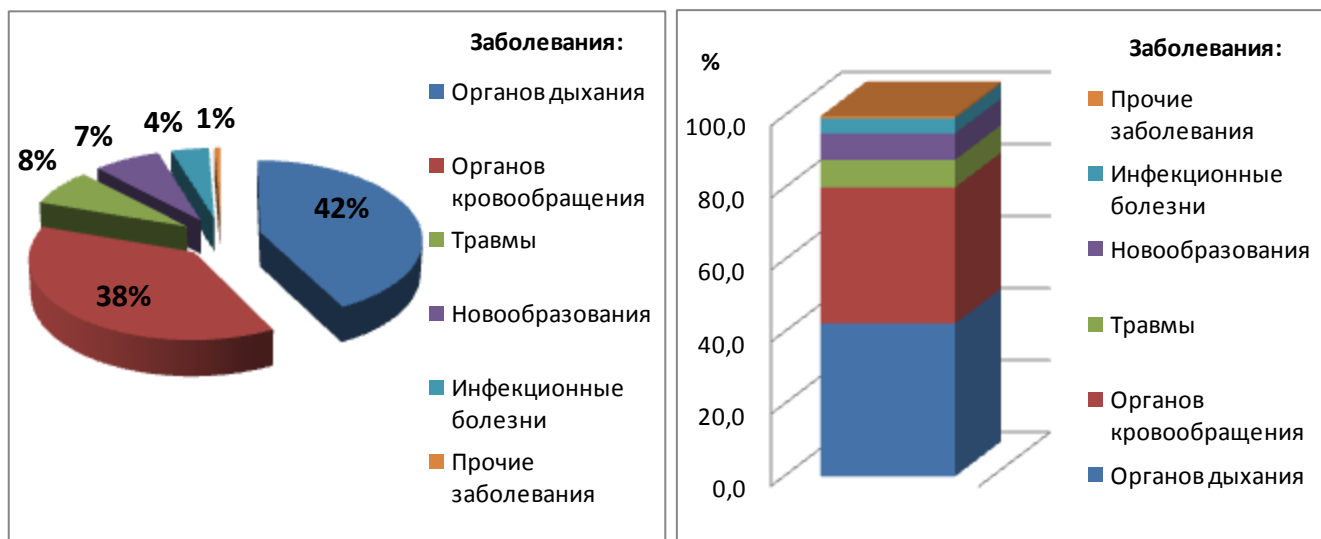


Рис. 9. Структура заболеваемости жителей города N в 2010 году (экстенсивные показатели, %).

Вывод: заболеваемость в районе N характеризуется повышенным уровнем болезней органов дыхания и сердечно-сосудистой системы по сравнению с другими классами заболеваний, причем их доли в структуре заболеваемости приблизительно равны.

Пример расчета и анализа показателей динамики численности студентов в Волгоградской области

Условие задачи: получены данные о численности студентов и населения Волгоградской области с 2004 по 2009 годы. В 2004 году число студентов составило 25 000 человек, в 2005 – 26 300 человек, в 2006 – 27 000, в 2007 – 26 800 человек, в 2008 – 28 300 в 2009 – 26 100 человек.

Задание: вычислить показатели динамики числа студентов, представить их графическим изображением, проанализировать и сделать вывод.

Решение: необходимо запустить программу Excel и на отдельном листе введите данные и решение задачи, как показано ниже, сохранить файл.

В программе Excel для вычисления динамических величин необходимо занести абсолютные данные из условия задачи в отдельные ячейки таблицы, а затем изготовить макет статистической таблицы и в соответствующих ячейках этого макета ввести расчетные формулы. Принцип вычисления динамических показателей

представлен в таблице 6, а графическое изображение динамики на основе показателя наглядности демонстрирует рисунок 10.

Таблица 6

Динамика количества студентов в Волгоградской области за 6 лет

Годы	Абсолютное значение числа студентов (чел.)	Динамика числа студентов в показателях:			
		абсолютного прироста или убыли (чел.)	наглядности (%)	роста/ убыли (%)	прироста/ убыли (%)
2004	25 000	-	100,0	-	-
2005	26 300	1300 =26300-25000	105,2 =26300/25000 *100	105,2 =26300/25000*100	5,2 =1300/25000*100
2006	27 000	700 =27000-26300	108,0 =27000/25000 *100	102,7 =27000/26300*100	2,7 =700/26300*100
2007	26 800	-200	107,2	99,3	-0,7
2008	28 300	1 500	113,2	105,6	5,6
2009	26 100	-2 200	104,4	92,2	-7,8

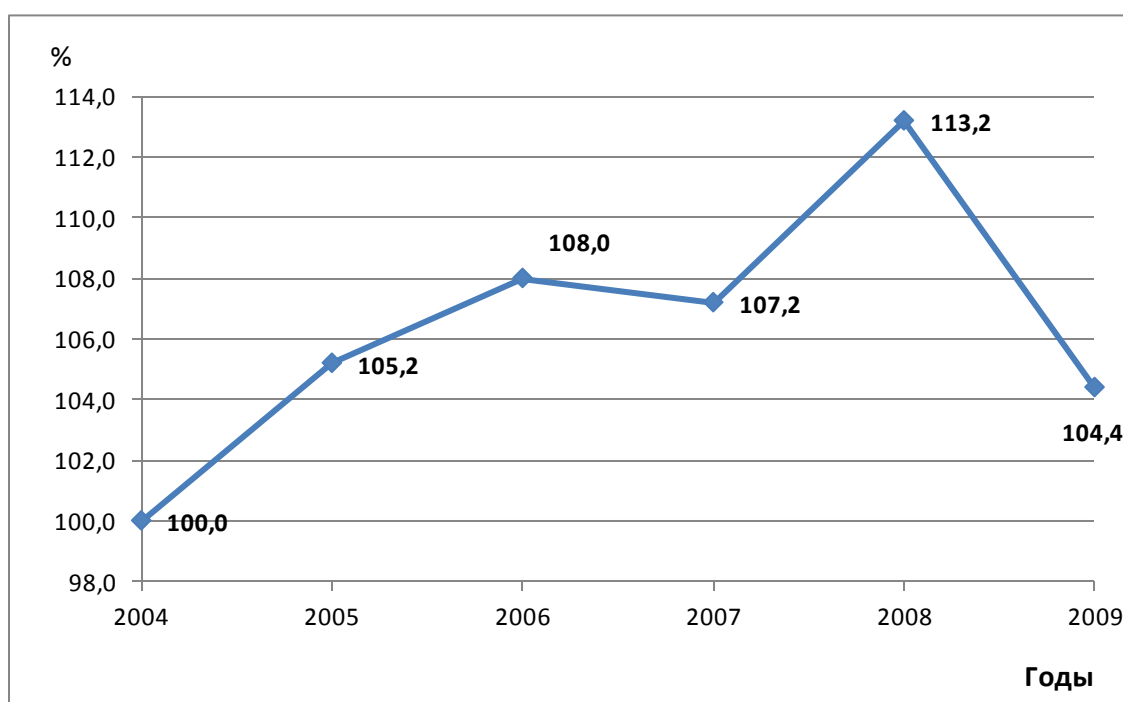


Рис. 10. Динамика числа студентов в Волгоградской области с 2004 по 2009 годы (показатель наглядности, %).

Вывод: изменение относительного числа студентов в Волгоградской области за 6 лет (с 2004 по 2009 годы) характеризуется увеличением до 2008 года с последующей тенденцией к снижению до уровня 2005 года.

ЗАДАНИЯ

В программе Excel выполните вычисление, изобразите диаграммами и проанализируйте относительные величины.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «ОтнВеличины» и выполните задачи варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки решения задач преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

- а) в районе города N с населением 40 тысяч человек в 2010 году впервые зарегистрировано 35600 случаев заболеваний, в том числе заболеваний органов дыхания 14500 случаев, болезней органов кровообращения 8000 случаев, травмы 9100 случаев, новообразования 1350 случаев, инфекционные заболевания 1500 случаев прочие заболевания 1150 случаев. Составьте простую статистическую таблицу. Вычислите интенсивные показатели заболеваемости на 1000 жителей. Определите структуру заболеваемости по группам болезней. Представьте графическое изображение интенсивных и экстенсивных показателей заболеваемости населения. Сформулируйте вывод.
- б) вычислите динамические показатели на основе данных таблицы 7, отобразите их графически, проанализируйте и сделайте вывод.

Таблица 7

Динамика заболеваемости населения района с 2006 по 2010 год

Годы	Число впервые зарегистрированных заболеваний	Динамика заболеваемости населения в показателях:			
		абсолютного прироста/убыли (чел.)	наглядности (%)	роста (%)	прироста (%)
2006	34500		100	-	-
2007	34600				
2008	34900				
2009	35000				
2010	35600				

Вариант 2

- а) в городе M с численностью взрослого населения 200 тысяч человек в 2010 году впервые установлена инвалидность 16300 жителям старше 18 лет, в том числе

инвалидов 3-й группы 7500 человек, 2-й группы – 6200 человек, 1-й группы – 2600 человек. Составьте простую статистическую таблицу. Вычислите интенсивные показатели инвалидности на 10 тысяч жителей. Определите возрастную структуру инвалидности. Представьте графическое изображение интенсивных и экстенсивных показателей инвалидности. Сформулируйте вывод.

- б) вычислите динамические показатели на основе данных таблицы 8, отобразите их графически, проанализируйте и сделайте вывод.

Таблица 8

Динамика инвалидности взрослого населения города М с 2006 по 2010 год

Годы	Число инвалидов	Динамика инвалидности населения в показателях:			
		абсолютного прироста/убыли (чел.)	наглядности (%)	роста (%)	прироста (%)
2006	15300		100	-	-
2007	15500				
2008	16000				
2009	16200				
2010	16300				

Вариант 3

- а) в одном из районов города А с населением в 20 тысяч человек в 2010 году зарегистрировано 5650 обращений к врачам-эндокринологам, из них по поводу заболеваний щитовидной железы 1950 обращений, сахарного диабета 1330 обращений, нарушение питания и обмена веществ - 1420, других эндокринных заболеваний - 950 обращений. Составьте простую статистическую таблицу. Вычислите интенсивные показатели на тысячу жителей. Определите структуру болезненности по группам заболеваний. Представьте графическое изображение интенсивных и экстенсивных показателей болезненности. Сформулируйте вывод.
- б) вычислите динамические показатели на основе данных таблицы 9, отобразите их графически, проанализируйте и сделайте вывод.

Таблица 9

Динамика обращаемости к врачам эндокринологам района с 2006 по 2010 год

Годы	Число обращений	Динамика обращаемости населения в показателях:			
		абсолютного прироста/убыли (чел.)	наглядности (%)	роста (%)	прироста (%)
2006	3210		100	-	-
2007	3185				
2008	3180				
2009	4190				
2010	5650				

Вариант 4

- а) в городе М с населением 500850 человек в 2010 году родилось живыми 5 870 детей, из них в возрасте до одного года умерло 103 ребенка, в том числе 8

смертельных исходов зарегистрировано от инфекционных и паразитарных болезней, от болезней органов дыхания умерло 13 детей, от врожденных аномалий и пороков развития - 25 детей, от состояний, возникающих в перинатальном периоде - 45 детей, от несчастных случаев - 5, от прочих заболеваний - 7. Составьте простую статистическую таблицу. Вычислите интенсивные показатели младенческой смертности на тысячу населения. Определите структуру смертности по группам заболеваний. Представьте графическое изображение показателей младенческой смертности. Сформулируйте вывод.

- б) вычислите динамические показатели на основе данных таблицы 10, отобразите их графически, проанализируйте и сделайте вывод.

Таблица 10

Динамика заболеваемости новорожденных в городе М с 2006 по 2010 год

Годы	Число детей первого года жизни, заболевших в течение года	Динамика заболеваемости новорожденных в показателях:			
		абсолютного прироста/убыли (чел.)	наглядности (%)	роста (%)	прироста (%)
2006	121		100	-	-
2007	122				
2008	112				
2009	110				
2010	103				

Вариант 5

- а) в города А с населением 1 100 450 человек в 2010 году впервые зарегистрировано 901 020 случаев заболеваний, в том числе заболеваний органов дыхания 200 300 случаев, болезней органов кровообращения 178 120 случаев, травм - 114 930 случаев, новообразований - 93520 случаев, прочих заболеваний - 120450 случаев. Составьте простую статистическую таблицу. Вычислите интенсивные показатели заболеваемости на 1000 жителей. Определите структуру заболеваемости по группам болезней. Представьте графическое изображение интенсивных и экстенсивных показателей заболеваемости. Сформулируйте вывод.
- б) вычислите динамические показатели заболеваемости на основе данных таблицы 11, отобразите их графически, проанализируйте и сделайте вывод.

Таблица 11

Динамика заболеваемости населения города А с 2006 по 2010 год

Годы	Число впервые зарегистрированных заболеваний	Динамика заболеваемости населения в показателях:			
		абсолютного прироста/убыли (чел.)	наглядности (%)	роста (%)	прироста (%)
2006	890 000		100	-	-
2007	878 000				
2008	890 000				
2009	892 340				
2010	901 020				

Вариант 6

- а) в городе М с населением 1 500 850 человек в 2010 году родилось живыми 13 450 детей, из них в возрасте до одного года умерло 245 детей, в том числе 20 смертельных исходов зарегистрировано от инфекционных и паразитарных болезней, от болезней органов дыхания умерло 32 ребенка, от врожденных аномалий и пороков развития - 59 детей, от состояний, возникающих в перинатальном периоде - 95 детей, от несчастных случаев - 15, от прочих заболеваний - 24. Составьте простую статистическую таблицу. Вычислите интенсивные показатели младенческой смертности на 1 тысячу населения. Определите структуру младенческой смертности по группам заболеваний. Представьте графическое изображение интенсивных и экстенсивных показателей. Сформулируйте вывод.
- б) вычислите динамические показатели на основе данных таблицы 12, отобразите их графически, проанализируйте и сделайте вывод.

Таблица 12

Динамика младенческой смертности в городе М с 2006 по 2010 год

Годы	Число детей, умерших на первом году жизни	Динамика младенческой смертности в показателях:			
		абсолютного прироста/убыли (чел.)	наглядности (%)	роста (%)	прироста (%)
2006	140		100	-	-
2007	145				
2008	132				
2009	125				
2010	110				

Вопросы для самоконтроля

1. Для чего применяются относительные величины?
2. Назовите основные типы относительных величин и методику их вычисления?
3. Принцип вычисления интенсивных показателей?
4. Как вычисляются экстенсивные относительные величины?
5. Особенности вычисления показателей соотношения?
6. Дайте определение динамического ряда.
7. Назовите основные показатели динамического ряда.
8. Для чего применяют графические изображения?
9. Перечислите основные виды графических изображений.
10. Какой тип показателей применяется для построения внутрестолбиковой диаграммы?
11. Какие показатели принято изображать секторной диаграммой?
12. На основе каких показателей изготавливают столбиковые диаграммы?

13. Какие показатели изображают линейной диаграммой?

Рекомендуемая литература:

1. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>
3. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.
4. Симонович С.В. Информатика. Базовый курс: Учебник для вузов. 3-е изд. Стандарт третьего поколения. – СПб.: Питер, 2011. – 640 с.: ил.

VI. Вариационные ряды, средние величины, вариабельность признака

Особое место в статистическом анализе принадлежит определению среднего уровня изучаемого признака или явления. Средний уровень признака измеряют средними величинами.

Средняя величина характеризует общий количественный уровень изучаемого признака и является групповым свойством статистической совокупности. Она нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и выдвигает на первый план основное, типичное свойство изучаемого признака.

Средние величины широко используются:

1. Для оценки состояния здоровья населения: характеристики физического развития (рост, вес, окружность грудной клетки и пр.), выявления распространенности и длительности различных заболеваний, анализа демографических показателей (естественного движения населения, средней продолжительности предстоящей жизни, воспроизводства населения, средней численности населения и др.).
2. Для изучения деятельности лечебно-профилактических учреждений, медицинских кадров и оценки качества их работы, планирования и определения потребности населения в различных видах медицинской помощи (среднее число обращений или посещений на одного жителя в год, средняя длительность пребывания больного в стационаре, средняя продолжительность обследования больного, средняя обеспеченность врачами, койками и пр.).
3. Для характеристики санитарно-эпидемиологического состояния (средняя запыленность воздуха в цехе, средняя площадь на одного человека, средние нормы потребления белков, жиров и углеводов и т. д.).
4. Для определения медико-физиологических показателей в норме и патологии, при обработке лабораторных данных, для установления достоверности результатов выборочного исследования в социально-гигиенических, клинических, экспериментальных исследованиях.

Вычисление средних величин выполняется на основе вариационных рядов. **ВАРИАЦИОННЫЙ РЯД** – это однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления.

Количественная вариация может быть двух типов: прерывная (дискретная) и непрерывная.

Прерывный (дискретный) признак выражается только целым числом и не может иметь никаких промежуточных значений (например, число посещений, численность населения участка, число детей в семье, степень тяжести болезни в баллах и др.).

Непрерывный признак может принимать любые значения в определенных пределах, в том числе и дробные, и выражается лишь приближенно (например, вес – для взрослых можно ограничиться килограммами, а для новорожденных – граммами; рост, артериальное давление, время, потраченное на прием больного, и т. д.).

Цифровое значение каждого отдельного признака или явления, входящего в вариационный ряд, называется вариантом и обозначается буквой V . В математической литературе встречаются и другие обозначения, например x или y .

Вариационный ряд, где каждая варианта указана один раз, называется ПРОСТЫМ. Такие ряды используются в большинстве статистических задач в случае компьютерной обработки данных.

При увеличении числа наблюдений, как правило, встречаются повторяющиеся значения вариант. В этом случае создается СТРУППИРОВАННЫЙ ВАРИАЦИОННЫЙ РЯД, где указывается число повторений (частота, обозначается буквой « p »).

РАНЖИРОВАННЫЙ ВАРИАЦИОННЫЙ РЯД состоит из вариант, расположенных в порядке возрастания или убывания. Как простой, так и сгруппированный ряды могут быть составлены с ранжированием.

ИНТЕРВАЛЬНЫЙ ВАРИАЦИОННЫЙ РЯД составляют с целью упрощения последующих вычислений, выполняемых без использования компьютера, при очень большом числе единиц наблюдения (более 1000).

НЕПРЕРЫВНЫЙ ВАРИАЦИОННЫЙ РЯД включает значения вариант, которые могут выражаться любыми значениями.

Если в вариационном ряде значения признака (варианты) заданы в виде отдельных конкретных чисел, то такой ряд называют ДИСКРЕТНЫМ.

Общими характеристиками значений признака, отражаемого в вариационном ряду, являются средние величины. Среди них наиболее применяемые: средняя арифметическая величина M , мода M_o и медиана M_e . Каждая из этих характеристик своеобразна. Они не могут подменить друг друга и лишь в совокупности достаточно полно и в сжатой форме представляют собой особенности вариационного ряда.

МОДОЙ (M_o) называют значение наиболее часто встречающейся варианты.

МЕДИАНА (M_e) – это значение варианты, делящей ранжированный вариационный ряд пополам (с каждой стороны медианы находится половина вариант). В редких случаях, когда имеется симметричный вариационный ряд, мода и медиана равны между собой и совпадают со значением средней арифметической.

Наиболее типичной характеристикой значений вариант является средняя арифметическая величина (M). В математической литературе она обозначается \bar{X} .

Средняя арифметическая величина (M, \bar{X}) – это общая количественная характеристика определенного признака изучаемых явлений, составляющих качественно однородную статистическую совокупность. Различают среднюю арифметическую простую и взвешенную. Средняя арифметическая простая вычисляется для простого вариационного ряда путем суммирования всех вариант и делением этой суммы на общее количество вариант, входящих в данный вариационный ряд. Вычисления проводятся по формуле:

$$M = \frac{\sum V}{n},$$

где: M - средняя арифметическая простая;

$\sum V$ - сумма вариант;

n - число наблюдений.

В сгруппированном вариационном ряду определяют взвешенную среднюю арифметическую. Формула ее вычисления:

$$M = \frac{\sum Vp}{n},$$

где: M - средняя арифметическая взвешенная;

$\sum Vp$ - сумма произведений вариант на их частоты;

n - число наблюдений.

При большом числе наблюдений и в случае ручных вычислений может применяться способ моментов. Независимо от выбранного способа вычислений, полученные значение средней арифметической будут одинаковы.

Средняя арифметическая имеет следующие свойства:

- сумма отклонений вариант от средней ($\sum d$) равна нулю (см. табл. 15);
- при умножении (делении) всех вариант на один и тот же множитель (делитель) средняя арифметическая умножается (делится) на тот же множитель (делитель);
- если прибавить (вычесть) ко всем вариантам одно и то же число, средняя арифметическая увеличивается (уменьшается) на это же число.

Средние арифметические величины, взятые сами по себе, без учета вариабельности рядов, из которых они вычислены, могут не в полной мере отражать свойства вариационного ряда, в особенности когда необходимо сопоставление с другими средними. Близкие по значению средние могут быть получены из рядов с различной степенью рассеяния. Чем ближе друг к другу отдельные варианты по своей количественной характеристике, тем меньше РАССЕЯНИЕ (КОЛЕБЛЕМОСТЬ, ВАРИАБЕЛЬНОСТЬ) ряда, тем типичнее его средняя.

Основными параметрами, которые позволяют оценить вариабельность признака, являются:

- Размах;
- Амплитуда;
- Среднее квадратическое отклонение;
- Коэффициент вариации.

Приблизительно о колеблемости признака можно судить по размаху и амплитуде вариационного ряда. Размах указывает на максимальную (V_{\max}) и минимальную (V_{\min}) варианты в ряду. Амплитуда (A_m) является разностью этих вариант: $A_m = V_{\max} - V_{\min}$.

Основной, общепринятой мерой колеблемости вариационного ряда являются ДИСПЕРСИЯ (D). Но наиболее часто применяется более удобный параметр, вычисляемый на основе дисперсии - среднее квадратическое отклонение (σ). Оно учитывает величину отклонения (d) каждой варианты вариационного ряда от его средней арифметической ($d = V - M$).

Поскольку отклонения вариант от средней могут быть положительными и отрицательными, то при суммировании они дают значение «0» ($\sum d = 0$). Чтобы избежать этого, величины отклонения (d) возводятся во вторую степень и

усредняются. Таким образом, дисперсия вариационного ряда является средним квадратом отклонений вариант от средней арифметической и вычисляется по формуле:

$$D = \frac{\sum d^2}{n}.$$

Она является важнейшей характеристикой вариабельности и применяется для вычисления многих статистических критериев.

Поскольку дисперсия выражается квадратом отклонений, ее величина не может использоваться в сопоставлении со средней арифметической. Для этих целей применяется СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ, которое обозначается знаком «Сигма» (σ). Оно характеризует среднее отклонение всех вариант вариационного ряда от средней арифметической величины в тех же единицах, что и сама средняя величина, поэтому они могут использоваться совместно.

Среднее квадратическое отклонение определяют по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}}.$$

Указанная формула применяется при числе наблюдений (n) больше 30. При меньшем числе n значение среднего квадратического отклонения будет иметь погрешность, связанную с математическим смещением ($n - 1$). В связи с этим, более точный результат может быть получен с помощью учета такого смещения в формуле расчета стандартного отклонения:

$$s = \sqrt{\frac{\sum d^2}{n-1}}.$$

СТАНДАРТНОЕ ОТКЛОНЕНИЕ (s) – это оценка среднеквадратического отклонения случайной величины X относительно её математического ожидания на основе несмещённой оценки её дисперсии.

При значениях $n > 30$ среднее квадратическое отклонение (σ) и стандартное отклонение (s) будут одинаковыми ($\sigma = s$). Поэтому в большинстве практических пособий эти критерии рассматриваются как равнозначные. В программе Excel вычисление стандартного отклонения может быть выполнено функцией =СТАНДОТКЛОН(диапазон). А с целью расчета среднего квадратического отклонения требуется создать соответствующую формулу.

Среднее квадратическое или стандартное отклонение позволяет определить, насколько значения признака могут отличаться от среднего значения. Предположим, существуют два города с одинаковой средней дневной температурой в летний период. Один из этих городов расположен на побережье, а другой на континенте. Известно, что в городах, расположенных на побережье, различия дневных температур меньше, чем у городов, расположенных внутри континента. Поэтому среднее квадратическое отклонение дневных температур у прибрежного города будет меньше, чем у второго города. На практике это означает, что средняя температура воздуха каждого конкретного дня в городе, расположенного на континенте будет сильнее отличаться от среднего значения, чем в городе на побережье. Кроме того стандартное отклонение позволяет оценить возможные отклонения температуры от средней с требуемым уровнем вероятности.

Согласно теории вероятности, в явлениях, подчиняющихся нормальному закону распределения, между значениями средней арифметической, среднего квадратического отклонения и вариантами существует строгая зависимость (ПРАВИЛО ТРЕХ СИГМ). Например, 68,3% значений варьирующего признака находятся в пределах $M \pm 1\sigma$, 95,5% — в пределах $M \pm 2\sigma$ и 99,7% — в пределах $M \pm 3\sigma$.

Величина среднего квадратического отклонения позволяет судить о характере однородности вариационного ряда и исследуемой группы. Если величина среднего квадратического отклонения небольшая, то это свидетельствует о достаточно высокой однородности изучаемого явления. Среднюю арифметическую в таком случае следует признать вполне характерной для данного вариационного ряда. Однако слишком малая величина сигмы заставляет думать об искусственном подборе наблюдений. При очень большой сигме средняя арифметическая в меньшей степени характеризует вариационный ряд, что говорит о значительной вариабельности изучаемого признака или явления или о неоднородности исследуемой группы. Однако сопоставление величины среднего квадратического отклонения возможно только для признаков одинаковой размерности. Действительно, если сравнивать разнообразие веса новорожденных детей и взрослых, мы всегда получим более высокие значения сигмы у взрослых.

Сравнение вариабельности признаков различной размерности может быть выполнено с помощью КОЭФФИЦИЕНТА ВАРИАЦИИ. Он выражает разнообразие в процентах от средней величины, что позволяет производить сравнение различных признаков. Коэффициент вариации в медицинской литературе обозначается знаком «С», а в математической «v» и вычисляется по формуле:

$$C = \frac{\sigma}{M} \times 100\% .$$

Значения коэффициента вариации менее 10% свидетельствует о малом рассеянии, от 10 до 20% – о среднем, более 20% – о сильном рассеянии вариант вокруг средней арифметической.

Средняя арифметическая величина, как правило, вычисляется на основе данных выборочной совокупности. При повторных исследованиях под влиянием случайных явлений средняя арифметическая может изменяться. Это обусловлено тем, что исследуется, как правило, только часть возможных единиц наблюдения, то есть выборочная совокупность. Информация обо всех возможных единицах, представляющих изучаемое явление, может быть получена при изучении всей генеральной совокупности, что не всегда возможно. В то же время с целью обобщения данных эксперимента представляет интерес величина средней в генеральной совокупности. Поэтому для формулировки общего вывода об изучаемом явлении, результаты, полученные на основе выборочной совокупности, должны быть, перенесены на генеральную совокупность статистическими методами.

Чтобы определить степень совпадения выборочного исследования и генеральной совокупности, необходимо оценить величину ошибки, которая неизбежно возникает при выборочном наблюдении. Такая ошибка называется

«ОШИБКОЙ РЕПРЕЗЕНТАТИВНОСТИ» или «Средней ошибкой средней арифметической». Она фактически является разностью между средними, полученными при выборочном статистическом наблюдении, и аналогичными величинами, которые были бы получены при сплошном исследовании того же объекта, т.е. при изучении генеральной совокупности. Поскольку выборочная средняя является случайной величиной, такой прогноз выполняется с приемлемым для исследователя уровнем вероятности. В медицинских исследованиях он составляет не менее 95%.

Ошибку репрезентативности нельзя смешивать с ошибками регистрации или ошибками внимания (описки, просчеты, опечатки и др.), которые должны быть сведены до минимума адекватной методикой и инструментами, применяемыми при проведении эксперимента.

Величина ошибки репрезентативности зависит как от объема выборки, так и от вариабельности признака. Чем больше число наблюдений, тем ближе выборка к генеральной совокупности и тем меньше ошибка. Чем более изменчив признак, тем больше величина статистической ошибки.

На практике для определения ошибки репрезентативности в вариационных рядах пользуются следующей формулой:

$$m = \frac{\sigma}{\sqrt{n-1}},$$

где: m – ошибка репрезентативности;

σ – среднее квадратическое отклонение;

n – число наблюдений в выборке.

Из формулы видно, что размер средней ошибки прямо пропорционален среднему квадратическому отклонению, т. е. вариабельности изучаемого признака, и обратно пропорционален корню квадратному из числа наблюдений.

При выполнении статистического анализа на основе вычисления относительных величин построение вариационного ряда не является обязательным. При этом определение средней ошибки для относительных показателей может выполняться по упрощенной формуле:

$$m = \sqrt{\frac{Pq}{n-1}},$$

где: P – величина относительного показателя, выраженного в процентах, промилле и т.д.;

q – величина, обратная P и выраженная как $(1-P)$, $(100-P)$, $(1000-P)$ и т. д., в зависимости от основания, на которое рассчитан показатель;

n – число наблюдений в выборочной совокупности.

Однако, указанная формула вычисления ошибки репрезентативности для относительных величин может применяться только в том случае, когда значение показателя меньше его основания. В ряде случаев расчета интенсивных показателей такое условие не соблюдается, и показатель может выражаться числом более 100% или 1000‰. В такой ситуации выполняется построение вариационного ряда и

вычисление ошибки репрезентативности по формуле для средних величин на основе среднего квадратического отклонения.

Прогнозирование величины средней арифметической в генеральной совокупности

Ошибка репрезентативности позволяет определить возможные значения средней величины в генеральной совокупности с требуемой вероятностью. Такой прогноз выполняется с указанием двух значений – минимального и максимального, в пределах которых может колебаться искомая средняя величина генеральной совокупности. Эти значения называются «Доверительные границы».

Прогнозирование величины средней арифметической в генеральной совокупности выполняется с указанием двух значений – минимального и максимального. Эти крайние значения возможных отклонений, в пределах которых может колебаться искомая средняя величина генеральной совокупности, называются «Доверительные границы».

Постулатами теории вероятностей доказано, что при нормальном распределении признака с вероятностью 99,7%, крайние значения отклонений средней будут не больше величины утроенной ошибки репрезентативности ($M \pm 3m$); в 95,5% – не больше величины удвоенной средней ошибки средней величины ($M \pm 2m$); в 68,3% – не больше величины одной средней ошибки ($M \pm 1m$). Отметим, что это утверждение справедливо только для признака, подчиняющегося нормальному закону распределения Гаусса. Такой закон выражается формулой:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(V-M)^2}{2\sigma^2}}, -\infty < x < \infty$$

- где: $f(x)$ – функция распределения;
- σ - среднее квадратическое отклонение;
- e – основание натурального логарифма;
- V - случайная величина (варианта);
- M – средняя величина.

Графически эта функция имеет форму колокола (рис. 11).

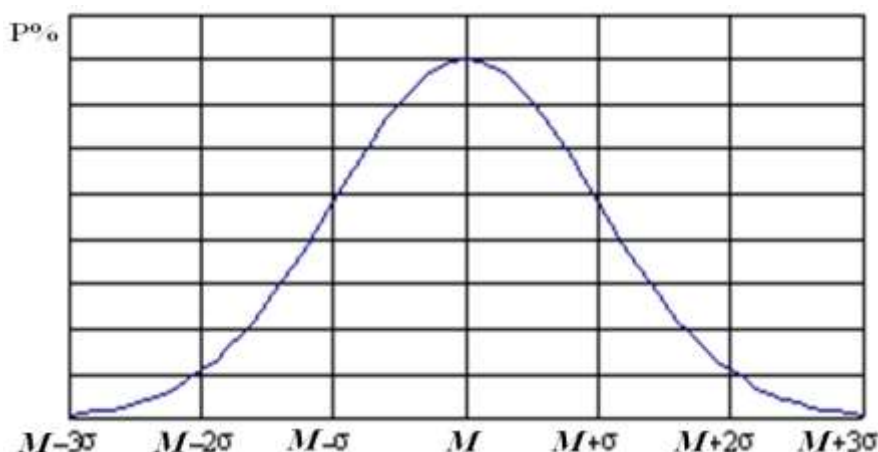


Рис. 11. Функция плотности вероятности случайной величины для нормального закона распределения Гаусса.

Большинство экспериментальных исследований, в том числе и в области медицины, связано с измерениями, результаты которых могут принимать практически любые значения в заданном интервале, поэтому, как правило, описываются моделью непрерывных случайных величин. В связи с этим в медицине наиболее часто применяются статистические методы, основанные на принципах непрерывных распределений. Одним из таких распределений, имеющим основополагающую роль в медицинской статистике, является нормальное распределение Гаусса.

Использование допущения о принадлежности медико-биологических данных закону распределения Гаусса объясняется рядом причин.

1. Прежде всего, модель нормального распределения подходит для описания большинства экспериментальных данных. Следует отметить, что не существует распределений эмпирических данных, которые в точности совпадали бы с нормальным законом распределения, поскольку нормально распределенная случайная величина находится в пределах от $-\infty$ до $+\infty$, чего никогда не встречается на практике.

Проводятся ли измерения веса, роста и других физиологических параметров организма человека - везде на результаты оказывает влияние очень большое число случайных факторов (естественные причины и ошибки измерения). Причем, как правило, действие каждого из этих факторов незначительно. Установлено, что результаты именно в таких случаях будут распределены практически по закону распределения Гаусса.

2. Многие распределения, связанные со случайной выборкой, при увеличении объема выборочного исследования переходят в нормальное.

3. Нормальное распределение хорошо подходит в качестве приближенного описания других непрерывных распределений (например, асимметричных).

4. Нормальное распределение обладает рядом благоприятных математических свойств, обеспечивших его широкое применение в статистике.

В то же время следует отметить, что в медицинских данных встречается много экспериментальных распределений, описание которых моделью нормального распределения невозможно. Для этого в статистке разработаны методы, которые принято называть «Непараметрическими». Поэтому выбор статистического метода, который предполагается применять для обработки данных конкретного эксперимента, должен производиться в зависимости от принадлежности полученных данных к нормальному закону распределения.

Анализ характера распределения опытных данных (его еще называют проверкой на нормальность распределения) осуществляется по каждому параметру, регистрируемому в эксперименте. Чтобы уверенно судить о соответствии распределения параметра нормальному закону, необходимо анализировать достаточно большое число единиц наблюдения (не менее 30 значений).

Проверка гипотезы на соответствие закона распределения признака нормальному закону распределения выполняется с помощью гистограммы распределения частот (графика), а также ряда статистических критериев. Среди них:

- Критерий асимметрии (β);

- Критерий проверки на эксцесс (γ);
- Критерий Шапиро – Уилкса (W).

Для нормального распределения критерии асимметрии и эксцесса принимают значение 0. Если распределение смещено вправо $\beta > 0$ (положительная асимметрия), при $\beta < 0$ - график распределения смещен влево (отрицательная асимметрия). Критерий асимметрии проверяет форму кривой распределения. В случае нормального закона $\gamma = 0$. При $\gamma > 0$ кривая распределения острее, если $\gamma < 0$ пик более сглаженный, чем функция нормального распределения.

Для проверки на нормальность по критерию Шапиро – Уилкса требуется найти значение этого критерия по статистическим таблицам при необходимом уровне значимости и в зависимости от числа единиц наблюдения (степеней свободы). Приложение 1. Гипотеза о нормальности отвергается при малых значениях этого критерия, как правило, при $W < 0,8$.

Пример построения вариационных рядов, вычисления средних величин, создания графика распределения признака и проверки на нормальность распределения.

Задача: Для выявления общей характеристики частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделении №1 больницы выполнено 16 измерений пульса у детей: Иванов Василий – 120 ударов в минуту, Сидоров Костя – 130 уд/мин, Казаков Василий – 115 уд/мин, Быков Иван - 120 уд/мин, Ежков Дмитрий - 120 уд/мин, Попов Александр - 125 уд/мин, Комин Алексей - 110 уд/мин, Ким Эльдар - 125 уд/мин, Грибин Петр - 115 уд/мин, Котов Дмитрий - 120 уд/мин, Носов Семен - 125 уд/мин, Лобанов Семен - 135 уд/мин, Бычков Евгений - 115 уд/мин, Мокин Юрий - 130 уд/мин, Захаров Михаил - 125 уд/мин, Чернов Павел - 120 уд/мин.

Задание: а) составьте простой не ранжированный вариационный ряд, вычислите среднюю арифметическую величину и определите степень рассеяния вариантов в вариационном ряду;

б) выполните упорядочение (ранжирование) ряда по возрастанию и определите медиану;

в) составьте сгруппированный ранжированный вариационный ряд, постройте график распределения признака, определите моду и среднюю взвешенную величину;

г) определите вид распределения признака и статистические критерии нормальности распределения;

д) вычислите ошибку репрезентативности и доверительные границы колебания средней арифметической в генеральной совокупности.

Решение: необходимо запустить программу Excel, создать новый лист, и на этом листе ввести условие и выполнить вычисления, как показано на рис.12, сохранить файл.

	A	B	C	D	E	F	G	
1	Расчет средних величин							
2								
3	Частота пульса детей, поступивших в отделение №1 больницы.							
5		V	d	d ²		Расчет мастером "Анализ данных - Описательная статистика		
6	1 Иванов	120	1,88	3,52		Столбец1		
7	2 Сидоров	130	-8,13	66,02				
8	3 ...	115	6,88	47,27		Среднее	121,875	
9	4 ...	120	1,88	3,52		Стандартная ошибка	1,63777	
10	5 ...	120	1,88	3,52		Медиана	120	
11	6 ...	125	-3,13	9,77		Мода	120	
12	7 ...	110	11,88	141,02		Стандартное отклонение	6,551081	
13	8 ...	125	-3,13	9,77		Дисперсия выборки	42,91667	
14	9 ...	115	6,88	47,27		Эксцесс	-0,16979	
15	10 ...	120	1,88	3,52		Асимметричность	0,209598	
16	11 ...	125	-3,13	9,77		Интервал	25	
17	12 ...	135	-13,13	172,27		Минимум	110	
18	13 ...	115	6,88	47,27		Максимум	135	
19	14 ...	130	-8,13	66,02		Сумма	1950	
20	15 ...	125	-3,13	9,77		Счет	16	
21	16 ...	120	1,88	3,52		Наибольший(1)	135	
22	Сумма=	1950	0,00	643,75		Наименьший(1)	110	
23	n=	16		643,75		Уровень надежности(95,0%)	3,490827	
24	M ₁ =	121,9						
25	m ₁ =	1,64	$m = \frac{\delta}{\sqrt{n-1}}$		Сигма:	$\sigma = \sqrt{\frac{\sum d^2}{n}}$	6,343057	
26	C ₁ =	5,2				$\sigma = \sqrt{\frac{\sum d^2}{n-1}}$		
27						Стандартное отклонение	6,551081	
28	Уровень значимости		0,05			Довер-й интервал=	3,209971	
29	Довер-е границы		от: 118,6			до: 125,2		
30	Асимметрия		0,209598					
31	Эксцесс=		-0,16979					
32	Вывод:	Средняя частота пульса пациентов 2-го отделения с вероятностью 95,5% составляет от 118,6 до 125,2 ударов в минуту, вариабельность малая.						

Рис. 12. Пример построения простого вариационного ряда и вычисления средних величин в программе Excel.

а) Построение простого вариационного ряда частоты пульса детей, поступивших в отделение №1 больницы. Простой вариационный ряд представляет собой таблицу, в которой подлежащим является изучаемый признак, обозначаемый знаком V (варианта). Все полученные в эксперименте данные вносят в таблицу в порядке их записи в журнал регистрации (условия задачи). Если значения вариант повторяются, то каждое из них заносится в отдельную ячейку таблицы. Затем вычисляют сумму вариант, среднюю арифметическую, отклонения (d) каждой варианты от средней величины и квадрат отклонения (d²) соответствующими формулами (таблица 13).

Таблица 13

Простой не ранжированный вариационный ряд

	V	d=V-M	d ²
1 Иванов Василий	120	-1,88	3,52
2 Сидоров Костя	130	8,13	66,02
3 ...	115	-6,88	47,27
4 ...	120	-1,88	3,52
5 ...	120	-1,88	3,52
6 ...	125	3,13	9,77
7 ...	110	-11,88	141,02
8 ...	125	3,13	9,77
9 ...	115	-6,88	47,27
10 ...	120	-1,88	3,52
11 ...	125	3,13	9,77
12 ...	135	13,13	172,27
13 ...	115	-6,88	47,27
14 ...	130	8,13	66,02
15 ...	125	3,13	9,77
16 ...	120	-1,88	3,52
Сумма (Σ) =	1950	0,00	643,75
	n=16		

Вычисление средней арифметической (M) в простом вариационном ряду выполняется по формуле:

$$M = \frac{\sum V}{n} = 1950/16 = \mathbf{121,9} \text{ уд/мин.}$$

В программе Excel для вычисления средней арифметической может применяться функция =СРЗНАЧ(Диапазон ячеек). Использование этой функции даст такой же результат.

Определение степени рассеяния вариант в вариационном ряду с помощью среднего квадратического отклонения (σ - Сигма) и коэффициента вариации (С). Среднее квадратическое отклонения вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}} = \text{КОРЕНЬ}(643,75 / 16) = \mathbf{\pm 6,343}.$$

Для упрощения расчета среднего квадратического отклонения при $n > 30$ в программе Excel может использоваться функция =СТАНДОТКЛОН(Диапазон). Если выполнить вычисления этой функцией в нашем вариационном ряду, то будет получено значение стандартного отклонения = 6,55. Это значение отличается от величины среднего квадратического отклонения, что обусловлено малым числом наблюдений ($n = 16$).

Коэффициент вариации (C) вычисляется формулой:

$$C = \frac{\sigma}{M} \times 100\% = 6,343/121,9 * 100 = \underline{5,2\%}, \text{ что является малым рассеянием } (<10\%).$$

Вариабельность признака (рассеяние) оценивается как малая при $C < 10\%$, средняя при $10\% < C < 20\%$, высокая при $C > 20\%$

Вывод: средняя частота пульса пациентов изучаемой группы составляет 121,9 ударов в минуту, вариабельность пульса малая.

б) Преобразование не ранжированного вариационного ряда в ранжированный (таблица 14) производится в имеющей таблице простого ряда с помощью команды сортировки по столбцу V . С этой целью необходимо предварительно выделить диапазон ячеек таблицы, включающий данные из сортируемого столбца, а затем на вкладке «Главная» нажать кнопку «Сортировка и фильтр». При выполнении сортировки требуется автоматически расширить диапазон сортируемых ячеек.

Таблица 14

Ранжированный вариационный ряд

	V	$d=V-M$	d^2
1	110	-11,88	141,02
2	115	-6,88	47,27
3	115	-6,88	47,27
4	115	-6,88	47,27
5 Иванов Василий	120	-1,88	3,52
6	120	-1,88	3,52
7	120	-1,88	3,52
8	120	-1,88	3,52
9	120	-1,88	3,52
10	125	3,13	9,77
11	125	3,13	9,77
12	125	3,13	9,77
13	125	3,13	9,77
14 Сидоров Костя	130	8,13	66,02
15	130	8,13	66,02
16	135	13,13	172,27
Сумма(Σ) = =СУММ(Диапазон ячеек)	1950	0	643,75
Средняя арифметическая = =СРЗНАЧ(Диапазон ячеек)	121,875		

Для определения медианы (Me) в ранжированном вариационном ряду находят значение варианты, расположенной в центре ряда.

Me = варианта, занимающая срединное положение (если ряд состоит из четного числа вариантов, медианой является полусумма двух центральных вариантов) = **120 уд/мин.**

В программе Excel для определения медианы применяется функция =МЕДИАНА(Диапазон ячеек).

в) Построение сгруппированного ранжированного ряда выполняется в отдельной таблице (таблица 15), которая должна включать столбец частоты повторений вариант (p). Данные из сгруппированного вариационного ряда используются для построения точечной диаграммы распределения частот. Он демонстрирует частоту встречаемости вариант (p), указываемую по вертикальной оси точечной диаграммы (рис. 13). Число повторений вариант может быт вычислено из диапазона ячеек простого вариационного ряда функцией =СЧЁТЕСЛИ(Диапазон; критерий).

Таблица 15

Сгруппированный ранжированный
вариационный ряд

	V	p
1	110	1
2	115	3
3	120	5
4	125	4
5	130	2
6	135	1
Σ =	735	16

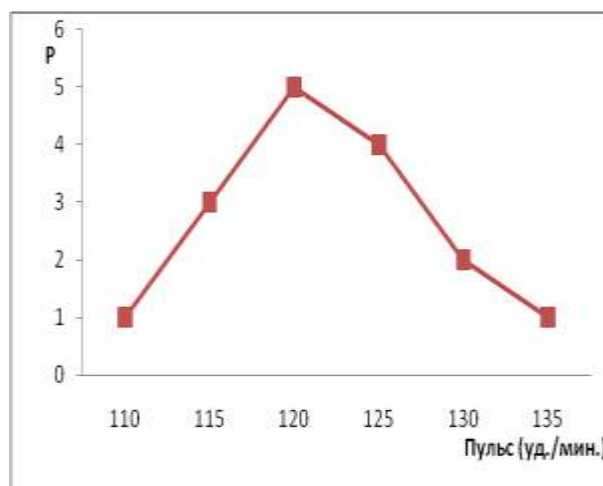


Рис. 13. График распределения признака.

Определение моды может выполняться в сгруппированном вариационном ряду выбором значения варианты, которая повторяется наибольшее число раз.

Mo = наиболее часто повторяющаяся варианта = **120 уд/мин** (встречается 5 раз).

В программе Excel для определения моды может применяться функция =МОДА(Диапазон данных).

ВНИМАНИЕ! Эта функция используется только в простом вариационном ряду.

В сгруппированном вариационном ряду средняя арифметическая вычисляется по модифицированной формуле и называется «Средняя взвешенная»:

$$M = \frac{\sum Vp}{n} = \frac{110*1+115*3+120*5+125*4+130*2+135*1}{1+3+5+4+2+1} = \frac{735}{16} = 121,9$$

Независимо от способа построения вариационного ряда и метода вычисления средней арифметической, для одинаковых данных значения их средней величины должны совпадать.

г) Оценка нормальности распределения выполняется с помощью графика распределения (рис. 13) по форме линии, которая должна приближаться к виду кривой Гаусса (функции нормального распределения). Статистические критерии нормальности распределения вычисляются с помощью функций или в модуле «Описательная статистика», который может быть вызван командой «Данные» - «Анализ данных» - «Описательная статистика».

Если указанная команда отсутствует разделе «Данные», необходимо выполнить установку надстройки программы Excel «Пакет анализа». Для этого требуется открыть меню настройки программы кнопкой «Office», выбрать «Параметры Excel» -> «Надстройки» -> «Перейти» -> «Пакет анализа» и «Пакет анализа - VBA» -> «ОК». Модуль анализа данных в программе Excel, как правило, не подключается при типовой установке пакета программ Microsoft Office. Поэтому для применения методов статистической обработки данных необходимо однократно выполнить указанную настройку программы.

Модуль «Описательная статистика» производит вычисления только в простых вариационных рядах. В связи с этим необходимо использовать диапазон ячеек таблицы, в котором составлен ПРОСТОЙ вариационный ряд созданный в пункте б) настоящей задачи. Результат вычислений, выполненный модулем «Описательная статистика», показан в таблице 16.

Таблица 16

Простой не ранжированный вариационный ряд и результат работы модуля «Описательная статистика»

	V	d=V-M	d ²
1 Иванов Василий	120	-1,88	3,52
2 Сидоров Костя	130	8,13	66,02
3 ...	115	-6,88	47,27
4 ...	120	-1,88	3,52
5 ...	120	-1,88	3,52
6 ...	125	3,13	9,77
7 ...	110	-11,88	141,02
8 ...	125	3,13	9,77
9 ...	115	-6,88	47,27
10 ...	120	-1,88	3,52
11 ...	125	3,13	9,77
12 ...	135	13,13	172,27
13 ...	115	-6,88	47,27
14 ...	130	8,13	66,02
15 ...	125	3,13	9,77
16 ...	120	-1,88	3,52
Σ=	1950	0,00	643,75

n=16

Результат выполнения команды «Анализ данных» -> «Описательная статистика».	
Столбец1	
Среднее	121,875
Стандартная ошибка	1,63777
Медиана	120
Мода	120
Стандартное отклонение	6,551081
Дисперсия выборки	42,91667
Эксцесс	-0,16979
Асимметричность	0,209598
Интервал	25
Минимум	110
Максимум	135
Сумма	1950
Счет	16
Наибольший(1)	135
Наименьший(1)	110
Уровень надежности(95,0%)	3,490827

Эксцесс = -0,17. Для нормального распределения эксцесс = 0. **Распределение близко к нормальному.**

Асимметричность = 0,2. Для нормального распределения асимметричность = 0. **Распределение близко к симметричному.**

Применяя стандартные функции =ЭКСС(Диапазон) и =СКОС(Диапазон) мы получим эти же значения.

д) Определение ошибки средней арифметической (m , ошибка репрезентативности или стандартная ошибка) выполняется по формуле:

$$m = \frac{\sigma}{\sqrt{n-1}} = 6,343 / \text{КОРЕНЬ}(16-1) = \underline{1,64}$$

Программа Excel позволяет вычислять этот критерий с помощью модуля «Описательная статистика». При этом ошибка средней арифметической (m) называется «Стандартная ошибка», так как в качестве среднего квадратического отклонения используется «Стандартное отклонение» (таблица 16). Поэтому в вариационных рядах с числом единиц наблюдения менее 30 возможно несущественное отличие значений полученных расчетной формулой и модулем описательной статистики.

Таким образом, с вероятностью $P > 95\%$ доверительные границы колебания средней арифметической в генеральной совокупности $M \pm 2m = 121,9 \pm 2 \cdot 1,64 = \underline{121,9 \pm 3,28 \text{ уд/мин}}$.

Для вычисления доверительных границ в программе Excel может применяться статистическая функция =ДОВЕРИТ(0,05; Значение стандартного отклонения; Число единиц наблюдения). При ее использовании с вычисленными ранее значениями задачи получим:

$$=ДОВЕРИТ(0,05; 6,55; 16) = 3,21.$$

Вывод: Средняя частота пульса пациентов 1-го отделения с вероятностью 95,5% составляет от 118,6 до 125,2 ударов в минуту.

Пример сравнения рассеяния вариационных рядов

Условие задачи: Для выявления общей характеристики частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделении №2 больницы выполнено 17 измерений пульса у детей: 1. Казаков Саша – 130 уд.в мин; 2. Литвинов Сережа – 135 уд.в мин.; 3...– 125; 4... – 115; 5 ...– 125; 6 ... – 125; 7 ... – 120; 8 ... – 125; 9 ... – 130; 10 ... – 120; 11 ... – 140; 12 ... – 145; 13 ... – 115; 14 ...– 130; 15 ... – 125; 16 ... – 120; 17 ... – 125.

Задание: а) создать простой, ранжированный вариационный ряды, определить средние величины и доверительные границы средней арифметической;
б) составить сгруппированный, ранжированный вариационный ряд и построить график распределения признака, выполнить проверку его на нормальность;
в) определить параметры вариабельности признака: амплитуду, размах, среднее квадратическое отклонение, коэффициент вариации и ошибку репрезентативности;
г) сравнить характеристики рассеяния вариационных рядов, используемых в примерах данного раздела.

Решение: необходимо запустить программу Excel, создать новый лист, ввести условие задачи, выполнить вычисления с помощью формул и модуля «Описательная статистика», как показано ниже, сохранить файл.

а) Построение простого ранжированного вариационного ряда (таблица 17) и вычисление их основных характеристик модулем «Описательная статистика».

Таблица 17

Простой ранжированный вариационный ряд и его характеристики

	V	d=V-M	d ²
1	115	11,18	124,91
2	115	11,18	124,91
3	115	11,18	124,91
4	120	6,18	38,15
5	120	6,18	38,15
6	125	1,18	1,38
7	125	1,18	1,38
8	125	1,18	1,38
9	125	1,18	1,38
10	125	1,18	1,38
11	125	1,18	1,38
12 Казаков	130	-3,82	14,62
13	130	-3,82	14,62
14	130	-3,82	14,62
15 Литвинов	135	-8,82	77,85
16	140	-13,82	191,09
17	145	-18,82	354,33
$\Sigma=$	2145	0,00	1126,47

n=17

M₂= **126,2**

m₂= **2,04**

Сигма (σ)= 8,390734

Результат выполнения команды «Анализ данных» -> «Описательная статистика».	
Столбец1	
Среднее	126,1764706
Стандартная ошибка	2,035051839
Медиана	125
Мода	125
Стандартное отклонение	8,390733685
Дисперсия выборки	70,40441176
Эксцесс	0,3573165
Асимметричность	0,659063476
Интервал	30
Минимум	115
Максимум	145
Сумма	2145
Счет	17
Наибольший(1)	145
Наименьший(1)	115
Уровень надежности(95,0%)	4,314116203

Вычисление коэффициента вариации: **C=8,39/126,2*100=6,6%**

Вывод: Средняя частота пульса пациентов 2-го отделения с вероятностью 95,5% составляет 126,2±2,04 ударов в минуту, вариабельность малая.

б) Проверка на нормальность распределения признака производится на основе сгруппированного вариационного ряда (таблица 18), который создается в новой таблице, как продемонстрировано в предыдущем примере. В этой таблице указывают величину вариант и число их повторений (частоту, p), сортируют по возрастанию (ранжируют), а затем определяют моду и медиану. Затем изготавливается точечная диаграмма распределения признака (рис. 14).

Таблица 18
Сгруппированный, ранжированный
вариационный ряд

	<i>V</i>	<i>p</i>
1	115	3
2	120	2
3	125	6
4	130	3
5	135	1
6	140	1
7	145	1
	$\Sigma p =$	17

n=17

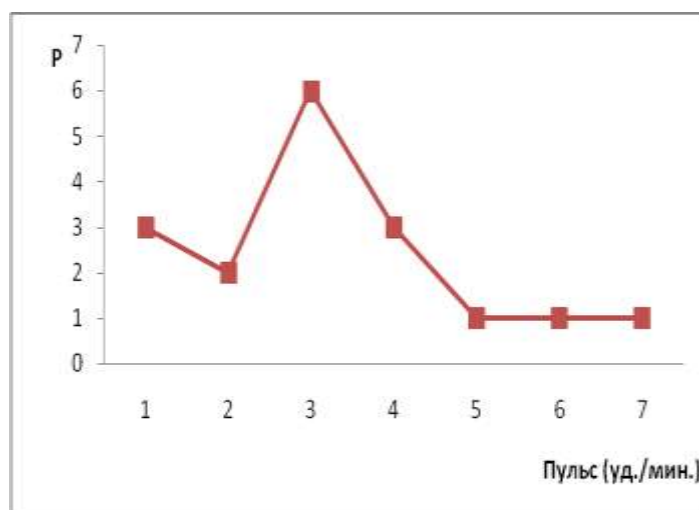


Рис. 14. График распределения признака.

Me= варианта, занимающая срединное положение =МЕДИАНА(Диапазон) = 125 уд/мин.

Mo= наиболее часто повторяющаяся варианта =МОДА(Диапазон) = 125 уд/мин.

в) Параметры вариабельности определяются по результатам обработки данных модулем «Описательная статистика» (таблица 17). Амплитуда (интервал) = 30 уд/мин, размах от 115 до 145 уд/мин, среднее квадратическое отклонение = 8,4 уд/мин, коэффициент вариации = **6,6%**, ошибка репрезентативности = ±2,04 уд/мин.

г) Сравнение рассеяния вариационных рядов, используемых в примерах данного раздела выполняется на основе коэффициента вариации первого (***C*₁=5,4%**) и второго вариационных рядов (***C*₂=6,6%**).

Вывод: вариабельность пульса пациентов 1-го отделения ниже, чем пациентов 2-го отделения больницы и в обоих случаях малая (<10%).

ЗАДАНИЯ

В программе Excel изготовьте вариационные ряды и выполните вычисление средних величин, оцените вариабельность признака.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя

«СредВеличины» и выполните задачи варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. **Сообщите преподавателю о выполнении задания.** После проверки решения задач преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

- а) В районе N, где расположена тепловая электростанция, в одной из точек жилого поселка было взято 100 проб атмосферного воздуха. Количество пыли в пробах воздуха измерялось следующими цифрами: 0,09 мг/м³ в 2-х пробах, 0,08 мг/м³ – 2 раза, 0,15 мг/м³ – 16 раз, 0,12 мг/м³ – 14 раз, 0,14 мг/м³ – 30 раз, 0,16 мг/м³ – 4 раза, 0,13 мг/м³ – 16 раз, 0,11 мг/м³ – 9 раз, 0,10 мг/м³ – 5 раз, 0,17 мг/м³ – 2 раза. Составьте простой, ранжированный вариационный ряд и определите среднесуточную концентрацию пыли, ее вариабельность, доверительные границы колебаний средней величины. Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.
- б) Сравните характер разнообразия массы тела у новорожденных, детей первого года жизни и семилетних, если известны следующие параметры:

Возраст	Средняя масса (M), кг	σ, кг
Новорожденные	3,4	±0,5
1 год	10,5	± 0,8
7 лет	22,9	±2,7

Вариант 2

- а) В питьевой воде, которой снабжаются дома жителей города N, определялась концентрация соединений фтора. В 4-х пробах было обнаружено 0,5 мг/л этих соединений, в 5-и – 0,6 мг/л, в 2-х – 0,3 мг/л, в 3-х – 0,4 мг/л, в 8-и – 0,8 мг/л, в 10-и – 0,9 мг/л, в 9-и – 1,2 мг/л, в 12-и – 1,1 мг/л, в 8-и – 1,3 мг/л, в 17-и – 1,0 мг/л, в 8-и – 1,5 мг/л, в 7-и – 1,6 мг/л, в 7-и – 0,7 мг/л, в 9-и – 1,4 мг/л, в 1-й – 0,2 мг/л. Составьте простой, ранжированный вариационный ряд, Определите среднюю концентрацию соединений фтора в питьевой воде и доверительные границы колебаний средней величины. Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.
- б) Сравните характер разнообразия лабораторных анализов с различной размерностью:

Наименование теста	Средний показатель	σ
Общий белок крови, мг%	6,8	±0,4
СОЭ, мм/ч	9	± 2
Лейкоциты	8000	±800

Вариант 3

- а) При стоматологическом обследовании группы школьников 4-х классов сельского района были получены следующие результаты: 2 школьника имели по 5 кариозных зубов, 28 школьников – по 1 кариозному зубу, 8 школьников – по 4 кариозных зуба, 1 школьник – 8 кариозных зубов, 20 школьников – по 3 кариозных зуба, 16 школьников – по 2 кариозных зуба и 6 школьников не имели пораженных

кариесом зубов. Составьте простой, ранжированный вариационный ряд, определите среднее число зубов пораженных кариесом у школьников района, степень вариабельности этого признака и доверительные границы колебаний средней величины. Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

- б) Сравните характер разнообразия антропометрических данных у мальчиков 7-и летнего возраста:

Показатель	М	σ
Рост, см	123,4	±4,9
Масса тела, кг	24,2	± 3,1
Окружность грудной клетки, см	60,1	±2,5

Вариант 4

- а) Перед сдачей экзамена у студентов определялась частота пульса. Были получены следующие данные: у 2-х студентов – 76 ударов в минуту, у 3-х студентов – 80 ударов в минуту, у 4-х студентов – 108 ударов в минуту, у 2-х студентов – 116 ударов в минуту, у 20-и студентов – 88 ударов в минуту, у 6-и студентов – 98 ударов в минуту, у 17-и студентов – 86 ударов в минуту, у 11-и студентов – 92 удара в минуту. Составьте простой, ранжированный вариационный ряд и определите среднюю частоту пульса у студентов перед экзаменом, степень вариабельности признака и доверительные границы колебаний средней величины. Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

- б) Сравните характер разнообразия антропометрических данных у девушек 17-и летнего возраста:

Показатель	М	σ
Рост, см	161,2	±5,1
Масса тела, кг	55,8	±7,2
Жизненная емкость легких, см ³	3400	±250

Вариант 5

- а) Исследовалась длительность лечения больных пневмонией в стационаре центральной районной больницы района N. Были получены следующие результаты: 10 дней лечилось 2 больных, 9 дней – 1 больной, 11 дней – 1 больной, 12 дней – 1 больной, 8 дней – 3 больных, 13 дней – 1 больной, 21 день – 3 больных, 7 дней – 1 больной, 22 дня – 3 больных, 14 дней – 2 больных, 20 дней – 5 больных, 15 дней – 2 больных, 16 дней – 3 больных, 17 дней – 4 больных, 19 дней – 8 больных, 18 дней – 7 больных. Составьте простой, ранжированный вариационный ряд и определите среднюю длительность лечения пневмонии, степень вариабельности признака и доверительные границы колебаний средней величины. Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

б) Сравните характер разнообразия антропометрических данных у 12-летних мальчиков:

Показатель	М, см	σ
Рост	142,0	$\pm 8,5$
Окружность грудной клетки	66,0	$\pm 4,0$
Окружность головы	50,0	$\pm 2,0$

Вариант 6

а) Исследовалась длина тела новорожденных девочек по данным родильного дома. Были получены следующие данные: у 8 девочек рост составил 48 см, у 6 девочек – 51 см, у 7 девочек – 53 см, у 1 девочки – 49 см, у 9 девочек – 52 см, у 8 девочек – 50 см, у 1 девочки – 47 см, у 3 девочек – 46 см, у 2 девочек – 54 см, у 1 девочки – 55 см, у 1 девочки – 56 см. Составьте простой, ранжированный вариационный ряд, определите среднюю длину тела новорожденных девочек, степень вариабельности признака и доверительные границы колебаний средней величины. Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

б) Сравните характер разнообразия антропометрических данных у 12-летних девочек:

Показатель	М	σ
Рост, см	140	$\pm 9,5$
Масса тела, кг	40	± 6
Жизненная емкость легких, см ³	2300	± 360

Вопросы для самоконтроля

1. Дайте определение вариационного ряда
2. С какой целью в медицинских исследованиях используются средние величины?
3. Дайте определение средней величины.
4. Назовите основные элементы вариационного ряда.
5. Как вычисляется простая средняя арифметическая?
6. Как вычисляется взвешенная средняя арифметическая?
7. Назовите основные свойства средней арифметической величины.
8. Что такое среднее квадратическое отклонение и что показывает эта характеристика?
9. Как вычисляется среднее квадратическое отклонение?
10. В каких случаях возникает необходимость вычисления коэффициента вариации? Приведите пример.
11. Как вычисляется коэффициент вариации.
12. Понятие ошибки репрезентативности.
13. Как вычисляется ошибка репрезентативности?
14. Какова формула вычисления ошибки репрезентативности для относительных величин?
15. Как определяются доверительные границы средней арифметической в

генеральной совокупности и с какой целью?

Рекомендуемая литература:

1. Гельман В.Я. Медицинская информатика. Практикум. СПб: Питер, 2001. – 420 с.
2. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.
3. Симонович С.В. Информатика. Базовый курс: Учебник для вузов. 3-е изд. Стандарт третьего поколения. – СПб.: Питер, 2011. – 640 с.: ил.
4. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>

VII. Проверка статистических гипотез, критерий Стьюдента

В научно-исследовательской практике часто требуется сопоставить средние арифметические, например, при сравнении результатов в контрольной и экспериментальной группах, при оценке показателей здоровья населения в различных местностях за несколько лет и т. д.

Методологической основой любого исследования является формулировка рабочей гипотезы. При этом основной целью исследования является получение данных, на основании которых выдвинутую еще до начала исследования (априори) гипотезу можно было бы принять, т.е. признать истинной, либо отвергнуть - признать ложной.

Выдвинутую гипотезу называют основной или нулевой (H_0). Гипотезу, которая противоречит нулевой и является ее логическим отрицанием, называют конкурирующей или альтернативной (H_1).

Гипотезы H_0 и H_1 предоставляют выбор только одного из двух вариантов. Например, если нулевая гипотеза предполагает, что среднее арифметическое $M = 15$, то логическим отрицанием будет $M \neq 15$. Коротко это записывается так: $H_0: M=15$; $H_1: M \neq 15$. В медико-биологических исследованиях при сравнении регистрируемых признаков в качестве нулевой гипотезы принимают гипотезу об отсутствии различий.

Например, при оценке токсичности какого-либо вещества обычно берутся две группы лабораторных животных. Подбираются животные одинакового возраста, пола, одинакового содержания и т. п. Таким образом, делается все, чтобы эти группы животных представляли собой единую, как можно более однородную статистическую совокупность, с тем, чтобы максимально снизить исходную вариабельность анализируемых данных. Оптимальным с этой точки зрения считается ситуация, когда отличия сравниваемых групп заключаются только в том, что одна из групп (опытная) подвергается воздействию токсического вещества, а другая (контрольная) - нет. В любом случае, произошли ли после воздействия токсического вещества изменения в опытной группе или нет, различия средних показателей в обеих группах обязательно будут. Вопрос состоит в следующем: являются ли эти различия только следствием выборочного исследования, или разница возникла из-за того, что произошли существенные сдвиги физиологических функций животных опытной группы, которые будут обнаруживаться всегда, т.е. в генеральной совокупности. Значит, проверяется вопрос: принадлежат ли животные опытной и контрольной групп к той же самой генеральной совокупности или опытная группа принадлежит к другой генеральной совокупности (совокупности с измененными физиологическими параметрами)?

Методы оценки достоверности различий средних величин позволяют установить, насколько выявленные различия существенны (носят ли они закономерный характер или являются результатом действия случайных причин). Эту оценку можно выполнить только с определенной степенью вероятности, когда после установленного уровня вероятности допущение о наличии различий могут считаться закономерными или, наоборот, отвергаются.

Выдвинутая гипотеза может оказаться правильной или неправильной. При ее статистической проверке может быть отвергнута правильная гипотеза. Вероятность совершить такую ошибку называют уровнем значимости. Этот параметр принято

обозначать знаком α (альфа) или p (пи). В биологии и медицине уровень значимости, как правило, принимают не выше 0,05. Это означает, что в 5 случаях из 100 (в 5%) мы рискуем отвергнуть правильную гипотезу. Соответственно, вероятность принятия такой гипотезы (P) равняется ($P = 1 - p$) 0,95 (или 95%).

Таким образом, статистическая значимость выборочных характеристик представляет собой меру уверенности в их «истинности». Уровень значимости находится в убывающей зависимости от надежности результата. Более высокая статистическая значимость соответствует более низкому уровню доверия к найденной в выборке средней величине. Именно уровень значимости представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю генеральную совокупность.

Выбор порога уровня значимости, выше которого результаты отвергаются как статистически не подтвержденные, во многом произвольный. Как правило, окончательное решение обычно зависит от традиций и накопленного практического опыта в данной области исследований. Верхняя граница $p < 0,05$ статистической значимости содержит довольно большую вероятность ошибки (5%). Поэтому в тех случаях, когда требуется особая уверенность в достоверности полученных результатов, принимается значимость $p < 0,01$ или даже $p < 0,001$.

В практике медико-биологических исследований наиболее часто используются следующие значения показателей значимости: 0,1; 0,05; 0,01; 0,001. Традиционная интерпретация уровней значимости, принятая в этих исследованиях, представлена в таблице 19.

Таблица 19

Интерпретация уровня значимости (p)

Величина уровня значимости (p)	Интерпретация
$\geq 0,1$	Данные согласуются с нулевой гипотезой (H_0), различия не подтверждены
$\geq 0,05$	Есть сомнения в истинности как нулевой (H_0), так и альтернативной гипотез (H_1)
$< 0,05$	Нулевая гипотеза (H_0) может быть отвергнута.
$\leq 0,01$	Нулевая гипотеза (H_0) может быть отвергнута. Сильный аргумент
$\leq 0,001$	Нулевая гипотеза (H_0) наверняка не подтверждается. Очень сильный аргумент

Приблизительно о наличии достоверных различий между средними величинами можно судить по их доверительным границам. Если обнаруживается пересечение верхней границы одного из интервалов и нижней границы второго, можно предположить, что полученная разница средних является случайной и может не повториться в следующих экспериментах с вероятностью, которая использовалась при вычислении этих границ (как правило, 95%).

Если изучаемый признак подчиняется закону нормального распределения Гауса, может использоваться расчет критерия достоверности Стьюдента (t) (коэффициента достоверности). Величина этого коэффициента определяется модулем отношения разности сравниваемых средних величин к ошибке их разности. Ошибка разности равна корню квадратному из суммы квадратов средних ошибок сравниваемых величин: $\sqrt{m_1^2 + m_2^2}$.

Таким образом, коэффициент достоверности (t) определяется по формуле:

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

где: M_1 – средняя арифметическая 1-го вариационного ряда,
 M_2 – средняя арифметическая 2-го вариационного ряда,
 m_1 – ошибка репрезентативности 1-го вариационного ряда,
 m_2 – ошибка репрезентативности 2-го вариационного ряда.

Для сравнения относительных величин (показателей) применяется модифицированная формула:

$$t = \frac{|P_1 - P_2|}{\sqrt{m_1^2 + m_2^2}},$$

где: P_1 – относительная величина (показатель) 1-й группы;
 P_2 – относительная величина (показатель) 2-й группы;
 m_1 – ошибка репрезентативности 1-го показателя;
 m_2 – ошибка репрезентативности 2-го показателя.

При этом ошибка репрезентативности относительной величины может быть вычислена по формуле:

$$m = \sqrt{\frac{Pq}{n-1}},$$

где: P – величина относительного показателя;
 q – величина, обратная P и вычисленная как $(1-P)$, $(100-P)$, $(100-P)$ и т. д., в зависимости от основания, на которое рассчитан показатель;
 n – число наблюдений.

В медико-биологических исследованиях, где число наблюдений больше 30, допускается использовать сравнение вычисленного значения t с критическим значением 2. Если t -критерий больше 2, тогда выявленные различия считаются закономерными (не случайными, достоверными), т.е. они статистически подтверждены с вероятностью более 95%. Если значение критерия меньше 2, то разница не доказана и носит случайный характер, статистически не подтверждается (вероятность менее 95%). При меньшем числе наблюдений значение критического уровня для сравнения с расчетным значением t -критерия необходимо искать в книгах с таблицами Стьюдента или вычислять в статистической компьютерной программе.

Пример определения достоверности различий между средними величинами по критерию Стьюдента

Задача: сравнить средние значения частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделениях №1, №2 (см. пример раздела VI).

Задания: а) оценить достоверность различий между ЧСС пациентов 1го и 2-го отделений с помощью доверительных границ;
б) вычислить критерий Стьюдента для сравнения ЧСС детей в этих отделениях, сделать вывод о достоверности различий средних величин.

Решение: необходимо запустить программу Excel, создать новый лист, ввести данные и выполнить вычисления, как показано на рис.15, сохранить файл.

Сравнение средних 2-х вариационных рядов критерием Стьюдента.											
1-й вариационный ряд:					2-й вариационный ряд:						
Частота пульса детей, поступивших в отделение №1 больницы в 20... году					Частота пульса детей, поступивших в отделение №2 больницы в 20... году						
V	d	d2	Расчет мастером "Анализ данных" - Описательная статистика		V	d	d2	Расчет мастером "Анализ данных" - Описательная статистика			
1 Иванов Василий	120	1,88	3,62	Среднее	121,975	1 Казаков	130	-8,82	14,62	Среднее	126,18
2 Сидоров Костя	130	-8,13	66,02	Стандартная ошибка	1,63777	3 Литвин	135	-8,82	77,85	Стандартная ошибка	2,0351
3	115	6,88	47,27	Медиана	120	4	115	11,18	124,91	Медиана	126
4	120	1,88	3,62	Мода	120	5	125	1,18	1,38	Мода	126
5	120	1,88	3,62	Стандартное отклонение	6,65108	6	125	1,18	1,38	Стандартное отклонение	8,3907
6	125	-3,13	9,77	Дисперсия выборки	42,9167	7	120	6,18	38,15	Дисперсия выборки	70,404
7	110	11,88	141,02	Экцесс	-0,1698	8	125	1,18	1,38	Экцесс	0,3973
8	125	-3,13	9,77	Асимметричность	0,2098	9	115	11,18	124,91	Асимметричность	0,6591
9	115	6,88	47,27	Интервал	25	10	130	-8,82	14,62	Интервал	30
10	120	1,88	3,62	Минимум	110	11	140	-13,82	191,09	Минимум	118
11	125	-3,13	9,77	Максимум	135	12	145	-18,82	354,33	Максимум	148
12	135	-13,13	172,27	Сумма	1950	13	115	11,18	124,91	Сумма	2148
13	115	6,88	47,27	Свет	18	14	130	-8,82	14,62	Свет	17
14	130	-8,13	66,02	Наибольший(t)	135	15	125	1,18	1,38	Наибольший(t)	148
15	125	-3,13	9,77	Наименьший(t)	110	16	120	6,18	38,15	Наименьший(t)	118
16	120	1,88	3,62	Уровень надежности(95,0%)	3,49083	17	125	1,18	1,38	Уровень надежности(95,0%)	4,3141
17	1860	0,00	643,75								
M1=	121,9	Сигма=	6,65	M2=	126,2	Сигма=	8,39073				
m1=	1,64			m2=	2,04						
C=	5,4 %	$\sigma = \sqrt{\frac{d^2}{n-1}}$		C=	6,6 %						
Вывод:	Средняя частота пульса составляет 121,9±1,62 ударов в минуту, вариабельность низкая.					Вывод: Средняя частота пульса пациентов 2-го отделения составляет 126,2±1,62 ударов в минуту, вариабельность низкая.					
Довер-е границы	от: 118,6			до: 125,2	Довер-е границы	от: 122,1			до: 130,2		
Сравнить среднюю частоту пульса детей, поступивших в 1-е и 2-е отделение больницы.											
Оценка достоверности различий по критерию Стьюдента											
t= 1,64667											
Округление до 2-х знаков после запятой «ОКРУГЛ(С41;2)											
t= 1,65 < 2,04											
Вывод: Различия средней частоты пульса пациентов 1-го и 2-го отделений недостоверны.											
Вычисление вероятности различий средних по t-критерию. Функция											
TTEST(массив1;массив2;хвосты;тип)											
Массив1 и Массив2 - вариационные ряды											
Хвосты = 2 Двустороннее распределение											
Тип = 3 Двухсторонний тест с неравными дисперсиями											
Значимость p = 0,318298 > 0,05 Недостоверно											
Вычисление критического значения критерия Стьюдента при n30											
СТЮДРАСПОБР(вероятность;степени_свободы) = СТЮДРАСПОБР(0,05;(16+17-2))											
2,03951											
Округление до 2-х знаков после запятой «ОКРУГЛ(148;2)											
2,04											

Рис. 15. Вычисление доверительных границ и критерия Стьюдента в программе Excel.

а) вычисление доверительных границ колебаний средних в каждом отделении при уровне значимости $p < 0,05$, т.е. с вероятностью прогноза более 95%, составляет $M \pm 2m$, где M – средняя арифметическая, m – ошибка репрезентативности.

По условию задачи в 1-м отделении $M_1=121,9$, $m_1=1,64$. Т.е. $121,9 \pm 2*1,64 = 121,9 \pm 3,28$ уд/мин. В ячейках таблицы вводятся формулы: $=121,9+3,28$ и $=121,9-3,28$. Получаем доверительные границы колебаний средней частоты пульса в 1-м отделении от 118,62 до 125,18 уд/мин.

Аналогично определяем доверительные границы средней ЧСС во 2-м отделении. По условию задачи $M_2=126,2$, $m_2=2,04$. Т.е. $126,29 \pm 2 * 2,04 =$

$126,2 \pm 4,08$ уд/мин. Формулы вычисления $=126,29+4,08$ и $=126,29-4,08$. Получаем доверительные границы колебаний средней частоты пульса в 2-м отделении от 122,21 до 130,37 уд/мин.

Величина доверительных границ частоты пульса в 2-х отделениях больницы позволяют утверждать, что при повторных экспериментах в 95% случаях будут получены средние величины, укладывающиеся в пределах вычисленных значений границ в 1-м отделении от 118,62 до 125,18 уд/мин, во 2-ом - от 122,21 до 130,37 уд/мин. Поскольку доверительные границы этих отделений имеют пересечение верхней границы 1-го и нижней границы 2-го отделений, можно предположить, что полученная разница средних является случайной и может не повториться в следующих экспериментах.

б) оценка достоверности различий средней частоты пульса детей, поступающих в 1-е и 2-е отделение больницы по критерию Стьюдента.

Формула вычисления критерия Стьюдента: $t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$,

где: M_1 – средняя арифметическая 1-го вариационного ряда - 121,8;
 M_2 – средняя арифметическая 2-го вариационного ряда - 126,2;
 m_1 – ошибка репрезентативности 1-го вариационного ряда - 1,64;
 m_2 – ошибка репрезентативности 2-го вариационного ряда - 2,04.

В программе Excel эта формула принимает вид:
 $= (121,8 - 126,2) / \text{КОРЕНЬ}(1,64^2 + 2,04^2) = -1,64667$.

Модуль числа может быть получен с помощью функции $=\text{ABS}(\text{Число}) = \text{ABS}(-1,64667) = 1,64667$. Округление числа выполняется функцией $=\text{ОКРУГЛ}(\text{Число}; \text{Разрядность}) = \text{ОКРУГЛ}(1,64667; 2) = \mathbf{1,65}$.

Вычисленное значение t -критерия (-1,65) оценивается по модулю числа (1,65) в сравнении с критическим значением, которое при числе наблюдений $n > 30$ составляет 2. При числе наблюдений $n < 30$ критическое значение находят по таблицам Стьюдента при степенях свободы $df = n_1 + n_2 - 2 = 16 + 17 - 2 = 31$. В программе Excel критическое значение критерия Стьюдента вычисляется функцией $= \text{СТЮДРАСПОБР}(\text{Уровень значимости } p; \text{Степени свободы } df) = \text{СТЮДРАСПОБР}(0,05; (16+17-2)) = \mathbf{2,04}$.

Если $t > 2,04$ – статистическая гипотеза о равенстве средних с уровнем значимости $p < 0,05$ опровергается, следовательно, истинной будет являться гипотеза об их различии. Если $t < 2,04$ – гипотеза равенства средних подтверждается.

В нашем примере получаем: $t = \mathbf{1,65 < 2,04}$.

Если в сравниваемых вариационных рядах равное число наблюдений ($n_1 = n_2$), программа Excel позволяет выполнить вычисления при помощи функции ТТЕСТ. $=\text{ТТЕСТ}(\text{Диапазон1}; \text{Диапазон2}; \text{Н}; \text{Тип})$, где:

Диапазон1 – первый вариационный ряд (множество данных);

Диапазон2 – второй вариационный ряд (множество данных);

N – число хвостов распределения (1 или 2), как правило, используется число 2. Если $N = 1$, то функция ТТЕСТ использует одностороннее распределение, при $N = 2$ используется двустороннее распределение.

Тип – цифра модификации теста 1, 2 или 3. Как правило используется цифра 3. Если указана цифра 1 – это парный тест для связанных выборок, 2 – двухвыборочный с равными дисперсиями, 3 – двухвыборочный с неравными дисперсиями.

В большинстве задач статистической обработки медицинских данных эта функция применяется с параметрами =ТТЕСТ(Диапазон1;Диапазон2;2;3), что считается более грубой оценкой, но вполне достаточной для опровержения нулевой гипотезы.

Функция ТТЕСТ возвращает уровень значимости основной гипотезы при сравнении 2-х числовых массивов, вычисленный по критерию Стьюдента. Он выражает вероятность того, что две выборки взяты из генеральных совокупностей, которые имеют одно и то же среднее.

В нашем случае можно выполнить вычисление этой функцией на основе данных 16-и человек в каждой группе. Получаем опытный уровень значимости 0,12. Это означает, что выдвинутая гипотеза о равенстве средних в генеральной совокупности подтверждается с вероятностью 12%. Поскольку значение опытного уровня значимости больше принятого критического уровня ($p=0,05$ или 5%), то альтернативная гипотеза о различии средних величин не может быть принята, и значит, различия не подтверждены. В такой ситуации можно провести дополнительное исследование с теми же условиями опыта, но с увеличенным числом единиц наблюдения, что на более качественном уровне подтвердит или опровергнет рабочую гипотезу.

Вывод: Различия средней частоты пульса пациентов 1-го и 2-го отделений **НЕ** достоверны. Значит, более высокая средняя частота пульса во 2-м отделении больницы (126,2 уд/мин) по сравнению с ЧСС в 1-м отделении (121,9 уд/мин) не подтверждается при уровне значимости $p=0,05$.

Пример сравнения относительных величин и определения достоверности различий между ними по критерию Стьюдента

Условие задачи: группа животных в количестве 120 особей получала препарат А. Из них у 98 животных произошло восстановление функций организма. Контрольная группа животных в составе 50 особей содержалась в аналогичных условиях без применения этого препарата, из них восстановление наблюдалось у 15 особей.

Задание: а) вычислить показатели частоты восстановления функций организма животных (интенсивные относительные величины) в 1-ой и 2-ой группах животных; б) вычислить ошибки репрезентативности относительных величин;

$$m_2 = \sqrt{\frac{15,31 * (100 - 15,31)}{50 - 1}} = 4,63\%.$$

в) определение доверительных границ относительных величин в каждой группе: при уровне значимости $p < 0,05$, т.е. с вероятностью прогноза более 95%, границы вычисляются по формуле $P \pm 2m$, где P – относительная величина, m – ошибка репрезентативности.

По условию задачи в 1-й группе животных $P_1 = 81,67$, $m_1 = 3,53$. Следовательно, $81,7 \pm 2 * 3,53 = 81,7 \pm 7,06\%$. Получаем доверительные границы колебаний относительных величин в 1-й группе от 74,6% до 88,7%, во 2-й группе от 20,7% до 39,3%. Поскольку доверительные границы не пересекаются, с вероятностью 95% справедливо утверждение, что полученная разница относительных величин не случайна.

г) вычисление критерия Стьюдента для относительных величин:
$$t = \left| \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \right|,$$

$$t = \text{ABS}((81,7 - 20,7) / \text{КОРЕНЬ}(3,53^2 + 4,63^2)) = 8,9 > 2.$$

Вывод: восстановление функций организма животных на фоне действия препарата А проявляется в 81,6%. Этот показатель достоверно выше, чем в контрольной группе животных, не получавших препарат, при уровне значимости $p < 0,05$.

ЗАДАНИЯ

В программе Excel выполните вычисление критерия Стьюдента и оцените достоверность различия средних величин.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «КритСтьюдента» и выполните задачи варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки решения задач преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

- а) В районе N, где расположена тепловая электростанция, в одной из точек жилого поселка было взято 50 проб атмосферного воздуха. Уровень пыли составил: 0,14 мг/м³ в 15-и пробах, 0,16 мг/м³ в 8-ти пробах, 0,13 мг/м³ в 2-х пробах, 0,2 мг/м³ в 15-ти пробах, 0,18 мг/м³ в 6-ти пробах, 0,17 мг/м³ в 4-х пробах. После установки золоуловителя количество пыли в пробах воздуха измерялось следующими цифрами: 0,09 мг/м³ в 2-х пробах, 0,08 мг/м³ в 2-х пробах, 0,05 мг/м³ в 16-ти пробах, 0,02 мг/м³ в 20-и пробах, 0,14 мг/м³ в 2-х пробах. Составьте простые вариационные ряды. Определите среднюю концентрацию пыли в атмосферном воздухе до и после установки золоуловителя. Определите, достоверно ли уменьшилась среднесуточная концентрация пыли после введения в действие золоуловителя с уровнем значимости $p < 0,05$?
- б) Группа больных в количестве 130 человек применяла при лечении лекарственный препарат Z в течение 5 дней. У 106 человек наступило полное выздоровление. Определите частоту выздоровления пациентов на 100 больных и доверительные границы с вероятностью безошибочного прогноза 95%, при которых может наступать выздоровление. Оцените достоверность отличия этого показателя от аналогичного в контрольной группе больных, если известно, что он составил $P = 58,3\%$, $m = \pm 0,63\%$.

Вариант 2

- а) В питьевой воде, которой снабжаются дома жителей города N, определялась концентрация соединений фтора. В 4-х пробах было обнаружено 0,5 мг/л этих соединений, в 5-и – 0,6 мг/л, в 2-х – 0,3 мг/л, в 3-х – 0,4 мг/л, в 8-и – 0,8 мг/л, в 10-и – 0,9 мг/л, в 9-и – 1,2 мг/л, в 12-и – 1,1 мг/л, в 8-и – 1,3 мг/л, в 17-и – 1,0 мг/л, в 8-и – 1,5 мг/л, в 7-и – 1,6 мг/л, в 7-и – 0,7 мг/л, в 9-и – 1,4 мг/л, в 1-й – 0,2 мг/л. Одновременно в городе M были получены следующие результаты: в 15-и пробах было обнаружено 0,1 мг/л соединений фтора, в 10-и – 0,09 мг/л, в 12-и – 0,2 мг/л, в 8-и – 0,05 мг/л, в 9-и – 0,08 мг/л, в 17-и – 0,15 мг/л, в 18-и – 0,13 мг/л, в 2-х – 1,1 мг/л, в 2-х – 0,01 мг/л, в 3-х – 1,0 мг/л, в 3-х – 0,02 мг/л, в 4-х – 0,03 мг/л, в 9-и – 0,3 мг/л, в 5-и – 0,4 мг/л, в 4-х – 0,7 мг/л. Составьте простые вариационные ряды. Определите среднюю концентрацию соединений фтора в питьевой воде городов N и M. Установите, достоверно ли отличается средняя концентрация фторидов в питьевой воде города N от уровня фтора в воде города M при уровне значимости $p < 0,05$?
- б) При обследовании 280 учащихся 3-х классов пяти школ района K обнаружено, что у 64 из них наблюдается нарушение осанки. Определите распространенность этих нарушений на 100 учеников и доверительные границы частоты нарушения осанки у школьников 3-х классов остальных школ района K с вероятностью безошибочного прогноза 95%. Оцените достоверность отличия этого показателя от аналогичного в соседнем районе, если известно, что он составил $P = 35,5\%$, $m = \pm 0,42\%$.

Вариант 3

- а) При обследовании группы школьников 4-х классов сельского района A было установлено, что 2 человека имели по 5 кариозных зубов, 28 человек – по 1

кариозному зубу, 8 человек – по 4 кариозных зуба, 1 человек – 8 кариозных зубов, 20 человек – по 3 кариозных зуба, 16 человек – по 2 кариозных зуба и 6 человек не имели пораженных кариесом зубов. При обследовании аналогичной группы школьников в районе Б были получены следующие результаты: 8 человек по 1 кариозному зубу, 15 человек – по 4 кариозных зуба, 26 человека имели по 5 кариозных зубов, 1 человек – 8 кариозных зубов, 10 человек – по 3 кариозных зуба, 11 человек – по 2 кариозных зуба и 2 человека не имели пораженных кариесом зубов. Составьте простые вариационные ряды. Определите средние величины поражения кариесом школьников районов А и Б. Установите, достоверно ли они отличаются при уровне значимости $p < 0,05$?

- б) При выборочном обследовании 220 рабочих одного из промышленных предприятий у 47 из них были выявлены гастроэнтерологические заболевания. Определите частоту встречаемости этих заболеваний на 100 обследованных и доверительные границы возможной частоты гастроэнтерологических заболеваний среди всех работающих в аналогичных условиях с уровнем вероятности 95%. Оцените достоверность отличия этого показателя от аналогичного показателя на другом предприятии, если известно, что он составил $P = 12,5\%$, $m = \pm 0,25\%$.

Вариант 4

- а) Перед сдачей экзамена у студентов определялась частота пульса. Были получены следующие данные: у 2-х студентов – 76 ударов в минуту, у 3-х студентов – 80 ударов в минуту, у 4-х студентов – 108 ударов в минуту, у 2-х студентов – 116 ударов в минуту, у 20-и студентов – 88 ударов в минуту, у 6-и студентов – 98 ударов в минуту, у 17-и студентов – 86 ударов в минуту, у 11-и студентов – 92 удара в минуту. Составьте простой вариационный ряд. Определите среднюю частоту пульса у студентов перед экзаменом. Достоверно ли отличается показатель частоты пульса перед экзаменом от частоты пульса у этих же студентов после экзамена (при уровне значимости $p < 0,05$), если известно, что она составляла у 19-и студентов – 78 ударов в минуту, у 3-х студентов – 76 ударов в минуту, у 3-х студентов – 75 ударов в минуту, у 4-х студентов – 75 ударов в минуту, у 2-х студентов – 85 ударов в минуту, у 6-и студентов – 82 удара в минуту, у 17-и студентов – 79 ударов в минуту, у 11-и студентов – 81 удар в минуту?
- б) Было осмотрено 185 учеников 5-х классов. У 26 из них обнаружена миопия. Определите распространенность миопии школьников 5-х классов на 100 учащихся и доверительные границы возможной частоты близорукости у школьников данного района с уровнем вероятности 95%. Оцените достоверность отличия распространенности миопии школьников района от аналогичного показателя в другом районе, если известно, что он составил $P = 25,5\%$, $m = \pm 0,31\%$.

Вариант 5

- а) Исследовалась длительность лечения больных пневмонией в стационаре центральной районной больницы района N. Были получены следующие результаты: 25 дней лечилось 2 больных, 26 дней – 1 больной, 11 дней – 1 больной, 12 дней – 1 больной, 23 дня – 3 больных, 13 дней – 1 больной, 21 день – 3 больных, 24 дня – 1 больной, 22 дня – 3 больных, 14 дней – 2 больных, 20 дней – 5 больных, 15 дней – 2

больных, 16 дней – 3 больных, 17 дней – 4 больных, 19 дней – 8 больных, 18 дней – 7 больных. Составьте простой ранжированный вариационный ряд. Рассчитайте среднюю длительность лечения пневмонии. Достоверно ли она отличается от аналогичного показателя больницы соседнего района М (при уровне значимости $p < 0,05$), если известно, что во этой больнице 25 дней лечился 1 больной, 21 день – 2 больных, 11 дней – 26 больных, 12 дней – 18 больных, 23 дня – 3 больных, 13 дней – 10 больных, 21 день – 3 больных, 24 дня – 1 больной, 22 дня – 3 больных, 14 дней – 6 больных, 20 дней – 5 больных, 15 дней – 7 больных, 16 дней – 5 больных, 17 дней – 4 больных, 19 дней – 1 больной, 18 дней – 1 больной?

- б) Исследовано 110 больных абсцессом легкого, у 36 из них обнаружена дистрофия пародонта. Определите распространенность этой патологии на 100 человек, доверительные границы возможной частоты дистрофии пародонта при абсцессе легкого с уровнем вероятности 95%. Оцените достоверность отличия распространенности этого заболевания от аналогичного показателя в контрольной группе пациентов, если известно, что он составил $P = 1,8\%$, $m = \pm 0,07\%$.

Вариант 6

- а) Исследовалась длина тела новорожденных девочек по данным родильного дома. Были получены следующие данные: у 8 девочек рост составил 48 см, у 6 девочек – 51 см, у 7 девочек – 53 см, у 1 девочки – 49 см, у 9 девочек – 52 см, у 8 девочек – 50 см, у 1 девочки – 47 см, у 2 девочек – 46 см, у 2 девочек – 54 см, у 1 девочки – 55 см, у 1 девочки – 56 см. Составьте простой ранжированный сгруппированный вариационный ряд, определите среднюю длину тела новорожденных девочек. Достоверно ли она отличается от длины тела новорожденных мальчиков (при уровне значимости $p < 0,05$), если по данным этого же родильного дома 1 мальчик родился с длиной тела 56 см, 1 мальчик – 55 см, 3 мальчика – 46 см, 7 мальчиков – 52 см, 2 мальчика – 45 см, 10 мальчиков – 50 см, 3 мальчика – 54 см, 5 мальчиков – 53 см, 5 мальчиков – 49 см, 11 мальчиков – 51 см, 2 мальчика – 48 см, 3 мальчика – 47 см?
- б) При выборочном обследовании 150 ткачих хлопчатобумажного комбината у 32 из них обнаружена гинекологическая патология. Определите распространенность этих заболеваний на 100 обследованных и доверительные границы возможной частоты этой патологии у всех работниц комбината с уровнем вероятности 95%. Оцените достоверность отличия распространенности гинекологической заболеваемости от аналогичного показателя другой фабрики, если известно, что она составила $P = 2,8\%$, $m = \pm 0,44\%$.

1. Как обозначаются статистические гипотезы.
2. Какая гипотеза используется при сравнении средних величин?
3. Как называется вероятность опровержения правильной гипотезы?
4. Какой уровень значимости используется в медико-биологических исследованиях?
5. Формула вычисления достоверности различий средних величин по критерию Стьюдента.
6. Методика вычисления критерия достоверности Стьюдента для относительных величин?
7. Оценка критерия Стьюдента.

Рекомендуемая литература:

1. Гельман В.Я. Медицинская информатика. Практикум. СПб: Питер, 2001. – 420 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика : Учебное пособие – 12-е изд., перераб. и доп. – М. : Юрайт, 2011. – 478 с. : ил.
3. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
4. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>
5. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.

VIII. Критерий Хи-квадрат

Анализ характера распределения данных (его еще называют проверкой на нормальность распределения) должен осуществляться для каждого учетного признака, и если установлено, что признак не является нормально распределенным, применение критерия достоверности Стьюдента не оправдано. Это, прежде всего, относится к дискретным и биномиальным данным, которые выражаются в баллах или строго определенными числовыми значениями.

Непараметрические критерии проверки статистических гипотез должны использоваться в том случае, когда распределение учетного признака отличается от нормального. Особенно эффективно применение непараметрических критериев при малых выборках ($n < 30$), а также при изучении качественных признаков. Большинство из них обладает достаточно высокой статистической мощностью (чувствительностью), однако несколько уступают аналогичным параметрическим критериям.

Наиболее часто в медицинских исследованиях применяется критерий достоверности Хи-квадрат (χ^2).

Формула вычисления критерия Хи-квадрат:

$$\chi^2 = (\mathcal{E} - T)^2 / T,$$

где: \mathcal{E} - эмпирическая частота появления признака, т.е. полученная в опыте;

T - теоретическая частота, рассчитанная по нулевой гипотезе (что было бы, если бы группы были одинаковы).

Под частотой понимается количество появлений какого-либо события. Обычно с частотой появления события имеют дело, когда переменные измерены в шкале наименований и другой их характеристики, кроме частоты, подобрать невозможно или сложно. Такие признаки применяются многими исследователями, которые используют балльную оценку величины явления, например: высокий, средний, низкий уровни и т.д.

Пример определения достоверности различий ЧСС в группах детей, поступающих в отделения больницы, по критерию Хи-квадрат

Задача: требуется сравнить частоту сердечных сокращений (ЧСС) детей 1-го года жизни, поступающих в отделения №1 и №2 больницы N (см. пример задачи в разделе VI).

Задание: определить достоверность различий частоты пульса детей, поступающих в 1-е и 2-е отделения больницы, по критерию Хи-квадрат и сделать вывод.

Решение: необходимо запустить программу Excel, создать новый лист и построить сгруппированные вариационные ряды, как показано в таблице 20, или скопировать

таблицы сгруппированных вариационных рядов, построенных в предыдущем разделе (см. раздел VI).

Таблица 20

Результаты измерения частоты пульса детей в 2-х отделениях больницы

1-й вариационный ряд: Частота пульса детей, поступивших в отделение №1 больницы в 20... году		
	V	p
1	110	1
2	115	3
3	120	5
4	125	4
5	130	2
6	135	1

2-й вариационный ряд: Частота пульса детей, поступивших в отделение №2 больницы в 20... году		
	V	P
1	115	3
2	120	2
3	125	6
4	130	3
5	135	1
6	140	1
7	145	1

Для вычислений создается дополнительная таблица, объединяющая данные двух вариационных рядов и содержащая расчетные формулы. Результат вычислений показан в таблице 21.

Таблица 21

Вычисление теоретических частот и критерия Хи-квадрат

ЧСС		1. Эмпирические частоты баллов (Э)			2. Теоретические частоты (Т) $T = \text{Всего} * \text{Итого} / n$			3. Вычисление $\chi^2 = (Э - Т)^2 / Т$		
		1-е отд-е	2-е отд-е	Всего	1-е отд-е	2-е отд-е	Всего	1-е отд-е	2-е отд-е	Всего
1	110	1	0	1	0,48	0,52	1	0,55	0,52	1,06
2	115	3	3	6	2,91	3,09	6	0,00	0,00	0,01
3	120	5	2	7	3,39	3,61	7	0,76	0,72	1,48
4	125	4	6	10	4,85	5,15	10	0,15	0,14	0,29
5	130	2	3	5	2,42	2,58	5	0,07	0,07	0,14
6	135	1	1	2	0,97	1,03	2	0,00	0,00	0,00
7	140	0	1	1	0,48	0,52	1	0,48	0,46	0,94
8	145	0	1	1	0,48	0,52	1	0,48	0,46	0,94
	Итого:	16	17	n=33	16	17	33	2,50	2,36	4,86

Расчеты выполняются в пять этапов:

- Объединение сравниваемых вариационных рядов и вычисление опытных (эмпирических) частот. Эмпирические частоты (Э) - это количество единиц наблюдения по баллам, вычисляем из вариационных рядов ручным подсчетом или функцией =СЧЁТЕСЛИ(Диапазон ячеек;Значение).
- Вычисление теоретических частот. Теоретические частоты (Т) рассчитывают из эмпирических частот как среднее значение в каждом отделении, например: $0,48 = 1 * 16 / 33$, $0,52 = 1 * 17 / 33$ и т.д. Итоги теоретических частот должны совпасть с итогами частот в эксперименте.

3. Вычисление опытного (эмпирического) значения критерия Хи-квадрат.

В ячейках каждого из сравниваемых столбцов по строкам таблицы используется формула: $(\mathcal{E} - T)^2 / T$, а затем суммируется строка «Итого» или столбец «Всего».

Общая формула вычислений имеет вид:

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(\mathcal{E}_{ij} - T_{ij})^2}{T_{ij}} = \underline{4,86}.$$

4. Определение критического значения критерия Хи-квадрат при выбранном уровне значимости и с учетом степени свободы (df), которая зависит от размера таблицы или непосредственно вероятность различий.

Первоначально выбирается уровень значимости, как правило, $p = 0,05$. Затем вычисляют степень свободы по формуле:

$$df = (R - 1) * (C - 1),$$

где: R – количество групп (в данном случае строк) в таблице ($R = 8$);

C – количество столбцов опытных данных ($C = 2$).

$$df = (8 - 1) * (2 - 1) = 7$$

Критическое значение определяется по таблице, приведенной в приложении 2, или вычисляется функцией =ХИ2ОБР(0,05;7), получаем величину = **14,06714**.

Для расчета вероятности различий используется функция =ХИ2ТЕСТ(Опытный интервал; Теоретический интервал). Её применение позволяет сократить вычисления, используя диапазоны ячеек с данными таблицы частот, и возвращает значение вероятности различий.

Расчетная вероятность различий по функции ХИ2ТЕСТ = **0,677=67,7%**.

5. Определение достоверности различий и формулировка вывода.

Производится сравнение опытного значения критерия Хи-квадрат с критическим значением или непосредственная оценка расчетного уровня значимости. Опытное значение критерия Хи-квадрат = 4,86, критическое = 14,07, т.е. опытное значение (4,86) МЕНЬШЕ критического значения (14,07), что указывает на случайный характер различий выборочных совокупностей.

Аналогичный вывод получаем на основе оценки расчетной вероятности различий: **0,677** > 0,05, т.е. вероятность гипотезы о равенстве совокупностей составляет 67,75%, что недостаточно для подтверждения различий при выбранном уровне значимости 5% ($p = 0,05$).

Вывод: различия частоты пульса в 2-х отделениях Недостойны при уровне значимости $p < 0,05$.

Вывод, сделанный на основе вычисления критерия Хи-квадрат, в основном согласуется с выводом, сделанным по критерию Стьюдента. Однако вероятности

того, что выборки взяты из одной генеральной совокупности, существенно отличаются. При использовании критерия Стьюдента эта вероятность составила 12% (0,12), а по критерию Хи-квадрат 67,7% (0,667). Значит, довод о равенстве совокупностей, полученный с применением критерия Хи-квадрат, более весомый.

ЗАДАНИЯ

В программе Excel выполните вычисление критерия Хи-квадрат и оцените достоверность различия вариационных рядов.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «КритХи-квадрат» и выполните задачу варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки решения задачи преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

В районе N, где расположена тепловая электростанция, в одной из точек жилого поселка было взято 50 проб атмосферного воздуха. Уровень пыли составил: 0,14 мг/м³ в 15-и пробах, 0,16 мг/м³ в 8-и пробах, 0,13 мг/м³ в 2-х пробах, 0,2 мг/м³ в 15-и пробах, 0,18 мг/м³ в 6-и пробах, 0,17 мг/м³ в 4-х пробах.

После установки золоуловителя количество пыли в пробах воздуха измерялось следующими цифрами: 0,09 мг/м³ в 2-х пробах, 0,08 мг/м³ в 2-х пробах, 0,05 мг/м³ в 16-и пробах, 0,02 мг/м³ в 20-и пробах, 0,14 г/м³ в 2-х пробах.

Определите, достоверно ли уменьшение запыленности после установки золоуловителя по критерию Хи-квадрат с уровнем значимости $p < 0,05$?

Вариант 2

В питьевой воде, которой снабжаются дома жителей города N, определяли концентрацию соединений фтора, в 2-х пробах было обнаружено 0,5 мг/л этих соединений, в 4-х пробах – 0,6 мг/л, в 8-и пробах – 0,9 мг/л, в 8-и пробах – 0,4 мг/л, в 16-и пробах – 0,8 мг/л, в 16-и пробах – 0,9 мг/л, в 20-и пробах – 1,2 мг/л, в 24-х пробах – 1,1 мг/л, в 40 пробах – 1,3 мг/л, в 50-и пробах – 1,0 мг/л, в 24-х пробах – 1,5 мг/л, в 20-и пробах – 1,6 мг/л, в 10-и пробах – 0,7 мг/л, в 8-и пробах – 1,4 мг/л, в 4-х пробах – 0,3 мг/л.

Одновременно в городе M были получены следующие результаты: в 20-и пробах было обнаружено 0,1 мг/л соединений фтора, в 15-и пробах – 0,09 мг/л, в 8-и пробах – 0,2 мг/л, в 8-и пробах – 0,05 мг/л, в 16-и пробах – 0,08 мг/л, в 10-и пробах – 0,15 мг/л, в 30-и пробах – 0,1 мг/л, в 12-и пробах – 1,1 мг/л, в 14-и пробах – 1,3 мг/л, в 5-ти пробах – 1,0 мг/л, в 4-х пробах – 1,5 мг/л, в 2-х пробах – 1,6 мг/л, в 1-й пробе – 0,7 мг/л, в 8-и пробах – 0,4 мг/л, в 4-х пробах – 0,3 мг/л.

Определите среднюю концентрацию фторидов в питьевой воде городов N и M. Установите, достоверно ли отличается уровни фтора в питьевой воде этих городов по критерию Хи-квадрат с уровнем значимости $p < 0,05$?

Вариант 3

При обследовании группы школьников 4-х классов сельского района А было установлено, что 8 человек имели по 5 кариозных зубов, 20 человек – по 1 зубу, 10 человек – по 4 зуба, 1 человек – 8 зубов, 20 человек – по 3 зуба, 16 человек – по 2 зуба и 15 человек не имели пораженных кариесом зубов.

При обследовании аналогичной группы школьников в районе Б были получены следующие результаты: 2 человека имели по 5 кариозных зубов, 28 человек – 1 зуб, 8 человек – по 4 зуба, 1 человек – 8 зубов, 20 человек – по 3 зуба, 16 человек – по 2 зуба и 6 школьников не имели пораженных кариесом зубов.

Сравните среднюю интенсивность поражения кариесом в районах А и Б, установите, достоверно ли отличаются районы по этому признаку на основе критерия Хи-квадрат с уровнем значимости $p < 0,05$.

Вариант 4

Перед сдачей экзамена у студентов определялась частота пульса (ЧСС). Были получены следующие данные: у 2-х студентов – 76 ударов в минуту, у 3-х студентов – 80 ударов в минуту, у 4-х студентов – 108 ударов в минуту, у 2-х студентов – 116 ударов в минуту, у 20-и студентов – 88 ударов в минуту, у 6-и студентов – 98 ударов в минуту, у 17-и студентов – 86 ударов в минуту, у 11-и студентов – 92 ударов в минуту.

У этих же студентов была измерена частота пульса после экзамена и получены следующие результаты: у 2-х студентов – 73 удара в минуту, у 3-х студентов – 75, у 4-х студентов – 80, у 2-х – студентов 81, у 20-и студентов – 74, у 6-и студентов – 82, у 17-и студентов – 86, у 11-и студентов – 85.

Сравните средние величины ЧСС студентов до и после экзамена и подтвердите достоверность различий этого признака по критерию Хи-квадрат с уровнем значимости $p < 0,05$.

Вариант 5

Исследовалась длительность лечения больных пневмонией в стационаре 2-х больниц города N. Были получены следующие результаты: в 1-й больнице 25 дней лечилось 10 больных, 26 дней – 8 больных, 11 дней – 1 больной, 12 дней – 1 больной, 23 дня – 5 больных, 13 дней – 1 больной, 21 день – 15 больных, 24 дня – 9 больных, 22 дня – 7 больных, 14 дней – 2 больных, 20 дней – 5 больных, 15 дней – 2 больных, 16 дней – 3 больных, 17 дней – 4 больных, 19 дней – 2 больных, 18 дней – 3 больных.

Во 2-й больнице 25 дней лечился 1 больной, 21 день – 2 больных, 11 дней – 26 больных, 12 дней – 18 больных, 23 дня – 3 больных, 13 дней – 10 больных, 21 день – 3 больных, 24 дня – 1 больной, 22 дня – 3 больных, 14 дней – 6 больных, 20 дней – 5 больных, 15 дней – 7 больных, 16 дней – 5 больных, 17 дней – 4 больных, 19 дней – 1 больной, 18 дней – 1 больной.

Рассчитайте среднюю длительность лечения пневмонии и определите достоверность различий этого показателя в 2-х больницах с помощью критерия Хи-квадрат с уровнем значимости $p < 0,05$.

Вариант 6

Изучалась длина тела новорожденных по данным родильного дома. Были получены следующие данные о новорожденных девочках: у 8 девочек рост составил 48 см, у 6 девочек – 51 см, у 7 девочек – 53 см, у 1 девочки – 49 см, у 9 девочек – 52 см, у 8 девочек – 50 см, у 3 девочек – 47 см, у 2 девочек – 46 см, у 2 девочек – 54 см, у 1 девочки – 55 см, у 1 девочки – 56 см.

Данные о новорожденных мальчиках: 9 мальчиков родились с длиной тела 56 см, 6 мальчиков – 55 см, 7 мальчиков – 58 см, 1 мальчик – 59 см, 9 мальчиков – 52 см, 8 мальчиков – 54 см, 2 мальчика – 50 см, 1 мальчик – 49 см, 2 мальчика – 53 см, 4 мальчика – 54 см, 2 мальчика – 51 см, 2 мальчика – 53 см.

Определите среднюю длину тела новорожденных девочек и мальчиков, установите, достоверно ли отличается этот признак у новорожденных женского и мужского пола, подтвердите различия по критерию Хи-квадрат с уровнем значимости $p < 0,05$.

Вопросы для самоконтроля

1. Особенности применения критерия соответствия Хи-квадрат.
2. Понятие о нулевой гипотезе.
3. Методика расчета ожидаемых чисел.
4. Принцип вычисления критерия Хи-квадрат.
5. Для чего выполняется определение критического уровня критерия Хи-квадрат?

Рекомендуемая литература:

1. Гельман В.Я. Медицинская информатика. Практикум. СПб: Питер, 2001. – 420 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика : Учебное пособие – 12-е изд., перераб. и доп. – М. : Юрайт, 2011. – 478 с. : ил.
3. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
4. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>
5. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.

IX. Метод стандартизации

Основным принципом подбора групп статистического наблюдения в случае необходимости сравнения между ними является однородность сравниваемых совокупностей по характеризующим признакам.

К характеризующим признакам относятся признаки, по которым выполняется группировка совокупностей. Они, как правило, оказывают влияние на результативные признаки и изучаемое явление в целом. Подбор групп с одинаковым уровнем характеризующих признаков позволяет исключить такое влияние и акцентировать внимание на изучаемых особенностях явления.

Например, общие интенсивные коэффициенты (рождаемости, смертности, младенческой смертности, заболеваемости и т.д.) правильно отражают частоту явлений при их сопоставлении лишь в том случае, если состав сравниваемых совокупностей однороден. Если же они имеют различный возрастно-половой или профессиональный состав, разное соотношение по нозологическим формам или другим характеризующим признакам, то, ориентируясь на общие показатели, сопоставляя их, можно сделать неправильный вывод об истинных причинах разницы показателей сравниваемых совокупностей.

Если же не удастся провести формирование однородных выборок, или уже после проведения исследования оказывается, что сравниваемые группы не одинаковы по характеризующим признакам, может использоваться МЕТОД СТАНДАРТИЗАЦИИ. Это метод расчета условных (стандартизованных) показателей, заменяющих общие интенсивные (или средние) величины в тех случаях, когда их сравнение затруднено из-за несопоставимости групп. Он позволяет определить, какие показатели были бы получены в случае однородности групп, т.е. устранить (элиминировать) возможное влияние различий в составе совокупностей по какому-либо признаку на величину сравниваемых интенсивных показателей. Следовательно, метод стандартизации применяется тогда, когда имеющиеся различия в составе сравниваемых совокупностей могут повлиять на размеры общих коэффициентов.

Стандартизованные показатели – это условные, гипотетические величины, они не отражают истинных размеров явлений, но свидетельствуют о том, каковы были бы значения сравниваемых интенсивных показателей, если бы были исключены различия в составах совокупностей.

Для того, чтобы устранить влияние неоднородности составов сравниваемых совокупностей на величину получаемых коэффициентов, их приводят к единому стандарту, то есть условно допускается, что состав совокупностей одинаков. В качестве стандарта можно принять величину среды одной или всех сравниваемых групп, уровень явления в этих группах или какой-либо близкой по существу третьей совокупности. От выбора стандарта зависит способ вычисления, т.е. модификация метода. Принято выделять прямой, обратный и косвенный варианты соответственно. Варианты используемого метода стандартизации не влияют на конечный вывод и определяются исследователем.

Методика выполнения расчетов прямого метода стандартизации состоит из 5 этапов:

- ПЕРВЫЙ ЭТАП. Вычисление общих показателей (относительных или средних величин) отдельно для каждой группы.
- ВТОРОЙ ЭТАП. Выбор стандарта осуществляется произвольно.
- ТРЕТИЙ ЭТАП. Вычисление ожидаемых величин, демонстрирующих уровень явления, который мог бы быть зафиксирован при отсутствии неоднородности в группах.
- ЧЕТВЕРТЫЙ ЭТАП. Определение стандартизованных показателей.
- ПЯТЫЙ ЭТАП. Сравнение групп по стандартизованным показателям и формулирование вывода.

Пример вычисления стандартизованных показателей прямым методом стандартизации

Условия задачи: проведено углубленное обследование жителей 2-х микрорайонов города N, при котором получены следующие данные. В 1-м районе обследовано 795 человек, из них 156 мужчины, 639 женщины. Выявлена стенокардия у 3 мужчин и 30 женщин. Во 2-м районе обследовано 720 человек (585 мужчин, 135 женщин). Заболевание выявлено у 15 мужчин и 12 женщин.

Задание: сравнить патологическую пораженность стенокардией в 2-х районах города.

Решение: необходимо запустить программу Excel, ввести данные условия задачи и выполнить вычисления по этапам, как показано ниже, сохранить файл.

1-й этап: создание статистической таблицы и вычисление общих относительных величин ($P_{\text{общ}}$), как показано в таблице 22.

Таблица 22

Вычисление относительных и стандартизованных величин

Районы:	Абсолютные данные				1. Патологическая пораженность на 100 обл. ($P_{\text{общ}}$)		2. Стандарт (1-й + 2-й районы)	3. Ожидаемые величины по районам	
	1-й		2-й		1-й	2-й		1-й	2-й
Пол	Обсл.	Заб.	Обсл.	Заб.	1-й	2-й		1-й	2-й
м	156	3	585	15	1,9	2,6	741	14,3	19,0
ж	639	30	135	12	4,7	8,9	774	36,3	68,8
Оба пола	795	33	720	27	4,2	3,8	1515	50,6	87,8
4. Стандартизованные величины ($P_{\text{ст.}}$) =								3,3	5,8

2-й этап: выбор стандарта. Для прямого метода стандартизации за стандарт выбирается число обследованных в обоих районах (уровень среды) среди мужчин и женщин отдельно, а затем их сумма, т.е. $741 = 156 + 585$ и т.д.

3-й этап: вычисление ожидаемых величин. Они вычисляются с помощью пропорций, которыми определяют возможное число заболевших, если бы в изучаемых районах был одинаков состав среды:

Из 100 болевших	-	$P_{\text{общ}}$	$x = \frac{P_{\text{общ}} * \text{Стандарт}}{100}$
Из числа стандарта	-	$X_{\text{Ожид. число болевших}}$	

Ожидаемая величина болевших мужчин района 1:

Из 100 муж. заболело	-	$P_{\text{общ}}$	$x = \frac{1,9 * 741}{100} = \underline{14,3}$
Из числа станд. муж.	-	$X_{\text{Ожид. число болевших}}$	

Ожидаемая величина болевших женщин района 1:

Из 100 жен. заболело	-	$P_{\text{общ}}$	$x = \frac{4,7 * 774}{100} = \underline{36,3}$
Из числа станд. жен.	-	$X_{\text{Ожид. число болевших}}$	

Аналогично составляются пропорции для района 2.

Ожидаемые величины по каждому району суммируются.

Для района 1: $14,3 + 36,3 = \underline{50,6}$.

Для района 2: $19,0 + 68,8 = \underline{87,8}$.

4-й этап: вычисление стандартизованных величин. Расчет выполняется по принципам вычисления относительных величин, но на основе ожидаемых значений с помощью составления пропорции:

Из общего числа обслед.	-	болело число ожидаемых	$P_{\text{см.}} = \frac{\text{Ожид. вел.} * 100}{\text{Общее число обслед.}}$
Из 100 чел.	-	$P_{\text{см.}}$	

Для района 1:

Из 1515 чел	-	болело 50,6	$P_{\text{см1.}} = \frac{50,6 * 100}{1515} = \underline{3,3}$
Из 100 чел.	-	$P_{\text{см.}}$	

Для района 2:

Из 1515 чел	-	болело 87,8	$P_{\text{см2.}} = \frac{87,8 * 100}{1515} = \underline{5,8}$
Из 100 чел.	-	$P_{\text{см.}}$	

5-й этап: оценка достоверности стандартизованных величин, их сопоставление между собой и с относительными величинами и формулирование вывода.

Вычисление критерия достоверности Стьюдента выполняется с целью подтверждения гипотезы о закономерности выявленных различий.

Критерий достоверности Стьюдента для относительных величин рассчитывается по формуле:

$$t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}}$$

где: P_1 – относительная величина (показатель) 1-й группы;
 P_2 – относительная величина (показатель) 2-й группы;
 m_1 – ошибка репрезентативности 1-го показателя;

m_2 – ошибка репрезентативности 2-го показателя.

Для вычисления ошибки репрезентативности относительной величины используется формула:

$$m = \sqrt{\frac{Pq}{n-1}},$$

где: P – величина относительного показателя;

q – величина, обратная P и вычисленная как $(1-P)$, $(100-P)$, $(100-P)$ и т. д., в зависимости от основания, на которое рассчитан показатель;

n – число наблюдений.

В программе Excel значения этих критериев может быть вычислено следующими формулами:

m_1 : =КОРЕНЬ(3,3*(100-3,3)/(795-1)), результат 0,63;

m_2 : =КОРЕНЬ(5,8*(100-5,8)/(720-1)), результат 0,87;

t : =ABS((3,3-5,8)/КОРЕНЬ(0,63^2+0,87^2)), результат 2,3 > 2, т.е. различия достоверны.

Для удобства сопоставления полученных значений составляется таблица, в которой сравнивают интенсивные и стандартизованные показатели (таблица 23).

Таблица 23

Сопоставление интенсивных и стандартизованных величин

Показатели	Район 1	Район 2	Результаты сравнения
Интенсивные ($P_{\text{Общ.}}$)	4,2	3,8	Заболеваемость в 1-м районе > чем во 2-м районе
Стандартизованные ($P_{\text{Ст.}}$)	3,3± 0,63	5,8±0,87	Заболеваемость в 1-м районе < чем во 2-м районе

Вывод: сравнение стандартизованных показателей выявления стенокардии в районах 1 и 2 позволяет сделать заключение, что, если бы состав обследованных по полу в этих районах был одинаков, то показатель заболеваемости стенокардией в 1-ом районе был бы значительно ниже, чем во 2-ом (уровень значимости $p < 0,05$).

Из анализа общих интенсивных показателей такой вывод сделать нельзя, так как на общие показатели оказал влияние разный состав осмотренных по полу в этих районах. Из литературных данных известно, что стенокардия приблизительно в 2 раза чаще встречается среди мужчин, чем среди женщин. Поэтому в 1-ом районе, где были обследованы преимущественно мужчины, общий интенсивный показатель был завышен (4,2 на 100 обследованных), а во 2-ом районе – занижен (3,8 на 100 обследованных), так как были обследованы в основном женщины. Метод стандартизации позволил увидеть истинную картину заболеваемости, устранив влияние фактора, связанного с полом обследованных.

ЗАДАНИЯ

В программе Excel выполните вычисление стандартизованных показателей и сделайте вывод.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «Станд-я» и выполните задание варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки задания преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

Проведено изучение заболеваемости населения в городах N и M. В городе N проживало 30000 человек, из них 20000 мужчин и 10000 женщин. В течение года болели 5000 мужчин и 4000 женщин. В городе M проживало 65000 человек (25000 мужчин, 40000 женщин). Болели в течение года 7000 мужчин и 14000 женщин. Требуется рассчитать интенсивные и стандартизованные показатели заболеваемости (на 1000 жителей), проанализировать и оценить достоверность различий между ними, сделать вывод.

Вариант 2

Требуется сравнить физическое развитие учеников 8-10 классов в 2-х школах. В 1-й школе обследовано 195 человек, из них 31 мальчик и 164 девочки, выявлено, что отставание физического развития есть у 1-го мальчика и 5-ти девочек. Во 2-й школе обследовано 182 человека, из них 125 мальчиков и 57 девочек, имели отставание 5 мальчиков и 4 девочки. Вычислите интенсивные и стандартизованные показатели частоты отклонений физического развития детей в этих школах, (на 100 детей) сравните их и оцените достоверность различий между ними, сделайте вывод.

Вариант 3

Необходимо сравнить качество протезирования зубов в 2-х стоматологических поликлиниках. В 1-й поликлинике изготовлено 400 протезов, из них 35 мостовидных и 365 единичных; плохая фиксация протезов зафиксирована у 1-го пациента с мостовидным и 5-ти пациентов с единичными протезами. Во 2-й поликлинике установлено 350 протезов, из них 250 мостовидных и 100 единичных; нарушение фиксации обнаружено 12-и пациентов с мостовидными и 5-и с единичными протезами. Вычислите интенсивные и стандартизованные показатели

качества протезирования в поликлиниках (на 100 пациентов), проанализируйте и оцените достоверность различий между ними, сделайте вывод.

Вариант 4

Сравните загрязненность воздуха по количеству проб, не отвечающих санитарным требованиям, в 2-х районах города N. В 1-м районе взято 300 проб, из них 30 в жилых кварталах и 270 вблизи автомобильных дорог; обнаружено несоответствие требованиям в 2-х квартальных пробах и 5 у автодороги. Во 2-м районе взято 200 проб, из них 150 квартальных и 50 у дороги; имели отклонение от норматива загрязненности 12 квартальных и 5 проб рядом с автодорогой. Требуется рассчитать интенсивные и стандартизованные показатели загрязненности воздуха в 2-х районах (на 100 проб), проанализировать и оценить достоверность различий между ними, сделать вывод.

Вариант 5

Требуется сравнить качество работы 2-х лечебно-профилактических учреждений (ЛПУ) по числу послеоперационных осложнений. В 1-й клинике проведено 600 хирургических операций, из них 55 полостных в общем отделении и 545 малоинвазивных в эндоскопическом отделении; послеоперационные осложнения зафиксированы у 20-и пациентов, перенесших полостную операцию, и у 5-ти пациентов из эндоскопического отделения. Во 2-й клинике выполнено 350 операций, из них 250 полостных и 100 малоинвазивных; осложнения были зафиксированы у 5-и пациентов общего отделения и у 4-х - эндоскопического. Вычислите интенсивные и стандартизованные показатели послеоперационных осложнений в 2-х ЛПУ (на 100 операций), проанализируйте и оцените достоверность различий между ними, сделайте вывод.

Вариант 6

Сравните загрязненность 2-х водоемов по количеству проб, не отвечающих санитарным требованиям. В 1-м водоеме взято 300 проб, из них 30 поверхностных и 270 с глубины 3 метра; обнаружено несоответствие требованиям в 5-ти поверхностных пробах и 7-ми глубинных. Во 2-м водоеме взято 320 проб, из них 270 поверхностных и 50 глубинных; имели отклонение от норматива загрязненности 8 поверхностных и 3 глубинные пробы. Необходимо рассчитать интенсивные и стандартизованные показатели загрязненности водоемов (на 100 проб), проанализировать и оценить достоверность различий между ними, сделать вывод.

Вопросы для самоконтроля

1. В каких случаях применяется метод стандартизации?
2. В чем заключается суть метода стандартизации?
3. Какие варианты (модификации) метода стандартизации вы знаете?
4. Этапы прямого метода стандартизации.

5. О чем свидетельствуют стандартизованные показатели?
6. Как оцениваются стандартизованные показатели?

Рекомендуемая литература:

1. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>.

Х. Дисперсионный анализ

Трудно представить любое медицинское исследование (социально-гигиеническое, гигиеническое, клиническое, экспериментальное и др.), в котором не ставилась бы в той или иной мере задача определения силы влияния различных факторов на размеры изучаемого признака. Определяя различия средних арифметических двух выборочных групп наблюдения путем расчета критерия Стьюдента t , который позволяет выявить статистическую значимость разницы средних, исследователь подразумевает, как само собой разумеющееся, что группы исследуемых совершенно однородны и отличаются только по одному какому-то признаку или методу воздействия на них.

Между тем на практике это условие соблюдается далеко не всегда. На изучаемое явление и, следовательно, его средний уровень оказывают влияние многочисленные факторы, как постоянные (планируемые или сознательно выделяемые для их изучения), так и случайные (неопределенные). Например, больные гипертонической болезнью, отобранные по полу, возрасту, стадии и длительности заболевания, помимо болезни, подвергаются воздействию других неучтенных факторов, в результате чего у разных больных наблюдается различный уровень артериального давления.

При изучении явлений, сравнении их друг с другом в поисках сходства и различий необходимо обращать внимание не только на величину средних, но и на разнообразие вариант, а также вариабельность изучаемых признаков. Исследователь может встретить вариационные ряды, не отличающиеся по центральной тенденции (размеру средней арифметической), но различные по степени варьирования. И наоборот - ряды, одинаковые по величине разброса вариант, могут различаться по размерам средней арифметической. Установление значимости различий средних арифметических, измерение степени влияния факторов и их градаций на варьирующий (результативный) признак наиболее эффективно достигаются путем применения дисперсионного анализа. Впервые основа дисперсионного анализа была разработана известным английским статистиком Р. Фишером в 1925 году.

ДИСПЕРСИОННЫЙ АНАЛИЗ – это метод в статистической математике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий вариабельности признака в исследуемой совокупности. В литературе также встречается обозначение ANOVA (от англ. ANalysis Of VAriance). Он базируется на определении степени рассеяния (дисперсии) оцениваемых признаков в нескольких группах. Это позволяет измерить силу влияния отдельных факторов на значения показателей.

Известно, что величина отдельных признаков представляет собой результат воздействия разнообразных факторов, различных по силе влияния. Одни факторы имеют значительно большую силу влияния, другие - меньшую. Причем, как правило, факторы сами воздействуют друг на друга, сочетая свое влияние, иногда усиливают действие друг друга, иногда, наоборот, погашают это действие. Преимуществом дисперсионного анализа является то, что он дает возможность изучить и сравнить роль каждого из них.

В отличие от дисперсионного анализа другие общепринятые в медицинских исследованиях статистические методы обработки, как правило, предусматривают

проведение попарных сравнений, что приводит к огромному объему расчетов и часто не дает полной оценки.

Сущность дисперсионного анализа заключается в изучении статистического влияния одного или нескольких факторов на результативный признак.

ФАКТОР - это влияние, воздействие или состояние, которое отражается на размерах и разнообразии результативного признака.

РЕЗУЛЬТАТИВНЫЙ ПРИЗНАК - это элементарное свойство объектов, изучаемое как результат влияния факторов: организованных в исследовании (обозначаются знаком x) и всех остальных, случайных, неорганизованных, не учтенных в данном исследовании (обозначаются знаком z).

ГРАДАЦИИ ФАКТОРА - это степень его воздействия, в том числе отсутствие воздействия (нулевое значение) в контрольной группе или состояние объектов изучения (пол, возраст и др.).

ДИСПЕРСИОННЫЙ КОМПЛЕКС - это совокупность градаций изучаемых данных (групп объектов наблюдения, выбранных исследователем) с вычисленными значениями относительных или средних величин по каждой градации.

При изучении количественных признаков в градации дисперсионного комплекса заносятся числовые результаты измерения изучаемого признака у каждого отдельного объекта. При изучении качественных признаков в градации комплекса заносится число объектов с наличием признака и общее число объектов.

Дисперсионные комплексы, составленные по принципу случайного отбора, называются РАНДОМИЗИРОВАННЫМИ.

Статистическое влияние - это отражение в разнообразии результативного признака того разнообразия фактора (его градаций), которое организовано в исследовании.

Сумма основных и случайных факторов составит общие факторы (y). Результативный признак изучается как результат воздействия факторов, организованных в исследовании (x) и неорганизованных (z).

Общее влияние как раз и определяет влияние всех организованных и неорганизованных (случайных) факторов, определивших такое развитие признака, которое наблюдалось в дисперсионном комплексе. Общее влияние служит базой для определения доли влияний - факториальных и случайных.

Факториальное влияние - это простое или комбинированное статистическое влияние изучаемых (учтенных) факторов.

Случайное влияние - это действие тех факторов, которые не учтены в дисперсионном комплексе и составляют общий фон, на котором действуют учитываемые факторы.

Таким образом, дисперсионный анализ исследует важнейшее свойство совокупности – разнообразие (вариабельность, дисперсию) признака. Для этого выделяется три вида разнообразия: межгрупповое, внутригрупповое и общее. Межгрупповое разнообразие зависит от влияния исследуемого фактора, по которому выделяется каждая группа. Иными словами, межгрупповое разнообразие - это различие средних в каждой группе. Внутригрупповое разнообразие зависит от

силы влияния каких-то неучтенных случайных факторов. Общее разнообразие складывается из межгруппового и внутригруппового.

В основе дисперсионного анализа лежит предположение о том, что одни переменные могут рассматриваться как причины (факторы, независимые переменные): f_1, \dots, f_k , а другие как следствия (зависимые переменные). Независимые переменные называют иногда регулируемыми факторами именно потому, что в эксперименте исследователь имеет возможность варьировать ими и анализировать получающийся результат.

Основной целью дисперсионного анализа является исследование значимости различия между группами с помощью сравнения дисперсий. Разделение общей дисперсии на несколько источников позволяет сравнить дисперсию, вызванную различием между группами, с дисперсией, вызванной внутригрупповой изменчивостью. При истинности нулевой гипотезы (о равенстве средних в нескольких группах наблюдений, выбранных из генеральной совокупности), оценка дисперсии, связанной с внутригрупповой изменчивостью, должна быть близкой к оценке межгрупповой дисперсии. Если выполняется сравнение средних в двух выборках, дисперсионный анализ даст тот же результат, что и обычный t-критерий Стьюдента. Однако, помимо этого, он позволяет оценить степень такого влияния, а также может использоваться при сопоставлении более чем 2-х групп.

Сущность дисперсионного анализа заключается в расчленении общей дисперсии (D или SS) изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак. Обозначение SS - это сокращение от фразы "суммы квадратов" (Англ. Sum of Squares). Оно чаще всего используется в зарубежных источниках.

Сравнивая компоненты дисперсии друг с другом посредством F-критерия Фишера, можно определить, какая доля общей вариативности результативного признака обусловлена действием регулируемых факторов.

Критерий Фишера экспериментальных (эмпирических) данных ($F_{Эмп.}$) вычисляется как отношение среднего квадрата дисперсии (MS), обусловленной изучаемым фактором, к среднему квадрату случайной дисперсии:

$$F_{Эмп.} = \frac{MS_{Факт.}}{MS_{Случ.}}$$

где: $F_{Эмп.}$ – критерий Фишера, вычисленный в исследуемой совокупности,

$MS_{Факт.}$ – средний квадрат дисперсии, обусловленной изучаемым фактором,

$MS_{Случ.}$ – средний квадрат дисперсии, обусловленной случайными факторами.

Для оценки достоверности полученного результата вычисленный на экспериментальных данных критерий Фишера сравнивают с его критическим значением ($F_{Крит.}$) для принятого уровня вероятности (p) и степеней свободы (df).

С целью уменьшения объема вычислений в программе Excel может применяться настройка «Анализ данных» и ее модуль «Однофакторный дисперсионный анализ».

Пример задачи на выявление влияния изучаемого фактора с помощью дисперсионного анализа

Условие задачи: три различные группы из шести испытуемых получили списки из десяти слов. Первой группе слова предъявлялись с низкой скоростью - 1 слово в 5 секунд, второй группе со средней скоростью - 1 слово в 2 секунды, и третьей группе с большой скоростью - 1 слово в секунду. Результаты измерений представлены в таблице 24.

Таблица 24

Результаты запоминания слов, предъявляемых испытуемым

№ испытуемого	Группа 1 (низкая скорость)	Группа 2 (средняя скорость)	Группа 3 (высокая скорость)
1	8	7	4
2	7	8	5
3	9	5	3
4	5	4	6
5	6	6	2
6	8	7	4
Суммы	43	37	24
Средние	7,17	6,17	4
Общая сумма	104		

Предполагается, что показатели запоминания и воспроизведения слов будут зависеть от скорости их предъявления. Поэтому перед выполнением статистического наблюдения были выдвинуты гипотезы:

- Основная (H_0): различия в объеме воспроизведения слов *между* группами являются не более выраженными, чем случайные различия *внутри* каждой группы.
- Альтернативная (H_1): Различия в объеме воспроизведения слов *между* группами являются более выраженными, чем случайные различия *внутри* каждой группы.

Задание: подтвердить одну из указанных выше гипотез.

Решение: необходимо запустить программу Excel, создать новый лист, на этом листе ввести данные и решение задачи, включающее несколько этапов вычислений, как показано ниже, сохранить файл.

Этапы выполнения дисперсионного анализа.

1. Подсчет $SS_{\text{факт.}}$ - вариативности признака, обусловленную действием исследуемого фактора (межгрупповое разнообразие):

$$SS_{\text{факт.}} = \frac{\sum T_c^2}{n} - \frac{(\sum x_i)^2}{N},$$

где: T_c – сумма индивидуальных значений по каждому из условий. Для нашего примера 43, 37, 24 (см. таблицу);

c – количество условий (градаций) фактора (=3);

n – количество испытуемых в каждой группе (=6);

N – общее количество индивидуальных значений (=18);

$(\sum x_i)^2$ – квадрат общей суммы индивидуальных значений (=104²=10816).

Отметим разницу между $\sum x_i^2$, в которой все индивидуальные значения сначала возводятся в квадрат, а потом суммируются, и $(\sum x_i)^2$, где индивидуальные значения сначала суммируются для получения общей суммы, а потом уже эта сумма возводится в квадрат.

Фактическая вариативность признака, вычисляется как:

$$SS_{\text{факт.}} = \frac{(43^2 + 37^2 + 24^2)}{6} - \frac{124^2}{18} = 31,44.$$

2. Вычисление $SS_{\text{общ.}}$ – общей вариативности признака:

$$\begin{aligned} SS_{\text{общ.}} &= \sum x_i^2 - \frac{(\sum x_i)^2}{N} = \\ &= 8^2 + 7^2 + 9^2 + 5^2 + 6^2 + 8^2 + 7^2 + 8^2 \dots + 2^2 + 4^2 - \frac{104^2}{18} = 63,11. \end{aligned}$$

3. Вычисление случайной (остаточной) величины дисперсии $SS_{\text{сл.}}$, обусловленной неучтенными факторами (внутригрупповое разнообразие):

$$SS_{\text{сл.}} = SS_{\text{общ.}} - SS_{\text{факт.}} = 63,11 - 31,44 = 31,67.$$

4. Определение числа степеней свободы $df_{\text{общ.}}$, $df_{\text{факт.}}$, $df_{\text{сл.}}$:

$$df_{\text{общ.}} = N - 1 = 18 - 1 = 17;$$

$$df_{\text{факт.}} = df_1 = C - 1 = 3 - 1 = 2;$$

$$df_{\text{сл.}} = df_2 = df_{\text{общ.}} - df_{\text{факт.}} = 17 - 2 = 15.$$

5. Математическое ожидание суммы квадратов или «средний квадрат», усредненная величина соответствующих сумм квадратов MS равна:

$$MS_{\text{факт.}} = \frac{SS_{\text{факт.}}}{df_{\text{факт.}}} = \frac{31,44}{2} = 15,72;$$

$$MS_{\text{сл.}} = \frac{SS_{\text{сл.}}}{df_{\text{сл.}}} = \frac{31,67}{15} = 2,11.$$

6. Значение статистики критерия $F_{\text{эмп.}}$ вычисляется по формуле:

$$F_{\text{эмп.}} = \frac{MS_{\text{факт.}}}{MS_{\text{сл.}}} = \frac{15,72}{2,11} = 7,45.$$

Для нашего примера имеем: $F_{\text{эмп.}} = \mathbf{7,45}$

7. Определение $F_{\text{крит.}}$ выполняется по статистическим таблицам (Приложение 3) для степеней свободы $df_1=k_1-2$ и $df_2=k_2-15$ и уровне значимости 0,05. Табличное значение статистики равно $F_{\text{крит.}} = \mathbf{3,68}$.

В программе Excel критическое значение критерия Фишера определяется функцией =FРАСПОБР(Уровень значимости; df_1 ; df_2) =FРАСПОБР(0,05;2;15) = 3,68232034.

8. Если $F_{\text{Эмп.}} < F_{\text{крит.}}$, то нулевая гипотеза принимается, в противном случае принимается альтернативная гипотеза. Для нашего примера $F_{\text{Эмп.}} > F_{\text{крит.}}$ ($7,45 > 3,68$), следовательно, принимается альтернативная гипотеза - влияние существует.

В программе Excel с помощью функции ФТЕСТ можно сразу вычислить вероятность различий двух массивов данных. Вводим в ячейку функцию =ФТЕСТ(Первый диапазон данных ; Второй диапазон данных).

Получаем вероятность $0,99999999 > 0,95$ (95%).

Аналогичные вычисления выполняются с помощью надстройки «Анализ данных» в модуле «Однофакторный дисперсионный анализ». Результат обработки данных задачи этой командой показан в таблице 25.

Таблица 25

Однофакторный дисперсионный анализ

ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Группа 1 (низкая скорость)	6	43	7,16667	2,166667		
Группа 2 (средняя скорость)	6	37	6,16667	2,166667		
Группа 3 (высокая скорость)	6	24	4	2		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	31,444	2	15,7222	7,4473684	0,00567184	3,682320344
Внутри групп	31,667	15	2,11111			
Итого	63,111	17				

Вывод: различия в объеме воспроизведения слов между группами являются более выраженными, чем случайные различия внутри каждой группы ($p < 0,05$). Таким образом, скорость предъявления слов влияет на объем их воспроизведения.

ЗАДАНИЯ

С помощью модуля «Однофакторный дисперсионный анализ» программы Excel оцените влияние одного из факторов на изучаемый признак, сформулируйте вывод.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «Дисп_анализ» и выполните задание варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки задания преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

В эксперименте на животных измерено время пробежки мышей по лабиринту на фоне различной концентрации препарата, стимулирующего нервную систему. Результаты измерений в секундах указаны в таблице 26.

Таблица 26

Результаты измерения времени пробежки мышей по лабиринту (сек.)

№ животного	Группа 1 (низкая концентрация)	Группа 2 (средняя концентрация)	Группа 3 (высокая концентрация)
1	8	7	4
2	7	8	5
3	9	5	3
4	5	4	6
5	6	6	2
6	8	4	3
7	7	7	4
8	8	6	2
9	9	7	4
10	8	7	3

Необходимо подтвердить влияние стимулирующего вещества с помощью дисперсионного анализа в программе Excel надстройкой «Анализ данных» модулем «Однофакторный дисперсионный анализ».

Вариант 2

На предприятии проведено изучение уровня травматизма с учетом фактора стажа работы сотрудников 5-и участков с близкими условиями труда, получены следующие данные (таблица 27).

Таблица 27

Уровень травматизма на 100 работающих

Участок	Стаж работы			
	до 5 лет	6-10 лет	11-15 лет	16 лет и более
1	11	8	6	4
2	12	9	7	7
3	10	6	6	5
4	10	9	7	7
5	13	8	5	3

Необходимо оценить влияние стажа работы на уровень травматизма с помощью дисперсионного анализа в программе Excel надстройкой «Анализ данных» модулем «Однофакторный дисперсионный анализ».

Вариант 3

Проведено изучение уровня загрязнения водоема в 10 точках с учетом времени года, получены следующие данные (таблица 28).

Таблица 28

Уровень загрязнения водоема

№ точки отбора	Концентрации (мг/м ³) по временам года			
	зима	весна	лето	осень
1	3	8	6	4
2	4	9	7	7
3	2	6	6	5
4	3	9	7	7
5	1	8	5	3
6	4	8	3	5
7	2	9	2	3
8	3	6	4	5
9	1	9	3	4
10	2	8	4	4

Требуется определить влияние времени года на уровень загрязнения водоема с помощью дисперсионного анализа в программе Excel надстройкой «Анализ данных» модулем «Однофакторный дисперсионный анализ».

Вариант 4

Проведено обследование 8 групп пациентов, которым выполнялось лечение кариеса с применением 3-х типов пломбировочного материала и с учетом времени выполнения работы врача, получены следующие данные (таблица 29).

Таблица 29

Время работы врача-стоматолога (мин)

№ группы пациентов	Тип пломбировочного материала		
	1-й	2-й	3-й
1	3	8	6
2	4	9	7
3	2	6	6
4	3	9	7
5	1	8	5
6	4	8	3
7	2	9	2
8	3	6	4

Необходимо подтвердить влияние типа используемого материала на время работы врача с помощью дисперсионного анализа в программе Excel надстройкой «Анализ данных» модулем «Однофакторный дисперсионный анализ».

Вариант 5

Проведено обследование длительности лечения 6 групп пациентов, которые получали терапию с применением лекарственного препарата, изготовленного тремя фармацевтическими компаниями, получены следующие данные (таблица 30).

Таблица 30

Длительность лечения (дней)

Пациент	Производитель лекарственного средства		
	1-й	2-й	3-й
1	5	6	9
2	4	7	9
3	6	6	8
4	7	8	7
5	4	6	10
6	5	7	6

Необходимо подтвердить влияние на длительность лечения пациентов применения лекарства изготовленного разными фармацевтическими компаниями с

помощью дисперсионного анализа в программе Excel надстройкой «Анализ данных» модулем «Однофакторный дисперсионный анализ».

Вариант 6

Выполнен отбор проб атмосферного воздуха населенного пункта для изучения уровня загрязнения фенолом с учетом расстояния от промышленной зоны, получены следующие данные (таблица 31).

Таблица 31

Уровень загрязнения воздуха фенолом

№ серии отбора проб	Концентрации (мг/м ³) на удалении			
	0,5 км	1 км	1,5 км	2 км
1	0,3	0,08	0,06	0,04
2	0,4	0,09	0,07	0,06
3	0,2	0,06	0,06	0,05
4	0,3	0,09	0,07	0,06
5	0,1	0,08	0,05	0,03
6	0,4	0,08	0,03	0,01
7	0,2	0,09	0,02	0,01
8	0,3	0,06	0,04	0,03
9	0,1	0,09	0,03	0,02
10	0,2	0,08	0,04	0,03
11	0,3	0,11	0,03	0,01
12	0,2	0,09	0,02	0,01

Требуется подтвердить влияние на уровень загрязнения фенолом фактора удаленности от промышленной зоны с помощью дисперсионного анализа в программе Excel надстройкой «Анализ данных» модулем «Однофакторный дисперсионный анализ».

Вопросы для самоконтроля

1. Дисперсионный анализ (определение).
2. Сущность дисперсионного анализа.
3. Объясните, что является результативным признаком в дисперсионном анализе?
4. Понятие «Фактор» в дисперсионном анализе.
5. Межгрупповое, внутригрупповое и общее разнообразие в дисперсионном анализе.
6. Независимые и зависимые переменные.
7. Методика вычисления критерия Фишера.

Рекомендуемая литература:

1. Гельман В.Я. Медицинская информатика. Практикум. СПб: Питер, 2001. – 420 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика : Учебное пособие – 12-е изд., перераб. и доп. – М. : Юрайт, 2011. – 478 с. : ил.
3. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
4. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>

XI. Метод корреляции

При проведении исследования в биологии или медицине, как правило, регистрируются множество учетных признаков. Представляет интерес вопрос об их взаимном изменении, т.е. обнаружение зависимостей между ними. Выявление наличия таких взаимосвязей является одной из важнейших задач любой науки, в том числе и медицины.

Различают две формы количественных связей между явлениями или процессами: функциональную и корреляционную. Под ФУНКЦИОНАЛЬНОЙ понимают такую связь, при которой любому значению одного из признаков соответствует строго определенное значение другого. В точных науках, таких, как физика, химия и другие, может быть установлена функциональная взаимосвязь. Например, зависимость площади круга от длины окружности в геометрии, или в физике длина пути, пройденной телом в свободном падении, от времени. Наиболее известным видом функциональной зависимости является линейная, которая выражается математической формулой: $y = ax + b$.

В биологии и медицине установить функциональную зависимость, как правило, не удастся. Объекты этих исследований имеют большую изменчивость и зависят от огромного числа факторов, измерить которые просто невозможно. В этом случае определяется наличие КОРРЕЛЯЦИОННОЙ связи, при которой значению каждой средней величины одного признака соответствует несколько значений другого взаимосвязанного с ним признака. Например: связь между ростом и массой тела человека. У группы людей с одинаковым ростом наблюдается различная масса тела, однако она варьирует в определенных пределах вокруг средней величины. Поэтому такую зависимость нужно оценивать с использованием понятия случайной величины с привлечением подходов теории вероятности. Такую форму зависимостей называют «Корреляционной».

При поиске зависимости между признаками может быть обнаружена взаимосвязь, различная по направлению и силе:

- Прямая (при увеличении одного признака увеличивается второй);
- Обратная (при увеличении одного признака второй уменьшается).

Степень взаимосвязи признаков по силе (тесноте) принято обозначать как:

- Отсутствие;
- Слабая;
- Средняя;
- Сильная;
- Полная.

Способами выявления корреляционной взаимосвязи между признаками являются:

- Визуальные (таблицы и графики).
- Статистические (корреляция и регрессия).

Следует подчеркнуть, что обнаружение корреляции между двумя признаками еще не говорит о существовании причинной связи между ними, а лишь указывает на

возможность таковой или на наличие фактора, определяющего изменение обеих переменных совместно.

Приёмы визуализации данных позволяют обнаружить корреляционную зависимость лишь при небольшом числе наблюдений и только приблизительно. Для обнаружения корреляционной взаимосвязи с помощью таблицы в ней располагают ранжированные вариационные ряды и затем определяют совместное изменение признаков. График более наглядно демонстрирует такую зависимость и позволяет оценить ее форму: линейная, параболическая, тригонометрическая и др.

Наиболее точным способом обнаружения взаимосвязи между признаками является вычисление коэффициента корреляции. В зависимости от природы обрабатываемых данных применяются параметрические или непараметрические методы вычисления этого коэффициента.

При вычислении коэффициента корреляции исследователь получает возможность судить о силе связи (степени сопряженности) и ее направлении, а также с требуемой долей вероятности делать вывод о проявлении этой связи в генеральной совокупности. Чем больше коэффициент корреляции, тем с большей степенью уверенности можно говорить о наличии корреляционной зависимости между признаками. Если каждому заданному значению одного признака соответствуют близкие друг к другу, тесно расположенные около средней величины значения другого признака, то связь является более тесной. Когда эти значения сильно варьируют, связь менее тесная. Таким образом, мера корреляции указывает, насколько тесно связаны между собой параметры.

Коэффициент корреляции может принимать значения от -1 до +1. Направление обнаруженной взаимосвязи определяют по знаку коэффициента корреляции. При его положительном значении обнаруженная связь является прямой, при отрицательном – обратной. Сила связи оценивается по модулю этого коэффициента. Условно выделяют следующие уровни корреляционной связи: отсутствие – 0; слабая – от 0 до 0,3; средняя – от 0,3 до 0,7; сильная – 0,7 и более; полная – 1. Однако обсуждать наличие корреляции имеет смысл только в тех случаях, когда она статистически значима ($p < 0,05$). Поэтому после вычисления коэффициента корреляции производится определение его ошибки репрезентативности и критерия достоверности.

Наиболее часто применяемыми в настоящее время методами обнаружения корреляции являются параметрический анализ по Пирсону и непараметрический анализ по Спирмену. Этими методами проверяется нулевая гипотеза (H_0) об отсутствии связи между параметрами. Если такая гипотеза отклоняется при заданном уровне значимости (p), можно говорить о наличии взаимосвязи между параметрами.

Корреляционный анализ по Пирсону используется при решении задачи исследования линейной связи двух нормально распределенных параметров. Кроме проверки на нормальность распределения каждого параметра, до проведения корреляционного анализа рекомендуется строить график в координатах оцениваемых параметров, чтобы визуально определить характер зависимости.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ ПИРСОНА (r_{xy}) или коэффициент линейной корреляции, был разработан в 90-х годах XIX века Карлом Пирсоном, Френсисом Эджуортом и Рафаэлем Уэлдоном в Англии. Он рассчитывается по формуле:

$$r_{xy} = \frac{COV_{xy}}{\sigma_x \sigma_y} = \frac{\sum(x - \bar{X})(y - \bar{Y})}{\sqrt{\sum(x - \bar{X})^2 \sum(y - \bar{Y})^2}},$$

где: r_{xy} – коэффициент линейной корреляции Пирсона;
 COV_{xy} – ковариация признаков x и y ;
 σ_x – среднее квадратическое отклонение признака x ;
 σ_y – среднее квадратическое отклонение признака y ;
 \bar{X} – средняя арифметическая признака x ;
 \bar{Y} – средняя арифметическая признака y .

В медицинской литературе встречается упрощенная запись этой формулы:

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}},$$

где: r_{xy} – коэффициент линейной корреляции Пирсона;
 d_x – отклонение каждой варианты признака x от средней этого признака:
 $d_x = x - M_x$,
 d_y – отклонение каждой варианты признака y от средней этого признака:
 $d_y = y - M_y$.

В программе Excel значение коэффициент линейной корреляции Пирсона может быть вычислено функцией =КОРРЕЛ(Диапазон ячеек 1-го ряда; Диапазон ячеек 2-го ряда).

Для прогнозирования уровня корреляции в генеральной совокупности определяют ошибку репрезентативности этого коэффициента m_r . Она вычисляется по формуле:

$$m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}},$$

где: m_r – ошибка репрезентативности коэффициента корреляции;
 r_{xy} – коэффициент линейной корреляции Пирсона;
 n – число парных вариантов.

Достоверность коэффициента линейной корреляции оценивается по коэффициенту Стьюдента (t_r), который вычисляется с использованием его ошибки:

$$t_r = \frac{r_{xy}}{m_r},$$

где: t_r – коэффициент достоверности Стьюдента;
 r_{xy} – коэффициент линейной корреляции Пирсона;
 m_r – ошибка репрезентативности коэффициента корреляции.

Если число парных вариантов $n > 30$, то при $t_r > 2$ связь считается достоверной при уровне значимости $p < 0,05$. Если число парных вариантов $n < 30$, то критическое значение $t_{r-крит.}$ находят по таблице критических значений Стьюдента при степени свободы $df = n - 2$. В программе Excel это значение вычисляется функцией =СТЮДРАСПОБР(Уровень значимости p ; Степени свободы df).

С целью уменьшения объема вычислений может применяться функция =КОРРЕЛ(Диапазон1; Диапазон2) или надстройка «Анализ данных» и ее модуль «Корреляционный анализ».

Отсутствие линейной корреляции еще не означает, что параметры полностью независимы. Связь между ними может быть нелинейной, или признаки, используемые в вычислениях, могут не подчиняться нормальному закону распределения. Поэтому, помимо вычисления коэффициента линейной корреляции, прибегают к использованию непараметрических коэффициентов корреляции. К ним относятся:

- Коэффициент ранговой корреляции Спирмена;
- Коэффициент ранговой корреляции Кендалла;
- Коэффициент корреляции знаков Фехнера;
- Коэффициент множественной ранговой корреляции (конкордации).

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ПО СПИРМЕНУ применяется для обнаружения взаимосвязи двух параметров, если распределение хотя бы одного из них отлично от нормального.

Каждому показателю x и y присваивается ранг. На основе полученных рангов рассчитываются их разности d . Затем вычисляется коэффициент корреляции (ρ) по формуле:

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

где: ρ – коэффициент корреляции Спирмена;
 d – разность рангов;
 n – число парных вариантов.

Ошибка репрезентативности коэффициента корреляции Спирмена определяется по формуле:

$$m_\rho = \sqrt{\frac{1 - \rho^2}{n - 2}},$$

а коэффициент достоверности Стьюдента:

$$t_\rho = \frac{\rho}{m_\rho},$$

где: t_ρ – коэффициент достоверности Стьюдента;
 ρ – коэффициент корреляции Спирмена;
 m_ρ – ошибка репрезентативности коэффициента корреляции Спирмена.

Оценка коэффициента корреляции Спирмена и его достоверности выполняется так же, как и коэффициента линейной корреляции Пирсона.

Пример решения задачи на выявление корреляционной зависимости

Условие задачи: выполнены измерения уровня запыленности на рабочих местах работников с учетом температуры в помещении (таблица 30).

Таблица 30

Результаты измерений запыленности на рабочем месте и температуры воздуха

Измерение на рабочем месте	Температура воздуха С°	Запыленность мг/м ³
1. Слесарь	20	0,2
2. Электрик	21	0,25
3. Сварщик	21	0,24
4. ...	19	0,08
5. ...	19	0,08
6. ...	19	0,07
7. ...	22	0,3
8. ...	22	0,28
9. ...	25	0,33
10. ...	24	0,31
11. ...	21	0,26
12. ...	21	0,27

Задание: определите силу и направление зависимости между температурой окружающей среды и уровнем запыленности помещения с помощью:

- таблицы;
- графического изображения взаимосвязи между признаками;
- коэффициента корреляции Пирсона;
- коэффициента корреляции Спирмена.

Решение: необходимо запустить программу Excel, создать новый лист, на этом листе ввести данные в таблицу, создать диаграмму и выполнить расчеты, как показано ниже, сохранить файл.

а) Для определения связи между признаками требуется выполнить ранжирование вариационных рядов и поместить их рядом друг с другом, как показано в таблице 31, а затем проанализировать совместное возрастание или убывание значений признаков.

Таблица 31

№ варианты	Параметр x (температура воздуха, С°)	Параметр y (запыленность, мг/м ³)
1	19	0,07
2	19	0,08
3	19	0,08
4	20	0,2
5	21	0,24
6	21	0,25
7	21	0,26
8	21	0,27
9	22	0,3
10	22	0,28
11	24	0,31
12	25	0,33

Вывод: в таблице наблюдается совместное прямое изменение двух изучаемых параметров.

б) С помощью графика может быть выполнена визуальная оценка формы взаимосвязи между признаками. Для этого в программе Excel необходимо создать точечную диаграмму с прямыми или гладкими отрезками между маркерами, как показано на рис. 17.

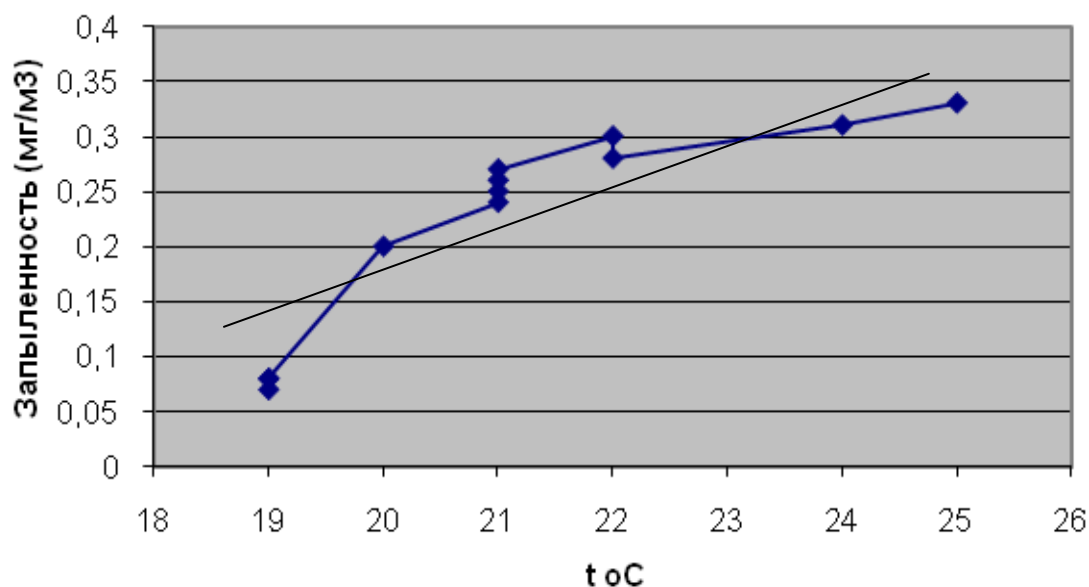


Рис. 17. Зависимость уровня запыленности от температуры воздуха в помещении.

Вывод: график зависимости совместного изменения двух изучаемых параметров показывает наличие взаимосвязи, которая приблизительно оценивается как линейная.

в) Вычисление и оценка коэффициента корреляции методом Пирсона (таблица 32).

Таблица 32

Вычисление отклонений вариант от средней арифметической

Варианта №	Температура воздуха (x)	Запыленность мг/м ³ (y)	$d_x = x - M_x$	$d_y = y - M_y$	$d_x \cdot d_y$	d_x^2	d_y^2
1	19	0,07	-2,2	-0,153	0,330	4,7	0,0233
2	19	0,08	-2,2	-0,143	0,309	4,7	0,0203
3	19	0,08	-2,2	-0,143	0,309	4,7	0,0203
4	20	0,2	-1,2	-0,023	0,026	1,4	0,0005
5	21	0,24	-0,2	0,018	-0,003	0,0	0,0003
6	21	0,25	-0,2	0,028	-0,005	0,0	0,0008
7	21	0,26	-0,2	0,038	-0,006	0,0	0,0014
8	21	0,27	-0,2	0,048	-0,008	0,0	0,0023
9	22	0,3	0,8	0,078	0,065	0,7	0,0060
10	22	0,28	0,8	0,058	0,048	0,7	0,0033
11	24	0,31	2,8	0,088	0,248	8,0	0,0077
12	25	0,33	3,8	0,108	0,412	14,7	0,0116
Средняя (M) = 21,2		0,223	Сумма (Σ) =		1,725	39,7	0,0976

$$n = 12$$

Коэффициент корреляции вычисляется по формуле:

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = \frac{1,725}{\sqrt{39,7 \times 0,0976}} = \mathbf{0,88}.$$

В программе Excel может использоваться функция =КОРРЕЛ(Диапазон1;Диапазон2) или модуль «Корреляция» из пакета анализа, который вызывается командой «Данные» - «Анализ данных». Он производит создание таблицы «Корреляционная матрица». В ней по диагонали указаны значения коэффициента корреляции между одинаковыми признаками, выражающиеся цифрой 1, и между разными признаками в нижней части матрицы. При использовании последнего способа расчета имеется возможность вычислить коэффициенты корреляции одновременно для нескольких признаков, если выделенный диапазон данных будут включены более двух столбцов. Результат вычислений, выполненный с помощью указанного модуля, приведен в таблице 33.

Таблица 33

Вычисление корреляционной матрицы модулем «Корреляция»

	Температура воздуха (x)	Запыленность мг/м ³ (y)
Температура воздуха (x)	1	
Запыленность мг/м ³ (y)	0,876588407	1

Оценка достоверности коэффициента корреляции выполняется с помощью ошибки репрезентативности корреляции (m_r) и критерия Стьюдента (t_r).

Ошибка репрезентативности корреляции вычисляется формулой:

$$m_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0,88^2}{12-2}} = \mathbf{0,152}, \text{ где: } n - \text{ число парных вариантов.}$$

Критерий достоверности Стьюдента для коэффициента корреляции рассчитывается по формуле:

$$t_r = \frac{r_{xy}}{m_r} = \frac{0,88}{0,152} = \mathbf{5,8 > 2}.$$

Сравнивая полученное значение критерия Стьюдента с критическим значением 2 можно оценить достоверность установленной корреляционной зависимости. При величине $t_r > 2$ она считается достоверной с уровнем значимости $p < 0,05$.

Вывод: зависимость изменения двух изучаемых параметров является сильной прямой и статистически достоверной при уровне значимости $p < 0,05$.

г) вычисление и оценка коэффициента корреляции методом Спирмена.

В таблице вариационных рядов производится подсчет рангов как показано в таблице 34.

Таблица 34

Вычисление рангов и суммы квадратов их отклонений

Варианта	Температура воздуха (x)	Запыленность мг/м ³ (y)	Ранг x	Ранг y	$d_r =$ ранг x - ранг y	d_r^2
1	19	0,07	1	1	0	0
2	19	0,08	1	2	-1	1
3	19	0,08	1	2	-1	1
4	20	0,2	4	4	0	0
5	21	0,24	5	5	0	0
6	21	0,25	5	6	-1	1
7	21	0,26	5	7	-2	4
8	21	0,27	5	8	-3	9
9	22	0,28	9	9	0	0
10	22	0,3	9	10	-1	1
11	24	0,31	11	11	0	0
12	25	0,33	12	12	0	0

$\Sigma=17$

Каждому из 12 чисел присваивается порядковый номер по возрастанию в соответствии с его значением. При этом наличие повторяющихся чисел влияет на ранг последующих чисел. Например, если в списке целых чисел трижды встречается число 19, имеющее ранг 1, число 20 будет иметь ранг 4 (ни одно из чисел не будет иметь ранги 2 и 3). Вычисление ранга в программе Excel возможно с помощью функции =РАНГ(Число; Диапазон; Порядок). Например: =РАНГ(C25;C\$24:C\$35;1). Затем вычисляется разность рангов $d_r = \text{ранг } x - \text{ранг } y$, она возводится в квадрат и суммируется.

Полученные значения используются в формуле вычисления коэффициента корреляции Спирмена:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 17}{12 \times (12^2 - 1)} = \mathbf{0,94}.$$

Вычисление ошибки репрезентативности коэффициента корреляции Спирмена:

$$m_\rho = \sqrt{\frac{1 - \rho^2}{n - 2}} = \sqrt{\frac{1 - 0,94^2}{12 - 2}} = 0,107.$$

Вычисление коэффициента достоверности Стьюдента для коэффициента корреляции:

$$t_\rho = \frac{\rho}{m_\rho} = \frac{0,94}{0,107} = \mathbf{8,76 > 2}.$$

Вывод: корреляционная связь двух изучаемых параметров является сильной прямой и статистически достоверной при уровне значимости $p < 0,05$.

ЗАДАНИЯ

В программе Excel определите силу и направление связи между двумя признаками с помощью таблицы, графического изображения, коэффициентов корреляции Пирсона и Спирмена.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «**Коррел-я**» и выполните задание варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки задания преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

Выполнены измерения признаков, характеризующих температуру в помещении на рабочих местах работников предприятия и концентрацию вредных веществ (таблица 35).

Таблица 35

Данные измерений на рабочих местах предприятия

Измерение на рабочем месте	Температура воздуха, С°	Концентрация вещества, мг/м ³
1. Слесарь	20	0,21
2. Электрик	21	0,26
3. Сварщик	21	0,25
4. ...	19	0,03
5. ...	19	0,04
6. ...	19	0,01
7. ...	22	0,31
8. ...	22	0,28
9. ...	25	0,36
10....	24	0,32
11....	21	0,21
12....	21	0,22

Определите силу и направление зависимости между температурой окружающей среды и концентрацией вредных веществ в помещении с помощью таблицы, графического изображения взаимосвязи между признаками, коэффициентов корреляции Пирсона и Спирмена, сделайте вывод.

Вариант 2

Выполнены измерения показателей физического развития школьников, характеризующих их рост стоя и объем грудной клетки (таблица 36).

Таблица 36

Данные физического развития школьников

Измерение	Рост, см	Объем грудной клетки, см
1. Чернов А.С.	151	70,8
2. Галкин М.В.	178	78,2
3. Попов А.М.	152	71,1
4. ...	160	73,2
5. ...	160	73,3
6. ...	178	78,2
7. ...	170	76,1
8. ...	170	76,3
9. ...	143	67,5
10....	170	76,1
11....	150	70,5
12....	172	76,6

Определите силу и направление зависимости между ростом и объемом грудной клетки с помощью таблицы, графического изображения взаимосвязи между признаками, коэффициентов корреляции Пирсона и Спирмена, сделайте вывод.

Вариант 3

Врачом футбольной команды выполнены измерения показателей деятельности системы кровообращения и тренированности спортсменов, измерены частота пульса и систолический объем сердечного выброса (таблица 37).

Таблица 37

Данные измерений показателей деятельности сердечно-сосудистой системы спортсменов

Измерение	Пульс, уд/мин	Объем сердечного выброса, мл
1. Васильев А.С.	78	58
2. Морозов Н.Р.	72	38
3. Родионов А.К.	78	63
4. ...	80	65
5. ...	72	35
6. ...	60	46
7. ...	72	59
8. ...	72	59
9. ...	72	50
10....	66	38
11....	72	40
12....	84	68

Определите силу и направление зависимости между пульсом и систолическим объемом с помощью таблицы, графического изображения взаимосвязи между признаками, коэффициентов корреляции Пирсона и Спирмена, сделайте вывод.

Вариант 4

В городе Н. было проведено изучение зависимости заболеваемости инфарктом миокарда по месяцам года от среднемесячной температуры воздуха (таблица 38).

Таблица 40

Заболеваемость инфарктом миокарда и температура воздуха по месяцам

Месяцы года	Заболеваемость инфарктом миокарда (на 10000 жителей)	Среднемесячная температура воздуха (С ⁰)
Январь	1,60	-7,1
Февраль	1,23	-7,7
Март	1,14	-5,8
Апрель	1,13	-4,1
Май	1,12	13,0
Июнь	1,02	14,9
Июль	0,91	18,8
Август	0,82	15,6
Сентябрь	1,06	9,0
Октябрь	1,22	6,0
Ноябрь	1,33	-1,0
Декабрь	1,40	-7,7

Определите силу и направление зависимости между заболеваемостью инфарктом миокарда и среднемесячной температурой воздуха с помощью таблицы, графического изображения взаимосвязи между признаками, коэффициентов корреляции Пирсона и Спирмена, сделайте вывод.

Вариант 5

Выполнены измерения показателей качества питьевой воды в городе Н, в том числе ее жесткость и концентрация кальция (таблица 37).

Таблица 37

Данные измерений показателей питьевой воды

№ измерения	Жесткость воды, градусы жесткости	Концентрация кальция в воде, мг/л
1.	4	28
2.	11	77
3.	27	191
4.	34	241
5.	8	56
6.	37	262

Определите силу и направление зависимости между жесткостью воды и количеством кальция в ней с помощью таблицы, графического изображения взаимосвязи между признаками, коэффициентов корреляции Пирсона и Спирмена, сделайте вывод.

Вариант 6

Выполнены измерения показателей физического развития студенток, в том числе их рост и объем грудной клетки (таблица 36).

Таблица 36

Данные физического развития студенток

Измерение	Рост, см	Объем грудной клетки, см
1. Галкина А.С.	159	71,8
2. Синицына М.В.	176	79,6
3. Попова В.С.	158	71,5
4. ...	165	73,2
5. ...	168	73,4
6. ...	179	78,2
7. ...	182	79,1
8. ...	176	76,3
9. ...	155	69,5
10....	171	76,1
11....	156	69,5
12....	172	76,6

Определите силу и направление зависимости между ростом и объемом грудной клетки студенток с помощью таблицы, графического изображения взаимосвязи между признаками, коэффициентов корреляции Пирсона и Спирмена, сделайте вывод.

Вопросы для самоконтроля

1. Виды связей между признаками, примеры.

2. Функциональная зависимость (определение).
3. Какие явления могут быть смоделированы функциональной зависимостью?
4. Корреляционная зависимость (определение).
5. Способы обнаружения корреляционной связи.
6. Оценка силы и направления связи по величине коэффициента корреляции.
7. Методика расчета коэффициента линейной корреляции.
8. Как оценивается достоверность коэффициента корреляции?
9. Методика расчета рангового коэффициента корреляции.

Рекомендуемая литература:

1. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>

ХII. Метод регрессии

МЕТОД РЕГРЕССИИ - это статистический способ поиска функции, которая позволяет по величине одного коррелируемого признака судить о величине другого. С помощью регрессии ставится задача выяснить, как количественно меняется одна величина при изменении другой величины на единицу. Для выполнения такого прогноза требуется определить коэффициент корреляции Пирсона, с использованием которого вычисляют коэффициент регрессии ($R_{y/x}$). Он участвует в создании регрессионной функции вида $y=ax+b$, которая применяется для прогнозирования требуемых параметров.

Коэффициент регрессии вычисляется по формуле:

$$R_{y/x} = r_{x/y} \frac{\sigma_y}{\sigma_x},$$

где: $R_{y/x}$ – коэффициент регрессии;

$r_{x/y}$ – коэффициент корреляции Пирсона;

σ_x – среднее квадратическое отклонение признака x ;

σ_y – среднее квадратическое отклонение признака y .

Среднее квадратическое отклонение (сигма) вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}},$$

а в программе Excel функцией =СТАНДОТКЛОН(Диапазон ячеек).

Значение коэффициента регрессии ($R_{y/x}$) в программе Excel может быть вычислено функцией =НАКЛОН(Диапазон_y; Диапазон_x).

Формула определения значения зависимого признака:

$$y = M_y + R_{y/x} (x - M_x),$$

где: y – зависимая переменная;

M_y – средняя признака y ;

$R_{y/x}$ – коэффициент регрессии;

x – значение измеренного признака;

M_x – средняя арифметическая признака x .

В программе Excel значение зависимой переменной (y) при заданном значении x может быть вычислено функцией =ПРЕДСКАЗ(x ; Диапазон_y; Диапазон_x).

После получения прогнозируемого значения (y) выполняется определение его доверительного интервала с целью экстраполяции данных на генеральную совокупность с уровнем значимости $p < 0,05$. Для этого вычисляется сигма регрессии $\sigma_{Ry/x}$, которая показывает меру вариабельности зависимого признака, вычисленного по уравнению регрессии, в генеральной совокупности.

Она определяется по формуле:

$$\sigma_{Ry/x} = \sigma_y \sqrt{1 - r_{xy}^2}.$$

Вычисление значения σ_y может производиться функцией = СТАНДОТКЛОН(Диапазон_y).

Пример прогнозирования значения одного признака по известному значению другого с помощью уравнения регрессии

Условие задачи: на основе данных, характеризующих уровень запыленности рабочих мест (см. раздел VIII), необходимо выполнить прогноз уровня пыли при температуре воздуха 23С⁰.

Задание: получить уравнение регрессии для зависимости между температурой окружающей среды и уровнем запыленности помещения, построить регрессионную функцию и вычислить значение уровня пыли при температуре воздуха 23С⁰. Определить сигму регрессии и доверительный интервал для прогнозируемого значения уровня пыли.

Решение: необходимо запустить программу Excel, создать новый лист, на этом листе ввести данные и выполнить требуемые этапы вычислений, как показано на рис.18, сохранить файл.

	A	B	C	D	E	F	G	H	I	J	K	L
1	1) Вычисление коэффициента корреляции Пирсона											
2	Варианта	Температура воздуха (x)	Запыленность мг/м3 (y)	d _x	d _y	d _x *d _y	d _x ²	d _y ²				
3	1	19	0,07	2,2	0,153	0,330	4,7	0,0233				
4	2	19	0,08	2,2	0,143	0,309	4,7	0,0203				
5	3	19	0,08	2,2	0,143	0,309	4,7	0,0203				
6	4	20	0,2	1,2	0,023	0,026	1,4	0,0005				
7	5	21	0,24	0,2	-0,018	-0,003	0,0	0,0003				
8	6	21	0,25	0,2	-0,028	-0,005	0,0	0,0008				
9	7	21	0,26	0,2	-0,038	-0,006	0,0	0,0014				
10	8	21	0,27	0,2	-0,048	-0,008	0,0	0,0023				
11	9	22	0,3	-0,8	-0,078	0,065	0,7	0,0060				
12	10	22	0,28	-0,8	-0,058	0,048	0,7	0,0033				
13	11	24	0,31	-2,8	-0,088	0,248	8,0	0,0077				
14	12	25	0,33	-3,8	-0,108	0,412	14,7	0,0116				
15	Средняя (M) =	21,2	0,223	Сумма (Σ) =		1,725	39,7	0,0976				
16	n=	12										
17												
18	$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$		$r_{xy} =$	0,88	=КОРРЕЛ(В3:В14;С3:С14)=			0,876588				
19	2) Вычисление коэффициента регрессии:											
20	$\sigma = \sqrt{\frac{\sum d^2}{n-1}}$		$\delta_x =$	1,90	=КОРЕНЬ(39,7/(12-1))=			1,9	=СТАНДОТКЛОН(В3:В14)=	1,9		
21			$\delta_y =$	0,09	=КОРЕНЬ(0,0976/(12-1))=			0,094	=СТАНДОТКЛОН(С3:С14)=	0,094		
22	$R_{x/y} = r_{xy} * \delta_y / \delta_x$		$R_{x/y} =$	0,043	=0,88*0,094/1,9=			0,044	=D18*D21/D20=	0,043		
23	3) Вычисление значения зависимого признака:											
24	$y = M_y + R_{x/y}(x - M_x)$		$x =$	23								
25			$y =$	0,302	=0,223+0,04(23-21,2)=				=ПРЕДСКАЗ(D24 ; y1:yn ; x1:xn)	0,302227		
26	4) Вычисление доверительных границ зависимого признака:											
27	$\delta_{R_{x/y}} = \delta_y * \sqrt{1 - r_{xy}^2}$		$\delta_{R_{x/y}} =$	0,045	=0,094*КОРЕНЬ(1-0,88^2)=			0,045	=D21*КОРЕНЬ(1-D18^2)=	0,045		
28	Доверительные границы (2б):		0,212	0,393								
29												
30	Вывод: при температуре воздуха 23С ⁰ запыленность составит 0,302±0,045											

Рис. 18. Регрессионный анализ в программе Excel.

Вычисления выполняются в четыре этапа:

1) уравнение регрессии может быть получено после вычисления коэффициента корреляции Пирсона с помощью таблицы отклонений (таблица 39) или функцией =КОРРЕЛ(Диапазон1;Диапазон2).

Таблица 39

Вычисление коэффициента корреляции Пирсона

Варианта	Температура воздуха (x)	Запыленность мг/м ³ (y)	$d_x = x - M_x$	$d_y = y - M_y$	$d_x * d_y$	d_x^2	d_y^2
1	19	0,07	2,2	0,153	0,330	4,7	0,0233
2	19	0,08	2,2	0,143	0,309	4,7	0,0203
3	19	0,08	2,2	0,143	0,309	4,7	0,0203
4	20	0,2	1,2	0,023	0,026	1,4	0,0005
5	21	0,24	0,2	-0,018	-0,003	0,0	0,0003
6	21	0,25	0,2	-0,028	-0,005	0,0	0,0008
7	21	0,26	0,2	-0,038	-0,006	0,0	0,0014
8	21	0,27	0,2	-0,048	-0,008	0,0	0,0023
9	22	0,3	-0,8	-0,078	0,065	0,7	0,0060
10	22	0,28	-0,8	-0,058	0,048	0,7	0,0033
11	24	0,31	-2,8	-0,088	0,248	8,0	0,0077
12	25	0,33	-3,8	-0,108	0,412	14,7	0,0116

Средняя (M) = 21,2 0,223 Сумма (Σ) = 1,725 39,7 0,0976
 $n = 12$

$$r_{xy} = \frac{\Sigma d_x d_y}{\sqrt{\Sigma d_x^2 \Sigma d_y^2}} = \frac{1,725}{\sqrt{39,7 \times 0,0976}} = \text{КОРРЕЛ}(x_1:x_n; y_1:y_n) = \mathbf{0,88}.$$

2) вычисление коэффициента регрессии ($R_{y/x}$) производится с использованием среднего квадратического отклонения (σ) и вычисленного ранее коэффициента корреляции ($r_{x/y}$):

$$\sigma = \sqrt{\frac{\Sigma d^2}{n-1}}.$$

$$\sigma_x = \text{КОРЕНЬ}(39,7/(12-1)) = \text{СТАНДОТКЛОН}(x_1:x_n) = 1,9.$$

$$\sigma_y = \text{КОРЕНЬ}(0,0976/12) = \text{СТАНДОТКЛОН}(y_1:y_n) = 0,094.$$

Подставив значения в формулу вычисления коэффициента регрессии:

$$R_{y/x} = r_{x/y} \frac{\sigma_y}{\sigma_x},$$

получим:

$$R_{y/x} = 0,88 * 0,094 / 1,9 = \mathbf{0,04}.$$

3) вычисление величины зависимого признака (y) при температуре 23C⁰:

$$y = M_y + R_{y/x}(x - M_x) \text{ При } x = 23C^0$$

$$y = 0,223 + 0,04(23 - 21,2) = \text{ПРЕДСКАЗ}(x; y_1:y_n; x_1:x_n) = \mathbf{0,30} \text{ мг/м}^3$$

4) вычисление доверительных границ колебаний зависимого признака в генеральной совокупности ($\sigma_{Ry/x}$):

$$\sigma_{Ry/x} = \sigma_y \sqrt{1 - r_{xy}^2} \quad \sigma_{Ry/x} = 0,094 * \text{КОРЕНЬ}(1 - 0,88^2) = \mathbf{0,045}.$$

Доверительные границы (2σ): **от 0,21 до 0,39** При $p < 0,05$

Вывод: при температуре воздуха 23С° запыленность составит от 0,21 до 0,39мг/м3.

ЗАДАНИЯ

С помощью программы Excel постройте уравнение регрессии для двух признаков, выполните вычисление значения одного из признаков по величине другого и доверительный интервал для полученного значения.

ВНИМАНИЕ! Скопируйте табличный документ с именем «Статистика–Фамилии студентов» из общей папки группы на рабочий стол своего компьютера. Откройте скопированный файл. Создайте новый лист, переименуйте его, указав имя «Регрессия» и выполните задание варианта, указанного преподавателем. Затем сохраните файл, закройте программу и скопируйте файл в общую папку своей группы. Сообщите преподавателю о выполнении задания. После проверки задания преподавателем удалите файл с рабочего стола своего компьютера.

Вариант 1

Выполнены измерения признаков, характеризующих температуру в помещении на рабочих местах работников предприятия и концентрацию вредных веществ (см. вариант 1 заданий раздела X). Постройте уравнение регрессии для зависимости между температурой окружающей среды и концентрацией вещества в помещении. Определите значение уровня пыли при температуре воздуха 23 С°, вычислите сигму регрессии и доверительный интервал для полученного значения уровня пыли.

Вариант 2

Выполнены измерения показателей физического развития школьников, характеризующих их рост стоя и объем грудной клетки (см. вариант 2 заданий раздела IX). Постройте уравнение регрессии для зависимости между ростом и объем грудной клетки. Определите значение объема грудной клетки при росте 175 см. Вычислите сигму регрессии и доверительный интервал для полученного значения роста.

Вариант 3

Выполнены измерения показателей деятельности сердечно-сосудистой системы и тренированности спортсменов, среди них частота пульса и

систолический объем сердечного выброса (см. вариант 3 заданий раздела X). Постройте уравнение регрессии для зависимости между пульсом спортсменов и систолическим объемом. Определите значение объема сердечного выброса при пульсе 75 уд/мин. Вычислите сигму регрессии и доверительный интервал для полученного значения пульса.

Вариант 4

В городе Н. было проведено изучение зависимости заболеваемости инфарктом миокарда по месяцам года в зависимости от среднемесячной температуры воздуха (см. вариант 4 заданий раздела X). Постройте уравнение регрессии для зависимости между среднемесячной температурой воздуха и уровнем заболеваемости инфарктом миокарда. Определите значение уровня заболеваемости инфарктом миокарда при температуре воздуха $+10\text{ C}^0$. Вычислите сигму регрессии и доверительный интервал для полученного значения показателя заболеваемости.

Вариант 5

Выполнены измерения показателей качества питьевой воды в городе Н (см. вариант 5 заданий раздела X). Постройте уравнение регрессии для зависимости между жесткостью воды и количеством кальция в ней. Определите концентрацию кальция в воде с жесткостью 14°Ж . Вычислите сигму регрессии и доверительный интервал для полученного значения концентрации кальция.

Вариант 6

Выполнены измерения показателей физического развития студенток, характеризующих их рост и объем грудной клетки (см. вариант 6 заданий раздела X). Постройте уравнение регрессии для зависимости между ростом и объемом грудной клетки студенток. Определите значение объема грудной клетки при росте студенток 160 см. Вычислите сигму регрессии и доверительный интервал для полученного значения объема грудной клетки.

Вопросы для самоконтроля

1. Дайте определение термина «Регрессия».
2. Значение коэффициента регрессии и методика его расчета.
3. Создание уравнения регрессии.
4. Вычисление и оценка сигмы регрессии.

Рекомендуемая литература:

1. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.

2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения [Электронный ресурс]: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. - 256 с. – Режим доступа: <http://www.studmedlib.ru>

Тестовые задания

1. ГРУППА ПРОГРАММ, ПРЕДНАЗНАЧЕННЫХ ДЛЯ СОЗДАНИЯ И РАБОТЫ С ТАБЛИЦАМИ, НАЗЫВАЕТСЯ ...

- 1) Электронные таблицы
- 2) Браузеры
- 3) Текстовые редакторы
- 4) Графические редакторы

Решение: группа программ, предназначенных для создания и работы с таблицами, называется «Электронные таблицы».

Правильный ответ: 1.

2. ОСНОВНЫМ НАЗНАЧЕНИЕМ ЭЛЕКТРОННЫХ ТАБЛИЦ (ТАБЛИЧНЫХ РЕДАКТОРОВ) ЯВЛЯЕТСЯ...

- 1) работа с упорядоченными числовыми данными, выполнение относительно несложных расчетов и создание диаграмм
- 2) это единственное средство проведения сложных расчетов и статистического анализа результатов научных исследований
- 3) создание деловой документации
- 4) просмотр страниц Интернет

Решение: основным назначением электронных таблиц (табличных процессоров) является работа с упорядоченными числовыми данными, выполнение относительно несложных расчетов и подготовка диаграмм.

Правильный ответ: 1.

3. ПРОГРАММА MICROSOFT EXCEL – ЭТО ...

- 1) табличный редактор
- 2) текстовый редактор
- 3) графический редактор
- 4) браузер

Решение: программа Microsoft Excel является табличным редактором.

Правильный ответ: 1.

4. ПАКЕТ ПРИКЛАДНЫХ ПРОГРАММ, В СОСТАВ КОТОРОГО ВХОДИТ ТАБЛИЧНЫЙ РЕДАКТОР EXCEL, НАЗЫВАЕТСЯ...

- 1) Open Office
- 2) Microsoft Office
- 3) Star Office
- 4) On Note

Решение: пакет прикладных программ, в состав которого входит табличный редактор Excel, называется Microsoft Office.

Правильный ответ: 2.

5. ПЕРЕНОС ТАБЛИЦ И РИСУНКОВ ДИАГРАММ ИЗ ПРОГРАММЫ EXCEL В ДЕЛОВЫЕ ДОКУМЕНТЫ, СОЗДАВАЕМЫЕ ПРОГРАММОЙ WORD, ...

- 1) возможен, но с потерей данных
- 2) не возможен
- 3) возможен без ограничений
- 4) возможен, но с ограничением размера таблицы

Решение: перенос таблиц и рисунков диаграмм из программы Excel в деловые документы, создаваемые программой Word, возможен без ограничений.

Правильный ответ: 3.

6. ИМЕНЕМ НОВОГО ДОКУМЕНТА, КОТОРОЕ ПО УМОЛЧАНИЮ ПРИСВАИВАЕТСЯ ТАБЛИЦЕ ПРИ ЗАПУСКЕ ПРОГРАММЫ EXCEL И ОТОБРАЖАЕТСЯ В ЗАГОЛОВКЕ ОКНА, ЯВЛЯЕТСЯ...

- 1) «Документ1»
- 2) «Книга1»
- 3) «Лист1»
- 4) «Страница1»

Решение: документ, создаваемый программой Excel, называется «Книгой», поэтому именем нового документа, которое по умолчанию присваивается табличному файлу при запуске программы и отображается в заголовке окна, является «Книга1».

Правильный ответ: 2.

7. ФАЙЛЫ ТАБЛИЧНЫХ ДОКУМЕНТОВ ПРОГРАММЫ MICROSOFT EXCEL ОБОЗНАЧАЮТСЯ РАСШИРЕНИЕМ ...

- 1) xls илиxlsx
- 2) zip или arj
- 3) htm или html
- 4) doc или docx

Решение: именем табличные документы, создаваемые в программе Excel, обозначаются расширением «xls» (для версий до 2007) или «xlsx» (для версий 2007 и 2010).

Правильный ответ: 1.

8. УСТАНОВКА ЗАЩИТЫ ЧАСТИ РАБОЧЕЙ КНИГИ В ПРОГРАММЕ EXCEL ВЫПОЛНЯЕТСЯ В РЕЖИМЕ ...

- 1) Главная
- 2) Вставка
- 3) Рецензирование
- 4) Вид

Решение: в программе Excel установка защиты части рабочей книги выполняется в режиме «Рецензирование».

Правильный ответ: 3.

9. В ПРОГРАММЕ EXCEL ЗАЩИТА ФАЙЛА ТАБЛИЧНОГО ДОКУМЕНТА С НАЗНАЧЕНИЕМ ПАРОЛЯ ДЛЯ ЕГО ОТКРЫТИЯ ВЫПОЛНЯЕТСЯ В РЕЖИМЕ ...

- 1) Подготовить
- 2) Вставка
- 3) Рецензирование
- 4) Вид

Решение: в программе Excel установка защиты файла табличного документа с назначением пароля на его открытие выполняется в режиме «Подготовить» командой «Зашифровать документ».

Правильный ответ: 1.

10. ЭЛЕМЕНТ ПРОГРАММЫ EXCEL, НАЗЫВАЕМЫЙ «РАБОЧИЙ ЛИСТ» – ЭТО...

- 1) сетка ячеек, предназначенных для ввода и отображения информации
- 2) место размещения кнопок для выполнения команд
- 3) строка ввода математических выражений для выполнения вычислений
- 4) диалоговое окно выполнения команды

Решение: элемент программы Excel, называемый рабочим листом – это сетка ячеек, предназначенных для ввода и отображения информации.

Правильный ответ: 1.

11. ИМЯ РАБОЧЕГО ЛИСТА В ПРОГРАММЕ EXCEL ОТОБРАЖАЕТСЯ ...

- 1) на панели инструментов
- 2) на корешке листа
- 3) в панели задач
- 4) в заголовке окна

Решение: в программе Excel имя рабочего листа отображается на ярлычке листа в нижней части окна.

Правильный ответ: 2.

12. СОВОКУПНОСТЬ ЛИСТОВ EXCEL СОСТАВЛЯЕТ:

- 1) рабочую книгу
- 2) рабочую область
- 3) рабочую страницу
- 4) рабочую ячейку

Решение: в программе Excel имя рабочего листа отображается на ярлычке в нижней части окна.

Правильный ответ: 1.

13. ПЕРЕКЛЮЧЕНИЕ К СЛЕДУЮЩЕМУ ЛИСТУ РАБОЧЕЙ КНИГИ EXCEL ПРОИЗВОДИТСЯ...

- 1) щелчком левой клавиши мыши в заголовке окна программы

- 2) протаскиванием движка в горизонтальной полосе прокрутки
- 3) щелчком левой клавиши мыши на ярлычке листа
- 4) щелчком правой клавиши мыши на ярлычке листа

Решение: переключение к следующему листу рабочей книги Excel производится щелчком левой клавиши мыши на ярлычке листа.

Правильный ответ: 3.

14. В ПРОГРАММЕ EXCEL ДЛЯ ОБОЗНАЧЕНИЯ СТОЛБЦОВ ТАБЛИЦЫ, КАК ПРАВИЛО, ПРИМЕНЯЮТСЯ ...

- 1) римские цифры
- 2) английские буквы и их сочетание
- 3) специальные символы
- 4) русские символы и их сочетание

Решение: в программе Excel для обозначения столбцов таблицы, как правило, применяются английские буквы и их сочетание.

Правильный ответ: 2.

15. В ПРОГРАММЕ EXCEL ДЛЯ ОБОЗНАЧЕНИЯ СТРОК ТАБЛИЦЫ ИСПОЛЬЗУЮТСЯ ...

- 1) арабские цифры
- 2) английские символы и их сочетание
- 3) римские цифры
- 4) русские символы и их сочетание

Решение: в программе Excel для обозначения строк таблицы используются арабские цифры.

Правильный ответ: 1.

16. МАКСИМАЛЬНО ВОЗМОЖНЫЙ РАЗМЕР ТАБЛИЦЫ НА ОДНОМ ЛИСТЕ ДОКУМЕНТА EXCEL 2003 СОСТАВЛЯЕТ...

- 1) 10 столбцов и 256 строк
- 2) 256 столбцов и 65,5 тысяч строк
- 3) 65,5 тысяч столбцов и 256 тысяч строк
- 4) 65,5 тысяч столбцов и 65,5 тысяч строк

Решение: в программе Excel 2003 максимально возможный размер таблицы на одном листе составляет 256 столбцов и 65,5 тысяч строк.

Правильный ответ: 2.

17. ЯЧЕЙКА ТАБЛИЦЫ В ПРОГРАММЕ EXCEL – ЭТО ...

- 1) кнопка панели инструментов программы
- 2) место пересечения столбцов и строк в рабочей области экрана
- 3) пункт меню
- 4) лист в книге

Решение: в программе Excel ячейка таблицы – это место пересечения столбцов и строк в рабочей области экрана.

Правильный ответ: 2.

18. АКТИВНАЯ ЯЧЕЙКА ТАБЛИЦЫ В ПРОГРАММЕ EXCEL ОБОЗНАЧАЕТСЯ ...

- 1) табличным курсором – черным прямоугольником на ячейке
- 2) изменением размера ячейки
- 3) цветом текста в ячейке
- 4) диапазоном адресов

Решение: активная ячейка таблицы в программе Excel обозначается табличным курсором – черным прямоугольником на ячейке.

Правильный ответ: 1.

19. СПОСОБАМИ ПЕРЕМЕЩЕНИЯ ТАБЛИЧНОГО КУРСОРА В ПРОГРАММЕ EXCEL ЯВЛЯЮТСЯ:

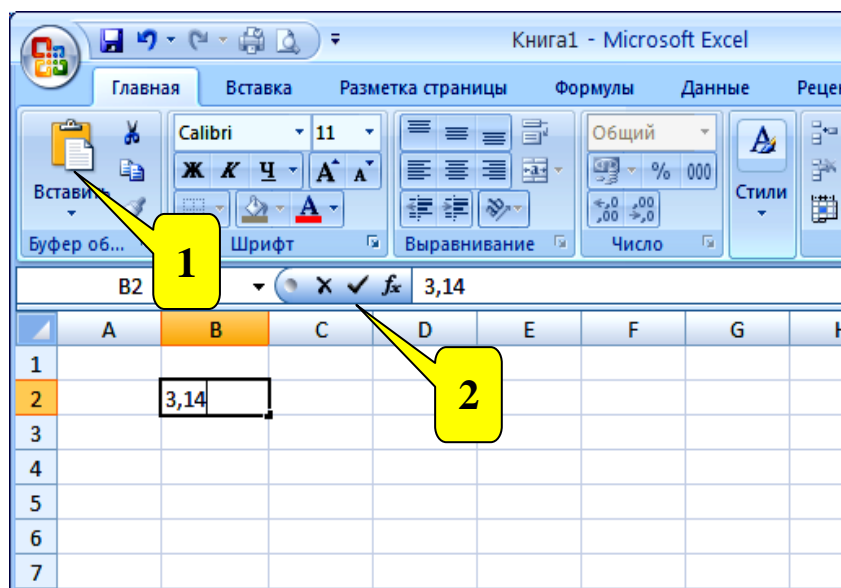
- 1) клавиши управления курсором (стрелки) клавиатуры
- 2) щелчок левой клавиши мыши на требуемой ячейке
- 3) применение клавиши Shift клавиатуры
- 4) протаскивание курсора мыши

Решение: способами перемещения табличного курсора в программе Excel являются клавиши управления курсором (стрелки) клавиатуры и щелчок левой клавиши мыши на ячейке.

Правильный ответ: 1, 2.

20. КОМАНДАМИ ЗАВЕРШЕНИЯ ВВОДА ДАННЫХ В ЯЧЕЙКУ B2 ТАБЛИЦЫ EXCEL (СМ. РИСУНОК) ЯВЛЯЮТСЯ:

- 1) клавиша Enter клавиатуры
- 2) комбинация клавиш клавиатуры «Alt + Shift»
- 3) кнопка 1 (см.рисунок)
- 4) кнопка 2 (см.рисунок)
- 5) клавиша F1 клавиатуры



Решение: командами завершения ввода данных в ячейку B2 таблицы Excel (см.рисунок) являются клавиша Enter клавиатуры и кнопка 2.

Правильный ответ: 1, 4.

21. АДРЕС ЯЧЕЙКИ ТАБЛИЦЫ В ПРОГРАММЕ EXCEL СОСТОИТ ИЗ ...

- 1) имени файла и даты его создания
- 2) имени владельца документа
- 3) буквы, обозначающей столбец, и цифры, обозначающей строку
- 4) формулы вычисления

Решение: адрес ячейки таблицы в программе Excel состоит из буквы, обозначающей столбец, и цифры, обозначающей строку.

Правильный ответ: 3.

22. В ПРОГРАММЕ EXCEL ДРОБНАЯ ЧАСТЬ ЧИСЛА (ДЕСЯТИЧНЫЕ ЗНАКИ) ОТДЕЛЯЕТСЯ ОТ ЦЕЛОЙ ЧАСТИ ...

- 1) пробелом
- 2) двоеточием
- 3) запятой
- 4) точкой

Решение: в программе Excel дробная часть числа (десятичные знаки) отделяется от целой части запятой.

Правильный ответ: 3.

23. В ТАБЛИЦЕ EXCEL НЕКОТОРОЕ КОЛИЧЕСТВО ЯЧЕЕК, ОБОЗНАЧЕННЫХ ПРЯМОУГОЛЬНОЙ ОБЛАСТЬЮ ВЫДЕЛЕНИЯ, НАЗЫВАЮТ ...

- 1) диапазоном ячеек
- 2) группой ячеек
- 3) множеством ячеек
- 4) рабочей ячейкой

Решение: в программе Excel некоторое количество ячеек, обозначенных прямоугольной областью выделения, называется диапазоном ячеек и обозначается начальным и конечным адресами, разделенными двоеточием.

Правильный ответ: 1.

24. ЕСЛИ ЯЧЕЙКА ТАБЛИЦЫ ПЕРЕТАСКИВАЕТСЯ МЫШЬЮ В ЯЧЕЙКУ, СОДЕРЖАЩУЮ ДАННЫЕ, ТО ПРОИСХОДИТ ...

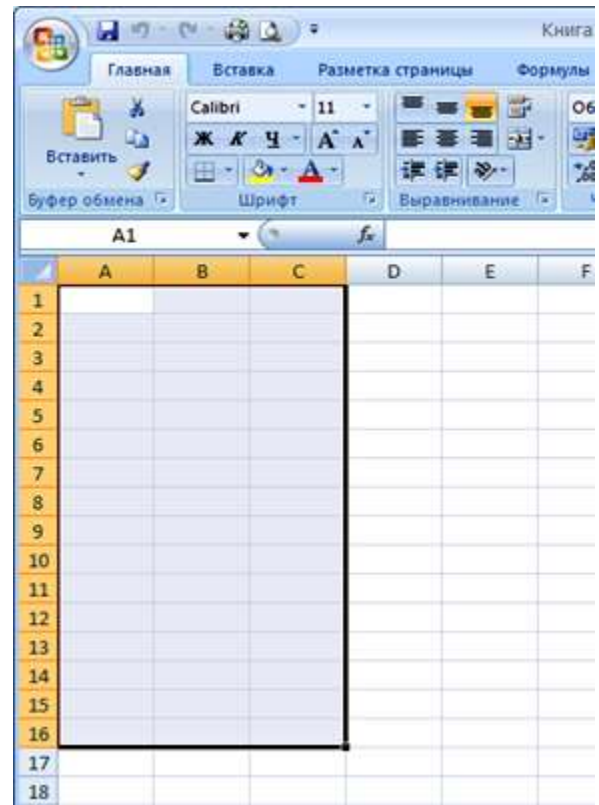
- 1) зависание программы
- 2) появление предупреждения о замене данных
- 3) замена старых данных на новые без предупреждения
- 4) вставляемая информация добавится к существующей

Решение: если ячейка таблицы перетаскивается мышью в ячейку, содержащую данные, то происходит появление предупреждения о замене данных.

Правильный ответ: 2.

25. В ПРОГРАММЕ EXCEL ВЫДЕЛЕНИЕ ДИАПАЗОНА АДРЕСОВ ТАБЛИЦЫ A1:C16 (СМ. РИСУНОК) ВЫПОЛНЯЕТСЯ СЛЕДУЮЩИМИ ДЕЙСТВИЯМИ:

- 1) указать курсором мыши на границу ячейки A1, нажать на левую клавишу мыши и, не отпуская ее, довести указатель мыши до ячейки C16
- 2) указать курсором мыши в центр ячейки A1, нажать на левую клавишу мыши и, не отпуская ее, довести указатель мыши до ячейки C16
- 3) щелкнуть левой клавишей мыши по ячейке A1, затем при нажатой клавише клавиатуры Ctrl щелкнуть левой клавишей мыши по ячейке C16
- 4) щелкнуть левой клавишей мыши по ячейке A1, затем при нажатой клавише клавиатуры Shift щелкнуть левой клавишей мыши по ячейке C16



Решение: в программе Excel выделение диапазона адресов таблицы A1:C16 (см. рисунок) выполняется следующими действиями: указать курсором мыши в центр ячейки A1, нажать на левую клавишу мыши и, не отпуская ее, довести указатель мыши до ячейки C16; или щелкнуть левой клавишей мыши по ячейке A1, затем при нажатой клавише клавиатуры Shift щелкнуть левой клавишей мыши по ячейке C16.

Правильный ответ: 2, 4.

26. ВЫДЕЛЕН ДИАПАЗОН ЯЧЕЕК A1:D3 ЭЛЕКТРОННОЙ ТАБЛИЦЫ EXCEL, ОН ВКЛЮЧАЕТ...

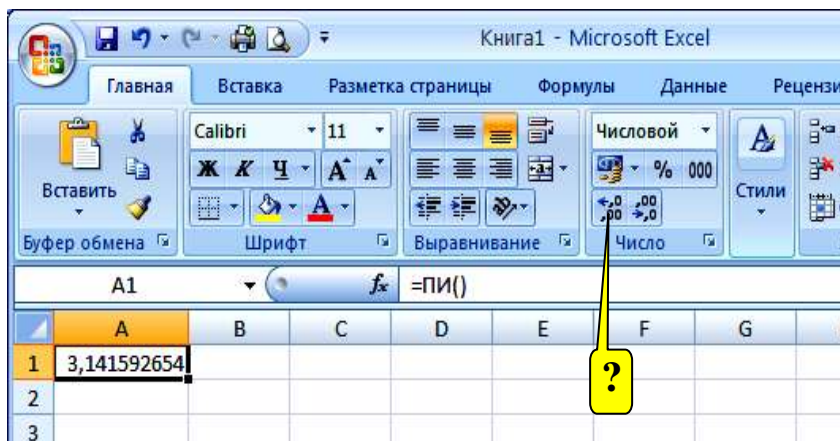
- 1) 2 ячейки
- 2) 6 ячеек
- 3) 9 ячеек
- 4) 12 ячеек

Решение: в программе Excel диапазон ячеек обозначается начальным и конечным адресом, разделенным знаком двоеточия, поэтому обозначенный диапазон A1:D3 включает 4 столбца и 3 строки, то есть 12 ячеек.

Правильный ответ: 4.

27. УКАЗАННАЯ КНОПКА ПАНЕЛИ ИНСТРУМЕНТОВ EXCEL (СМ. РИСУНОК) ИСПОЛЬЗУЕТСЯ ДЛЯ ...

- 1) округления числа до одного знака после запятой
- 2) увеличения числа знаков, отображаемых после запятой
- 3) деления числа на 10
- 4) умножения числа на 10



Решение: указанная кнопка панели инструментов Excel (см.рисунок) используется для увеличения числа знаков, отображаемых после запятой.

Правильный ответ: 2.

28. В ПРОГРАММЕ EXCEL ПРИЧИНОЙ ПОЯВЛЕНИЯ В ЯЧЕЙКЕ ЗНАКОВ РЕШЕТКИ (####) ЯВЛЯЕТСЯ ...

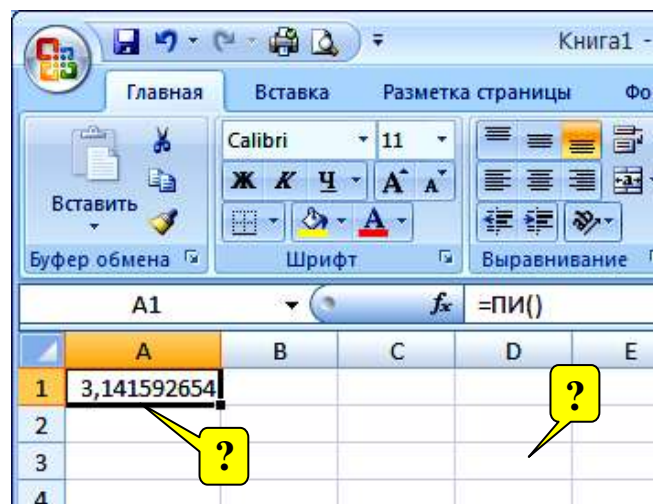
- 1) использование недопустимого типа данных, расположенных в ячейке, адрес которой указан в формуле
- 2) числовое значение не умещается по ширине ячейки
- 3) в формуле указан несуществующий или неправильный адрес ячейки
- 4) формула содержит вычисления включающие деление на ноль

Решение: если в программе Excel при вводе числа или формулы в ячейке отображаются знаки решетки (####), это указывает на недостаточную ширину ячейки для отображения числа. Для исправления ошибки требуется увеличить ширину столбца, уменьшить размер шрифта или его разрядность.

Правильный ответ: 2.

29. СПОСОБАМИ ПЕРЕМЕЩЕНИЯ ДАННЫХ ИЗ ЯЧЕЙКИ A1 В ЯЧЕЙКУ D3 (СМ. РИСУНОК) ЯВЛЯЮТСЯ:

- а) протаскивание контура ячейки A1 левой клавишей мыши в ячейку D3
- 2) протаскивание контура ячейки A1 левой клавишей мыши при нажатой клавише клавиатуры Ctrl в ячейку D3
- 3) протаскивание левой клавишей мыши маркера в правом нижнем углу ячейки A1 до ячейки D3



4) выполнением команд «Вырезать» и «Вставить»

Решение: способами перемещения данных из ячейки A1 в ячейку D3 (см.рисунок) являются: протаскивание контура ячейки A1 левой клавишей мыши в ячейку D3; выполнением команд «Вырезать» и «Вставить».

Правильный ответ: 1, 4.

30. СТРОКА ФОРМУЛ В MICROSOFT EXCEL ...

- 1) отображает наименование открытого листа
- 2) воспроизводит имя обрабатываемого файла
- 3) показывает адрес ячейки, обозначенной курсором и данные в ней
- 4) содержит служебную информацию о таблице

Решение: в программе Excel строка формул размещается в верхней части окна и показывает адрес активной ячейки, обозначенной курсором, а также данные и формулы в ней.

Правильный ответ: 3.

31. СТРОКА ФОРМУЛ В ОКНЕ ПРОГРАММЫ EXCEL, КАК ПРАВИЛО, РАЗМЕЩАЕТСЯ ...

- 1) в верхней части, ниже стандартной панели инструментов
- 2) справа, рядом с полосой прокрутки
- 3) внизу, над строкой состояния
- 4) слева, рядом с границей окна
- 5) в произвольном положении как отдельное окно

Решение: строка формул, как правило, расположена в верхней части окна ниже панели инструментов. Она предназначена для ввода и редактирования в ячейках таблицы текстовых данных, чисел и формул. Данные, внесенные в активную ячейку, дополнительно отображаются в этой строке.

Правильный ответ: 1.

32. СТРОКА ФОРМУЛ ПРОГРАММЫ EXCEL СЛУЖИТ ДЛЯ ...

- 1) ввода команд, управления программой
- 2) ввода и редактирования текстовых данных, чисел и формул в ячейках таблицы
- 3) сохранения таблицы в файле электронного документа
- 4) открытия страницы Интернет

Решение: строка формул, как правило, расположена в верхней части окна программы Excel. Она предназначена для ввода и редактирования в ячейках таблицы текстовых данных, чисел и формул. Данные, внесенные в активную ячейку, дополнительно отображаются в этой строке.

Правильный ответ: 2.

33. СТРОКА ФОРМУЛ ПРОГРАММЫ EXCEL ПОКАЗЫВАЕТ:

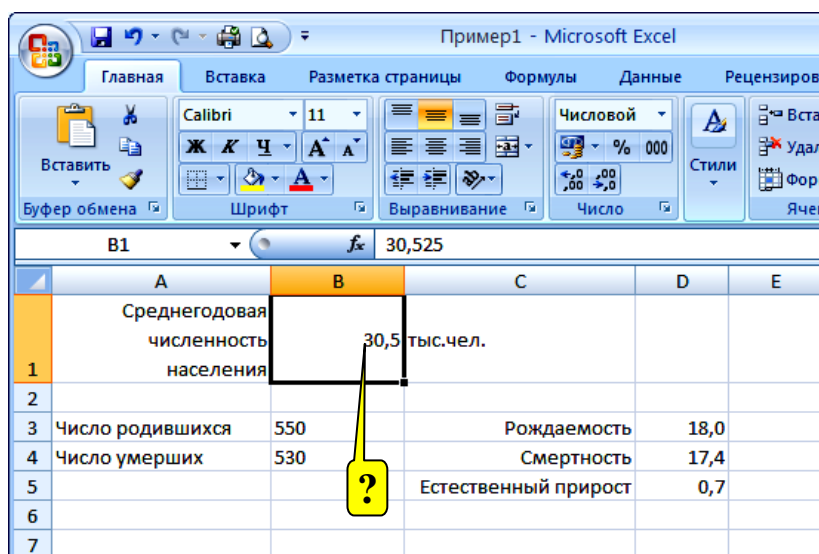
- 1) адрес ячейки, обозначенной табличным курсором
- 2) имя обрабатываемого файла
- 3) наименование открытого листа
- 4) данные и формулы, расположенные в активной ячейке

Решение: строка формул, как правило, расположена в верхней части окна программы Excel. Она предназначена для ввода и редактирования в ячейках таблицы текстовых данных, чисел и формул. Данные, внесенные в активную ячейку, дополнительно отображаются в этой строке

Правильный ответ: 1, 4.

34. В ЯЧЕЙКЕ В1 ОТОБРАЖЕНО ЧИСЛО 30,5 (СМ. РИСУНОК), ПРИ ЭТОМ В ВЫЧИСЛЕНИЯХ ИСПОЛЬЗУЕТСЯ ЗНАЧЕНИЕ ...

- 1) 30,5
- 2) 30,525
- 3) в зависимости от настройки программы: 30,5 или 30,525



Решение: в ячейке В1 отображено число 30,5 (см.рисунок), при этом в вычислениях используется значение 30,525.

Правильный ответ: 2.

35. ВВОД ФОРМУЛЫ В ЯЧЕЙКУ ТАБЛИЦЫ EXCEL НАЧИНАЕТСЯ С ...

- 1) знака «=» (знака равенства)
- 2) указания адреса ячейки
- 3) выполнения команды «Формат»
- 4) сохранения документа

Решение: в программе Excel ввод формулы в ячейку таблицы начинается с «=» (знака равенства).

Правильный ответ: 1.

36. ФОРМУЛА ВЫЧИСЛЕНИЯ В ПРОГРАММЕ EXCEL ВСЕГДА НАЧИНАЕТСЯ С СИМВОЛА ...

- 1) любого алфавитного знака
- 2) «'» (апостроф)
- 3) «~» (тильда)

4) «=» (равно)

Решение: одной из важнейших функций электронных таблиц является создание формул, позволяющих выполнять вычисления в таблице. Формула всегда должна начинаться со знака «=» (равно). Все элементы в формулах обозначаются латинскими буквами, без пробелов.

Правильный ответ: 4.

37. КОРРЕКТИРОВКА ФОРМУЛЫ, ВВЕДЕННОЙ В ЯЧЕЙКУ ПРОГРАММЫ EXCEL:

- 1) не возможна
- 2) выполняется в ячейке таблицы
- 3) происходит в отдельном диалоговом окне
- 4) выполняется в строке формул

Решение: корректировка формулы, введенной в ячейку таблицы Excel, выполняется в строке формул, которая расположена в верхней части окна ниже панели инструментов, или непосредственно в ячейке таблицы.

Правильный ответ: 2, 4.

38. В ПРОГРАММЕ EXCEL АДРЕС ЯЧЕЙКИ \$A\$1 ЯВЛЯЕТСЯ...

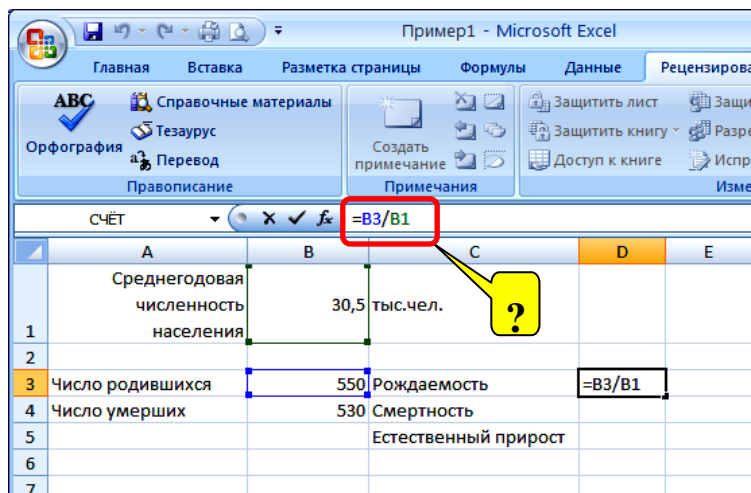
- 1) смешанным
- 2) абсолютным
- 3) пользовательским
- 4) относительным

Решение: в программе Excel добавление в адрес ячейки знака денежной единицы «\$», служит для указания абсолютных адресов, которые не изменяются при копировании.

Правильный ответ: 2.

39. ИЗМЕНЕНИЕ ТИПА АДРЕСА ЯЧЕЙКИ B1 В ФОРМУЛЕ (СМ. РИСУНОК) НА АБСОЛЮТНЫЙ ВЫПОЛНЯЕТСЯ С ПОМОЩЬЮ ...

- 1) ввода с клавиатуры #B#1
- 2) ввода с клавиатуры \$B\$1
- 3) ввода с клавиатуры &B&1
- 4) ввода с клавиатуры &B\$1
- 5) клавиши F4



Решение: строка формул, которая расположена в верхней части окна ниже панели инструментов, предназначенная для ввода и редактирования в ячейках таблицы текстовых данных, чисел и формул. Данные, внесенные в активную ячейку, дополнительно отображаются в этой строке. Изменение типа адреса ячейки В1 в формуле (см. рисунок) на абсолютный выполняется в строке формул с помощью ввода с клавиатуры $\$B\1 или клавишей F4.

Правильный ответ: 2, 3.

40. ФОРМУЛА ВЫЧИСЛЕНИЙ В ЯЧЕЙКЕ ТАБЛИЦЫ EXCEL МОЖЕТ ВКЛЮЧАТЬ ...

- 1) знак «=» (равно), а затем адреса числовых ячеек, константы и функции
- 2) только адреса ячеек, из которых извлекается число для вычислений
- 3) текстовое описание расчета
- 4) только функции вычислений

Решение: формула вычислений в ячейке таблицы Excel может включать знак «=» (равно), а затем адреса числовых ячеек, константы и функции.

Правильный ответ: 1.

41. ДЛЯ ВВОДА ФОРМУЛЫ ВЫЧИСЛЕНИЙ В ЯЧЕЙКУ ТАБЛИЦЫ EXCEL ПРИМЕНЯЮТСЯ СЛЕДУЮЩИЕ СПОСОБЫ:

- 1) ввод с клавиатуры знака «=» (равно), а затем расчетной части формулы
- 2) ввод текста описания расчета с указанием адресов ячеек, используемых в вычислениях
- 3) выполнение команды «Вставить функцию» с выбором способа вычисления в «Мастере функций»

Решение: формула вычислений в ячейке таблицы Excel может быть вставлена набором с клавиатуры знака равно (=), а затем вводом адресов ячеек, констант и функций. Дополнительным способом использования формул является применение «Мастера функций», который позволяет выбрать требуемый алгоритм вычислений из списка функций.

Правильный ответ: 1, 3.

42. В ПРОГРАММЕ EXCEL ВЫЧИСЛЕНИЕ СУММЫ ЗНАЧЕНИЙ В ДИАПАЗОНЕ ЯЧЕЕК ВЫПОЛНЯЕТСЯ ФУНКЦИЕЙ ...

- 1) СУММ()
- 2) МАКС()
- 3) ДИСП()
- 4) СЧЕТ()

Решение: в программе Excel сумма значений в диапазоне ячеек может быть вычислена функцией СУММ().

Правильный ответ: 1.

43. В ПРОГРАММЕ EXCEL ВЫБОР ФУНКЦИИ ИЗ СПИСКА ВЫПОЛНЯЕТСЯ С ПОМОЩЬЮ...

- 1) создания новой книги
- 2) мастера функций
- 3) надстроек
- 4) справочной системы

Решение: в программе Excel для облегчения ввода имени функции вычислений может применяться «Мастер функций», который позволяет выполнить выбор требуемой функции из соответствующей категории.

Правильный ответ: 2.

44. В ПРЕДСТАВЛЕННОМ ФРАГМЕНТЕ ЭЛЕКТРОННОЙ ТАБЛИЦЫ EXCEL ПРИ ИСПОЛЬЗОВАНИИ УКАЗАННОЙ ФОРМУЛЫ (СМ. РИСУНОК) ЗНАЧЕНИЕ В ЯЧЕЙКЕ В3 БУДЕТ РАВНО...

	A	B
1	3	2
2	4	3
3		=МАКС(A1:B2;A1+B2;A2+A1)

- 1) 5
- 2) 12
- 3) 7
- 4) 3

Решение: в программе Excel функция МАКС служит для вычисления максимального значения в диапазоне ячеек, указанных в скобках и перечисленных через точку с запятой, поэтому в ячейке В3 будет вычислено значение 7.

Правильный ответ: 3.

45. ЕСЛИ ФОРМУЛА ВЫЧИСЛЕНИЯ В ПРОГРАММЕ EXCEL СОДЕРЖИТ ОШИБКУ, ТО СООБЩЕНИЕ О НЕЙ ОТОБРАЖАЕТСЯ В ...

- 1) строке формул
- 2) панели инструментов
- 3) ярлыке листа
- 4) ячейке таблицы

Решение: в случае ввода в ячейку таблицы формулы с ошибкой, программа Excel сообщит о ней в ячейке таблицы.

Правильный ответ: 4.

46. В ПРОГРАММЕ EXCEL ПРИЧИНОЙ ПОЯВЛЕНИЯ ОШИБКИ #ЗНАЧ! ПРИ ВВОДЕ ФОРМУЛЫ ЯВЛЯЕТСЯ ...

- 1) использование недопустимого типа данных, расположенных в ячейке, адрес которой указан в формуле
- 2) числовое значение не умещается по ширине ячейки
- 3) в формуле указан несуществующий или неправильный адрес ячейки
- 4) формула содержит вычисления включающие деление на ноль

Решение: если при вводе формулы в программе Excel в ячейке отображается ошибка #ЗНАЧ!, это указывает на использование недопустимого типа данных в формуле, например: используется адрес текстовой ячейки. Для исправления ошибки требуется изменение формулы или корректировка данных в ячейке, адрес которой используется в формуле.

Правильный ответ: 1.

47. В ПРОГРАММЕ EXCEL ПРИЧИНОЙ ПОЯВЛЕНИЯ ОШИБКИ #ИМЯ? ЯВЛЯЕТСЯ ...

- 1) использование недопустимого типа данных, расположенных в ячейке, адрес которой указан в формуле
- 2) числовое значение не умещается по ширине ячейки
- 3) в формуле указан несуществующий или неправильный адрес ячейки
- 4) формула содержит вычисления включающие деление на ноль

Решение: если при вводе формулы в программе Excel в ячейке отображается ошибка #ИМЯ?, это указывает на использование несуществующего или неправильно указанного адреса ячейки. Для исправления ошибки необходимо проверить правильность обозначения адресов ячеек в формуле.

Правильный ответ: 3.

48. В ПРОГРАММЕ EXCEL ПРИЧИНОЙ ПОЯВЛЕНИЯ ОШИБКИ #ДЕЛ/0! ЯВЛЯЕТСЯ ...

- 1) использование недопустимого типа данных, расположенных в ячейке, адрес которой указан в формуле
- 2) числовое значение не умещается по ширине ячейки
- 3) в формуле указан несуществующий или неправильный адрес ячейки
- 4) формула содержит вычисления включающие деление на ноль

Решение: если при вводе формулы в программе Excel в ячейке отображается ошибка #ДЕЛ/0, это указывает на попытку деления на ноль. Для исправления ошибки необходимо проверить правильность формулы и содержимое адресов ячеек, используемых в ней.

Правильный ответ: 4.

49. ДЛЯ СОЗДАНИЯ РИСУНКА ДИАГРАММЫ В ПРОГРАММЕ EXCEL НЕОБХОДИМО ИСПОЛЬЗОВАТЬ РЕЖИМ ...

- 1) Главная
- 2) Вставка
- 3) Рецензирование
- 4) Вид

Решение: в программе Excel для создания рисунка диаграммы необходимо использовать режим «Вставка».

Правильный ответ: 2.

50. В EXCEL СУЩЕСТВУЮТ СЛЕДУЮЩИЕ ОСНОВНЫЕ ТИПЫ ДИАГРАММ:

- 1) Гистограмма
- 2) График
- 3) Круговая
- 4) Линейчатая
- 5) С областями
- 6) Точечная
- 7) Рисованная

Решение: в программе Excel в режиме «Вставка» содержатся команды создания основных типов диаграмм: «Гистограмма», «График», «Круговая», «Линейчатая», «С областями» и «Точечная».

Правильный ответ: 1, 2, 3, 4, 5, 6.

51. ТИП ДИАГРАММЫ СОДЕРЖАЩЕЙ ВЕРТИКАЛЬНЫЕ СТОЛБИКИ – ЭТО ...

- 1) Гистограмма
- 2) График
- 3) Линейчатая
- 4) Круговая

Решение: диаграмма, содержащая вертикальные столбики называется «Гистограмма».

Правильный ответ: 1.

52. ДЛЯ ПОДКЛЮЧЕНИЯ К ПРОГРАММЕ EXCEL ПАКЕТА СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ ПОЛЬЗОВАТЕЛЮ ТРЕБУЕТСЯ ВЫПОЛНИТЬ СЛЕДУЮЩЕЕ ДЕЙСТВИЕ ...

- 1) произвести установку модуля «Пакет анализа» из режима «Надстройки» программы
- 2) удалить программу Excel и произвести ее повторную стандартную установку
- 3) перезагрузить компьютер

Решение: для подключения к программе Excel пакета статистического анализа данных необходимо подключить модуль «Пакет анализа» из режима «Надстройки» основного меню программы.

Правильный ответ: 1.

53. ДЛЯ ВЫЧИСЛЕНИЯ СРЕДНЕГО АРИФМЕТИЧЕСКОГО ЗНАЧЕНИЯ ПРИМЕНЯЕТСЯ ФУНКЦИЯ ...

- 1) МЕДИАНА(...)
- 2) СРГАРМ(...)
- 3) СРЗНАЧ(...)
- 4) СРГЕОМ(...)

Решение: для вычисления среднего арифметического значения применяется функция СРЗНАЧ(диапазон ячеек).

Правильный ответ: 3.

54. ПОЛНЫМ ОПРЕДЕЛЕНИЕМ СТАТИСТИКИ ЯВЛЯЕТСЯ ...

- 1) наука об общих методах изучения массовых явлений
- 2) общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной в конкретных исторических условиях
- 3) универсальная наука, подвергающая количественному изучению все явления общества и природы
- 4) наука об особенностях деятельности медицинских учреждений в условиях рыночной экономики и страховой медицины

Решение: статистика - это общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной в конкретных исторических условиях.

Правильный ответ: 2.

55. ПРЕДМЕТАМИ ИЗУЧЕНИЯ МЕДИЦИНСКОЙ СТАТИСТИКИ ЯВЛЯЮТСЯ:

- 1) здоровье населения в целом и отдельных возрастно-половых групп;
- 2) выявление и установление зависимостей между уровнем здоровья и факторами окружающей среды
- 3) финансовые результаты деятельности медицинских учреждений
- 4) данные о сети, деятельности, кадрах учреждений здравоохранения
- 5) оценка статистической достоверности результатов медико-биологических, клинических и экспериментальных исследований

Решение: предметами изучения медицинской статистики являются: здоровье населения в целом и отдельных возрастно-половых групп; выявление и установление зависимостей между уровнем здоровья и факторами окружающей среды; данные о сети, деятельности, кадрах учреждений здравоохранения; оценка статистической достоверности результатов медико-биологических, клинических и экспериментальных исследований.

Правильный ответ: 1, 2, 4, 5.

56. ОБЪЕКТ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ – ЭТО ...

- 1) отдельная социальная или биологическая единица наблюдения, подлежащая углубленному изучению и регистрации ее признаков в специальной учетной форме (бланке)
- 2) место или территория, где осуществляется статистическое исследование
- 3) статистическая совокупность, состоящая из единиц, о которых должны быть собраны статистические сведения, взятая в определенных границах времени и пространства
- 4) отдельное ЛПУ, в котором проводится статистическое исследование

Решение: объект статистического наблюдения – это статистическая совокупность, состоящая из единиц, о которых должны быть собраны статистические сведения, взятая в определенных границах времени и пространства.

Правильный ответ: 3.

57. ОБЪЕКТОМ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ПРИ ИЗУЧЕНИИ ДЕТСКОГО УЛИЧНОГО ТРАВМАТИЗМА ЯВЛЯЕТСЯ ...

- 1) выборочная часть детского травматизма за определенный период
- 2) все случаи детского уличного травматизма на определенной территории за определенный период
- 3) все случаи уличного травматизма на определенной территории за определенный период
- 4) отдельное ЛПУ, в котором лечились пострадавшие

Решение: объектом статистического исследования при изучении детского уличного травматизма является все случаи детского уличного травматизма на определенной территории за определенный период.

Правильный ответ: 2.

58. ЕДИНИЦА СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ – ЭТО...

- 1) составная часть объекта наблюдения, подлежащая изучению и регистрации в соответствии с программой исследования
- 2) явление, которое подлежит детальному изучению, и все учетные признаки которого могут быть измерены только количественно
- 3) явление, которое подлежит детальному изучению и его учетные признаки должны носить только качественный, описательный характер
- 4) место или территория, где осуществляется статистическое исследование

Решение: единица статистического наблюдения – это составная часть объекта наблюдения, подлежащая изучению и регистрации в соответствии с программой исследования.

Правильный ответ: 1.

59. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ МОЖЕТ БЫТЬ:

- 1) текущим
- 2) нормированным
- 3) результативным
- 4) единовременным.

Решение: статистическое наблюдение может быть единовременным и текущим.

Правильный ответ: 1, 4.

60. МЕТОДАМИ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ ЯВЛЯЮТСЯ:

- 1) сплошное
- 2) выборочное
- 4) нормированное
- 5) результативное

Решение: методами статистического наблюдения являются: сплошное исследование, при котором изучаются все доступные единицы наблюдения; выборочное - изучается определенная часть единиц наблюдения, наиболее полно характеризующие статистическую совокупность в целом.

Правильный ответ: 1, 2.

61. СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ – ЭТО ...

- 1) группа, состоящая из большого числа относительно однородных элементов (единиц наблюдения), взятых вместе в известных границах времени и пространства
- 2) явление, которое подлежит детальному изучению, и все учетные признаки которого могут быть измерены только количественно
- 3) явление, которое подлежит детальному изучению и его учетные признаки должны носить только качественный, описательный характер
- 4) место или территория, где осуществляется статистическое исследование

Решение: статистическая совокупность - это группа, состоящая из большого числа относительно однородных элементов (единиц наблюдения), взятых вместе в известных границах времени и пространства. Необходимо различать два основных вида статистических совокупностей: генеральная и выборочная.

Правильный ответ: 1.

62. ГЕНЕРАЛЬНАЯ СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ – ЭТО ...

- 1) явление, которое подлежит детальному изучению, и все учетные признаки которого могут быть измерены только количественно
- 2) явление, которое подлежит детальному изучению и его учетные признаки должны носить только качественный, описательный характер
- 3) место или территория, где осуществляется статистическое исследование
- 4) набор всех возможных единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования

Решение: генеральная статистическая совокупность состоит из всех возможных единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования.

Правильный ответ: 4.

63. ВЫБОРОЧНАЯ СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ – ЭТО ...

- 1) явление, которое подлежит детальному изучению, и все учетные признаки которого могут быть измерены только количественно
- 2) часть генеральной совокупности, отобранная специальным методом и предназначенная для характеристики генеральной совокупности
- 3) место или территория, где осуществляется статистическое исследование
- 4) набор всех возможных единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования

Решение: выборочная статистическая совокупность – это часть генеральной совокупности, отобранная специальным методом и предназначенная для характеристики генеральной совокупности.

Правильный ответ: 2.

64. УЧЕТНЫЕ ПРИЗНАКИ - ЭТО ...

- 1) совокупность математических критериев, используемых при статистическом исследовании
- 2) медико-биологические характеристики, регистрируемые у единицы наблюдения в соответствии с целями и задачами исследования
- 3) относительные величины, сгруппированные по определенным признакам в статистическую таблицу
- 4) статистические показатели, характеризующие изучаемое явление

Решение: учетные признаки - это медико-биологические характеристики, регистрируемые у единицы наблюдения в соответствии с целями и задачами исследования.

Правильный ответ: 2.

65. В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ ИСПОЛЬЗУЮТСЯ УЧЕТНЫЕ ПРИЗНАКИ:

- 1) сходства
- 2) различия
- 3) факторные
- 4) результативные
- 5) интервальные

Решение: в медико-биологических исследованиях используются учетные признаки: сходства, различия, факторные, результативные.

Правильный ответ: 1, 2, 3, 4.

66. В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ ВСТРЕЧАЮТСЯ СЛЕДУЮЩИЕ ТИПЫ УЧЕТНЫХ ПРИЗНАКОВ:

- 1) непрерывные
- 2) качественные (описательные или атрибутивные)
- 3) ранжированные
- 4) количественные

Решение: в медико-биологических исследованиях встречаются следующие типы учетных признаков: качественные (описательные или атрибутивные) – выражающие словом изучаемое свойство, и количественные – демонстрирующие числом уровень признака.

Правильный ответ: 2, 4.

67. ПОСЛЕДОВАТЕЛЬНОСТЬ ЭТАПОВ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ:

- 1) статистическое наблюдение
- 2) разработка программы и составление плана статистического исследования
- 3) анализ результатов исследования
- 4) группировка и разработка статистического материала

Решение: последовательность этапов статистического исследования состоит из:
1-й - разработка программы и составление плана статистического

исследования; 2-й - статистическое наблюдение; 3-й - группировка и разработка статистического материала; 4-й - анализ результатов исследования.

Правильный ответ: 1-2, 2-1, 3-4, 4-3.

68. ОСНОВНЫМИ ВИДАМИ РАБОТ НА ПЕРВОМ ЭТАПЕ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ЯВЛЯЮТСЯ:

- 1) вычисление статистических показателей;
- 2) сопоставление статистических данных;
- 3) графическое изображение показателей;
- 4) шифровка (кодирование) статистического материала;
- 5) выявление закономерностей в изучаемых явлениях;
- 6) обобщение результатов исследования;
- 7) группировка статистического материала;
- 8) составление плана исследования;
- 9) подготовка программы исследования.

Решение: основными видами работ на первом этапе статистического исследования являются составление плана и подготовка программы исследования.

Правильный ответ: 8, 9.

69. ПЛАН СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ ...

- 1) установление качественных и количественных закономерностей в изучаемых явлениях
- 2) составление статистических таблиц с результатами сбора материала;
- 3) вопросы: что и в каком направлении изучать, с обозначением объекта и единиц наблюдения, учетных признаков, методов сбора, разработки и анализа материала
- 4) вопросы: где, когда, кто и как выполняет исследование
- 5) вопросы контроля за однородностью статистической совокупности и правила составления статистических таблиц

Решение: план статистического исследования включает вопросы: где, когда, кто и как выполняет исследование.

Правильный ответ: 4.

70. ПРОГРАММА СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ ...

- 1) установление качественных и количественных закономерностей в изучаемых явлениях
- 2) составление статистических таблиц с результатами сбора материала
- 3) вопросы: что и в каком направлении изучать, с обозначением объекта и единиц наблюдения, учетных признаков, методов сбора, разработки и анализа материала
- 4) вопросы: где, когда, кто и как выполняет исследование
- 5) вопросы контроля за однородностью статистической совокупности и правила составления статистических таблиц

Решение: программа исследования включает вопросы: что и в каком направлении изучать, с обозначением объекта и единиц наблюдения, учетных признаков, методов сбора, разработки и анализа материала.

Правильный ответ: 3.

71. СОСТАВЛЕНИЕ ПЛАНА И ПРОГРАММЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ПРОИЗВОДИТСЯ ...

- 1) на первом этапе
- 2) на втором этапе
- 3) на третьем этапе
- 4) на четвертом этапе

Решение: составление плана и программы производится на первом этапе статистического исследования.

Правильный ответ: 1.

72. ОСНОВНЫМ ВИДОМ РАБОТ НА ВТОРОМ ЭТАПЕ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ЯВЛЯЕТСЯ ...

- 1) вычисление статистических показателей;
- 2) сбор материала по программе исследования;
- 3) графическое изображение показателей;
- 4) шифровка (кодирование) статистического материала;
- 5) выявление закономерностей в изучаемых явлениях;
- 6) обобщение результатов исследования;
- 7) группировка статистического материала;
- 8) составление плана и программы исследования.

Решение: основным видом работ на втором этапе статистического исследования является сбор материала в соответствии с программой исследования.

Правильный ответ: 2.

73. ОСНОВНЫМИ ВИДАМИ РАБОТ НА ТРЕТЬЕМ ЭТАПЕ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ЯВЛЯЮТСЯ:

- 1) шифровка (кодирование) статистического материала
- 2) сбор материала по программе исследования
- 3) группировка статистического материала
- 4) вычисление статистических показателей
- 5) выявление закономерностей в изучаемых явлениях
- 6) обобщение результатов исследования
- 7) графическое изображение данных
- 8) составление плана и программы исследования

Решение: основными видами работ на третьем этапе статистического исследования являются: шифровка (кодирование) статистического материала; группировка; вычисление статистических показателей; графическое изображение данных.

Правильный ответ: 1, 3, 4, 7.

74. ПРИ ПРОВЕДЕНИИ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВЫЧИСЛЕНИЕ ПОКАЗАТЕЛЕЙ, А ТАКЖЕ ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ МАТЕРИАЛА ВЫПОЛНЯЮТСЯ ...

- 1) на втором этапе
- 2) на первом этапе
- 3) на третьем этапе
- 4) на четвертом этапе
- 5) на всех этапах

Решение: при проведении статистического исследования вычисление показателей, а также графическое представление материала выполняются на третьем этапе.

Правильный ответ: 3.

75. ОСНОВНЫМИ ВИДАМИ ГРУППИРОВОК ЕДИНИЦ НАБЛЮДЕНИЯ ЯВЛЯЮТСЯ:

- 1) вариационная
- 2) абсолютная
- 3) типологическая
- 4) относительная

Решение: основными видами группировок единиц наблюдения являются вариационная и типологическая.

Правильный ответ: 1, 3.

76. ОСНОВНЫМИ ВИДАМИ РАБОТ НА ЧЕТВЕРТОМ ЭТАПЕ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ЯВЛЯЮТСЯ:

- 1) вычисление статистических показателей
- 2) сопоставление статистических данных
- 3) графическое изображение показателей
- 4) шифровка (кодирование) статистического материала
- 5) выявление закономерностей в изучаемых явлениях
- 6) обобщение результатов исследования
- 7) группировка статистического материала

Решение: основными видами работ на четвертом этапе статистического исследования являются: сопоставление статистических данных; выявление закономерностей в изучаемых явлениях; обобщение результатов исследования.

Правильный ответ: 2, 5, 6.

77. ПРИ ПРОВЕДЕНИИ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ОБОБЩЕНИЕ ПОЛУЧЕННЫХ ДАННЫХ, ФОРМИРОВАНИЕ ВЫВОДОВ И ПРЕДЛОЖЕНИЙ ВЫПОЛНЯЕТСЯ ...

- 1) на втором этапе
- 2) на первом этапе
- 3) на третьем этапе

4) на четвертом этапе

5) на всех этапах

Решение: при проведении статистического исследования обобщение полученных данных, формирование выводов и предложений выполняется на четвертом этапе.

Правильный ответ: 4.

78. ОСНОВНЫМИ ВАРИАНТАМИ ПРАКТИЧЕСКОГО ИСПОЛЬЗОВАНИЯ РЕЗУЛЬТАТОВ МЕДИКО-СОЦИАЛЬНОГО ИССЛЕДОВАНИЯ ЯВЛЯЮТСЯ:

1) ознакомление аудитории с его результатами (лекции, доклады, семинары)

2) выпуск методических рекомендаций, приказов и инструкций

3) реорганизация деятельности лечебных учреждений

4) получение прибыли от оказания платных медицинских услуг

5) рационализаторские предложения, изобретения, открытия

Решение: основными вариантами практического использования результатов медико-социального исследования являются: ознакомление аудитории с его результатами (лекции, доклады, семинары); выпуск методических рекомендаций, приказов и инструкций; реорганизация деятельности лечебных учреждений; рационализаторские предложения, изобретения, открытия.

Правильный ответ: 1, 2, 3, 5.

79. ВИДАМИ СТАТИСТИЧЕСКИХ ТАБЛИЦ ЯВЛЯЮТСЯ:

1) простая

2) групповая

3) моментная

4) комбинационная

5) взвешенная

Решение: видами статистических таблиц являются: простая; групповая; комбинационная.

Правильный ответ: 1, 2, 4.

80. ОСНОВНЫМИ ЭЛЕМЕНТАМИ СТАТИСТИЧЕСКИХ ТАБЛИЦ ЯВЛЯЮТСЯ:

1) табличная последовательность

2) табличное подлежащее

3) табличные подразделы

4) табличное сказуемое

5) табличное распределение

Решение: основными элементами статистических таблиц являются табличное подлежащее и табличное сказуемое.

Правильный ответ: 2, 4.

81. СТАТИСТИЧЕСКАЯ ТАБЛИЦА, В КОТОРОЙ ПРЕДСТАВЛЕНА СВОДКА ДАННЫХ ПО ОДНОМУ ПРИЗНАКУ, НАЗЫВАЕТСЯ ...

1) простой

- 2) групповой
- 3) вариационной
- 4) комбинационной

Решение: статистическая таблица, в которой представлена сводка данных по одному признаку, называется простой.

Правильный ответ: 1.

82. СТАТИСТИЧЕСКАЯ ТАБЛИЦА, В КОТОРОЙ ПРЕДСТАВЛЕНЫ ДАННЫЕ ПО ДВУМ СВЯЗАННЫМ МЕЖДУ СОБОЙ ПРИЗНАКАМ, НАЗЫВАЕТСЯ ...

- 1) простой
- 2) групповой
- 3) вариационной
- 4) комбинационной

Решение: статистическая таблица, в которой представлены данные по двум связанным между собой признакам, называется групповой.

Правильный ответ: 2.

83. КОЛИЧЕСТВО СВЯЗАННЫХ МЕЖДУ СОБОЙ ПРИЗНАКОВ ИЗУЧАЕМОГО ЯВЛЕНИЯ, ВКЛЮЧАЕМЫХ В ГРУППОВУЮ ТАБЛИЦУ, СОСТАВЛЯЕТ ...

- 1) один
- 2) два
- 3) три
- 4) четыре и более

Решение: в групповой таблице представлены два связанных между собой признака.

Правильный ответ: 2.

84. СТАТИСТИЧЕСКАЯ ТАБЛИЦА, В КОТОРОЙ ПРЕДСТАВЛЕНЫ ДАННЫЕ ПО ТРЕМ И БОЛЕЕ СВЯЗАННЫМ МЕЖДУ СОБОЙ ПРИЗНАКАМ, НАЗЫВАЕТСЯ ...

- 1) простой
- 2) групповой
- 3) вариационной
- 4) комбинационной
- 5) выборочной

Решение: статистическая таблица, в которой представлены данные по трем и более связанным между собой признакам, называется комбинационной.

Правильный ответ: 4.

85. ТАБЛИЦА, ПОКАЗАННАЯ НА РИСУНКЕ, ЯВЛЯЕТСЯ ...

**ПРИЧИНЫ СМЕРТИ ПО ДАННЫМ
ПАТОЛОГОАНАТОМИЧЕСКИХ ВСКРЫТИЙ**

<i>Заболевания</i>	<i>Число патологоанатомических вскрытий</i>
<i>Инфаркт миокарда</i>	<i>394</i>
<i>Язвенная болезнь желудка</i>	<i>80</i>
<i>Кардиосклероз</i>	<i>156</i>
<i>Рак легкого</i>	<i>200</i>
<i>Итого:</i>	<i>830</i>

- 1) простой
- 2) групповой
- 3) вариационной
- 4) комбинационной
- 5) выборочной

Решение: таблица, показанная на рисунке, является простой т.к. в ней содержится сводка данных по одному признаку (число патологоанатомических вскрытий).

Правильный ответ: 1.

86. ТАБЛИЦА, ПОКАЗАННАЯ НА РИСУНКЕ, ЯВЛЯЕТСЯ ...

**ПРИЧИНЫ СМЕРТИ ПО ДАННЫМ
ПАТОЛОГОАНАТОМИЧЕСКИХ ВСКРЫТИЙ**

<i>Заболевания</i>	<i>Число несовпадений диагнозов</i>		
	<i>Пол</i>		<i>Всего</i>
	<i>М</i>	<i>Ж</i>	
<i>Инфаркт миокарда</i>	<i>20</i>	<i>44</i>	<i>64</i>
<i>Язвенная болезнь желудка</i>	<i>10</i>	<i>12</i>	<i>22</i>
<i>Кардиосклероз</i>	<i>22</i>	<i>14</i>	<i>36</i>
<i>Рак легкого</i>	<i>20</i>	<i>20</i>	<i>40</i>
<i>Итого:</i>	<i>72</i>	<i>90</i>	<i>162</i>

- 1) простой
- 2) групповой
- 3) вариационной
- 4) комбинационной
- 5) выборочной

Решение: таблица, показанная на рисунке, является групповой т.к. в ней содержится сводка данных по двум связанным признакам (несовпадение диагнозов при вскрытии и пол).

Правильный ответ: 2.

87. ТАБЛИЦА, ПОКАЗАННАЯ НА РИСУНКЕ, ЯВЛЯЕТСЯ ...

**ПРИЧИНЫ СМЕРТИ ПО ДАННЫМ
ПАТОЛОГОАНАТОМИЧЕСКИХ ВСКРЫТИЙ**

Заболевания	Возраст						Всего	
	0-16 лет		17-60 лет		61 и более			
	М	Ж	М	Ж	М	Ж	М	Ж
<i>Инфаркт миокарда</i>	0	0	10	22	42	64	52	86
<i>Язвенная болезнь желудка</i>	0	0	20	6	18	22	38	28
<i>Кардиосклероз</i>	2	4	5	15	16	36	23	55
<i>Рак легкого</i>	0	0	15	25	35	40	50	65
Итого:	2	4	50	68	111	162	163	234

- 1) простой
- 2) групповой
- 3) вариационной
- 4) комбинационной
- 5) выборочной

Решение: таблица, показанная на рисунке, является комбинационной т.к. в ней содержится сводка данных по трем связанным признакам (число вскрытий, пол и возраст).

Правильный ответ: 4.

88. ОСНОВНЫМИ ТРЕБОВАНИЯМИ, ПРЕДЪЯВЛЯЕМЫМИ К СТАТИСТИЧЕСКИМ ТАБЛИЦАМ, ЯВЛЯЮТСЯ:

- 1) название таблицы
- 2) номер таблицы
- 3) итоговые данные
- 4) отсутствие пустых ячеек
- 5) одинаковые размеры ячеек

Решение: основными требованиями, предъявляемыми к статистическим таблицам, являются: наличие названия таблицы расположенное на дней, номер таблицы; итоги по столбцам и строкам; отсутствие пустых ячеек.

Правильный ответ: 1, 2, 3, 4.

89. ВИДАМИ СТАТИСТИЧЕСКИХ ВЕЛИЧИН, КОТОРЫЕ ИСПОЛЬЗУЮТСЯ ПРИ ВЫПОЛНЕНИИ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ, ЯВЛЯЮТСЯ:

- 1) абсолютные
- 2) интенсивные
- 3) экстенсивные
- 4) соотношения
- 5) динамические
- 6) простые

Решение: видами статистических величин, которые используются при выполнении медико-биологических исследований, являются: абсолютные, интенсивные, экстенсивные, соотношения и динамические величины.

Правильный ответ: 1, 2, 3, 4, 5.

90. В ТАБЛИЦЕ (СМ. РИСУНОК) УКАЗАН СЛЕДУЮЩИЙ ВИД СТАТИСТИЧЕСКИХ ВЕЛИЧИН ...

СОВПАДЕНИЕ ДИАГНОЗОВ ЛПУ С ДАННЫМИ ПАТОЛОГОАТОМИЧЕСКИХ ВСКРЫТИЙ

<i>Причины смерти по данным ЛПУ</i>	<i>Число пат. анатомических вскрытий</i>	<i>Число несовпадений диагнозов</i>
<i>Инфаркт миокарда</i>	<i>394</i>	<i>64</i>
<i>Язвенная болезнь желудка</i>	<i>80</i>	<i>22</i>
<i>Кардиосклероз</i>	<i>156</i>	<i>36</i>
<i>Итого:</i>	<i>830</i>	<i>162</i>

- 1) абсолютные
- 2) интенсивные
- 3) экстенсивные
- 4) относительные

Решение: таблица, показанная на рисунке, содержит абсолютные величины.

Правильный ответ: 1.

91. ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ ПРИМЕНЯЮТСЯ ДЛЯ ...

- 1) сравнения статистических совокупностей
- 2) оценки вариабельность признака
- 3) определения достоверности различий между средними величинами

Решение: относительные величины применяются для сравнения статистических совокупностей.

Правильный ответ: 1.

92. ЭКСТЕНСИВНЫЕ ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ (ПОКАЗАТЕЛИ) ХАРАКТЕРИЗУЮТ ...

- 1) часть изучаемого явления во всей его совокупности (структуру явления)
- 2) частоту (распространенность) явления в изучаемой среде
- 3) количественные изменения изучаемого явления во времени
- 4) численное соотношение 2-х не связанных между собой совокупностей

Решение: экстенсивные относительные величины (показатели) характеризуют часть изучаемого явления во всей его совокупности (структуру явления).

Правильный ответ: 1.

93. ИНТЕНСИВНЫЕ ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ (ПОКАЗАТЕЛИ) ХАРАКТЕРИЗУЮТ ...

- 1) часть изучаемого явления во всей его совокупности (структуру явления)
- 2) частоту (распространенность) явления в изучаемой среде
- 3) количественные изменения изучаемого явления во времени

4) численное соотношение 2-х не связанных между собой совокупностей

Решение: интенсивные относительные величины (показатели) характеризуют частоту (распространенность) явления в изучаемой среде.

Правильный ответ: 2.

94. ПОКАЗАТЕЛИ СООТНОШЕНИЯ ХАРАКТЕРИЗУЮТ...

1) часть изучаемого явления во всей его совокупности (структуру явления)

2) частоту (распространенность) явления в изучаемой среде

3) количественные изменения изучаемого явления во времени

4) численное соотношение 2-х не связанных между собой совокупностей

Решение: показатели соотношения характеризуют численное соотношение 2-х не связанных между собой совокупностей.

Правильный ответ: 4.

95. ЧАСТОТУ (РАСПРОСТРАНЕННОСТЬ) ЯВЛЕНИЯ ХАРАКТЕРИЗУЕТ...

1) экстенсивный показатель

2) интенсивный показатель

3) показатель соотношения

4) показатель наглядности

Решение: частоту (распространенность) явления характеризует интенсивный показатель.

Правильный ответ: 2.

96. ПОКАЗАТЕЛЬ НАГЛЯДНОСТИ ИСПОЛЬЗУЕТСЯ ...

1) для сравнения изменений величин изучаемого явления во времени по отношению к исходному уровню

2) для характеристики структуры явления

3) для определения распространенности явления

Решение: показатель наглядности используется для сравнения изменений величин изучаемого явления во времени по отношению к исходному уровню.

Правильный ответ: 1.

97. ИЗ ПЕРЕЧИСЛЕННЫХ ПОКАЗАТЕЛЕЙ ЭКСТЕНСИВНЫМ ЯВЛЯЕТСЯ ...

1) доля заболеваний органов дыхания в общей массе заболеваний;

2) обеспеченность населения больничными койками;

3) заболеваемость дифтерией на 1000 жителей;

4) рост числа заболеваний в текущем году по отношению к предыдущему.

Решение: из перечисленных показателей экстенсивным является доля заболеваний органов дыхания в общей массе заболеваний.

Правильный ответ: 1.

98. ДЛЯ ОПРЕДЕЛЕНИЯ СТРУКТУРЫ ЗАБОЛЕВАЕМОСТИ ПРИМЕНЯЮТСЯ ...

1) показатели соотношения

- 2) интенсивные показатели
- 3) экстенсивные показатели
- 4) показатели наглядности

Решение: для определения структуры заболеваемости применяются экстенсивные показатели.

Правильный ответ: 3.

99. ПОКАЗАТЕЛЕМ СООТНОШЕНИЯ ИЗ ПЕРЕЧИСЛЕННЫХ ЯВЛЯЕТСЯ ...

- 1) доля заболеваний органов дыхания от общего числа заболеваний
- 2) обеспеченность населения больничными койками
- 3) заболеваемость дифтерией на 1000 жителей
- 4) рост числа заболеваний в текущем году по отношению к предыдущему

Решение: обеспеченность населения больничными койками является показателем соотношения.

Правильный ответ: 2.

100. ФОРМУЛА ВЫЧИСЛЕНИЯ ИНТЕНСИВНОГО ПОКАЗАТЕЛЯ, ХАРАКТЕРИЗУЮЩЕГО УРОВЕНЬ ЗАБОЛЕВАЕМОСТИ НАСЕЛЕНИЯ (УСЛОВНЫЕ ОБОЗНАЧЕНИЯ: ЧСЗ - ЧИСЛО СЛУЧАЕВ ЗАБОЛЕВАНИЯ, ЧЖ - ЧИСЛО ЖИТЕЛЕЙ НАСЕЛЕННОГО ПУНКТА, ЧВ - ЧИСЛО ВРАЧЕЙ) ...

- 1) $(ЧСЗ \cdot 1000) / ЧВ$
- 2) $(ЧСЗ \cdot ЧЖ) / 1000$
- 3) $(ЧСЗ \cdot 1000) / ЧЖ$
- 4) $(ЧЖ \cdot 1000) / ЧСЗ$

Решение: формула вычисления интенсивного показателя, характеризующего уровень заболеваемости населения: $(ЧСЗ \cdot 1000) / ЧЖ$, (условные обозначения: ЧСЗ - число случаев заболевания, ЧЖ - число жителей населенного пункта, ЧВ - число врачей).

Правильный ответ: 3.

101. ПОКАЗАТЕЛЕМ НАГЛЯДНОСТИ ЯВЛЯЕТСЯ ...

- 1) доля заболеваний органов дыхания от общего числа заболеваний;
- 2) обеспеченность населения больничными койками;
- 3) заболеваемость населения гриппом на 1000 жителей;
- 4) рост числа заболеваний в текущем году по отношению к предыдущему.

Решение: показателем наглядности является рост числа заболеваний в текущем году по отношению к предыдущему.

Правильный ответ: 4.

102. ДИНАМИЧЕСКИЕ ПОКАЗАТЕЛИ ХАРАКТЕРИЗУЮТ ...

- 1) часть изучаемого явления во всей его совокупности (структуру явления)
- 2) частоту (распространенность) явления в изучаемой среде
- 3) количественные изменения изучаемого явления во времени

4) численное соотношение 2-х не связанных между собой совокупностей

Решение: динамические показатели характеризуют количественные изменения изучаемого явления во времени.

Правильный ответ: 3.

103. ПОКАЗАТЕЛИ, КОТОРЫЕ МОЖНО ВЫЧИСЛИТЬ ПРИ АНАЛИЗЕ КАЧЕСТВА ДИАГНОСТИКИ ПО РЕЗУЛЬТАТАМ ВСКРЫТИЙ (СМ. ТАБЛИЦУ), ОТНОСЯТСЯ К ...

УТОЧНЕНИЕ ПРИЧИН СМЕРТИ ПО ДАННЫМ ПАТОЛОГОАНАТОМИЧЕСКИХ ВСКРЫТИЙ

<i>Причины смерти по данным ЛПУ</i>	<i>Число пат.анат. вскрытий</i>	<i>Число несовпадений диагнозов</i>
<i>Инфаркт миокарда</i>	<i>394</i>	<i>64</i>
<i>Язвенная болезнь желудка</i>	<i>80</i>	<i>22</i>
<i>Кардиосклероз</i>	<i>156</i>	<i>36</i>
<i>Рак легкого</i>	<i>200</i>	<i>40</i>
<i>Итого:</i>	<i>830</i>	<i>162</i>

- 1) интенсивным
- 2) наглядности
- 3) средним величинам
- 4) экстенсивным
- 5) соотношения

Решение: показатели, которые можно вычислить при анализе качества диагностики по результатам вскрытий (см.таблицу), относятся к экстенсивным относительным величинам.

Правильный ответ: 4.

104. ДЛЯ ГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ ИНТЕНСИВНЫХ ПОКАЗАТЕЛЕЙ, КАК ПРАВИЛО, ИСПОЛЬЗУЮТСЯ:

- 1) секторная диаграмма
- 2) столбиковая диаграмма
- 3) внутрестолбиковая диаграмма
- 4) линейная диаграмма
- 5) радиальная диаграмма

Решение: для графического изображения интенсивных показателей, как правило, используются столбиковая и линейная диаграммы.

Правильный ответ: 2, 4.

105. ДЛЯ ГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ ЭКСТЕНСИВНЫХ ПОКАЗАТЕЛЕЙ ИСПОЛЬЗУЮТСЯ:

- 1) секторная диаграмма
- 2) столбиковая диаграмма
- 3) внутрестолбиковая диаграмма

- 4) линейная диаграмма
- 5) радиальная диаграмма

Решение: для графического изображения экстенсивных показателей используются секторная и внутрисклонбиковая диаграммы.

Правильный ответ: 1, 3.

106. ДЛЯ ГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ ДИНАМИЧЕСКОГО РЯДА ИСПОЛЬЗУЮТСЯ:

- 1) секторная диаграмма
- 2) столбиковая диаграмма
- 3) внутрисклонбиковая диаграмма
- 4) линейная диаграмма
- 5) картограмма

Решение: для графического изображения динамического ряда используются столбиковая и линейная диаграммы.

Правильный ответ: 2, 4.

107. ДЛЯ ГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ ЦИКЛИЧЕСКИХ ПРОЦЕССОВ ИСПОЛЬЗУЕТСЯ ...

- 1) секторная диаграмма
- 2) столбиковая диаграмма
- 3) радиальная диаграмма
- 4) линейная диаграмма
- 5) картограмма

Решение: для графического изображения циклических процессов используется радиальная диаграмма.

Правильный ответ: 3.

108. КАРТОГРАММА - ЭТО ...

- 1) круг, разделенный на секторы, с цветной штриховкой
- 2) столбики с различной штриховкой
- 3) географическая карта с различной штриховкой
- 4) географическая карта с нанесенными на нее диаграммами

Решение: картограмма - это географическая карта с различной штриховкой.

Правильный ответ: 3.

109. КАРТОДИАГРАММА – ЭТО ...

- 1) круг, разделенный на секторы, с цветной штриховкой
- 2) столбики с различной штриховкой
- 3) географическая карта с различной штриховкой
- 4) географическая карта с нанесенными на нее диаграммами

Решение: картодиаграмма - это географическая карта с нанесенными на нее диаграммами.

Правильный ответ: 4.

110. ОСНОВНЫМИ ТРЕБОВАНИЯМИ, ПРЕДЪЯВЛЯЕМЫМИ К ДИАГРАММАМ, ЯВЛЯЮТСЯ:

- 1) название диаграммы
- 2) номер рисунка
- 3) наличие имени автора диаграммы
- 4) наличие количественного определителя представленных явлений рядом с диаграммой

Решение: основными требованиями, предъявляемыми к диаграммам, являются: название, помещаемое под рисунком; его порядковый номер; наличие количественного определителя представленных явлений рядом с диаграммой.

Правильный ответ: 1, 2, 4.

111. ОБЛАСТИ ВОЗМОЖНОГО ИСПОЛЬЗОВАНИЯ СРЕДНИХ ВЕЛИЧИН:

- 1) изучение состояния здоровья населения
- 2) анализ демографической ситуации
- 3) оценка деятельности лечебных учреждений
- 4) санитарно-эпидемиологические исследования
- 5) изучение экономических аспектов здравоохранения
- 6) рентгенографические исследования

Решение: области возможного использования средних величин: изучение состояния здоровья населения; анализ демографической ситуации; оценка деятельности лечебных учреждений; санитарно-эпидемиологические исследования; изучение экономических аспектов здравоохранения.

Правильный ответ: 1, 2, 3, 4, 5.

112. ВАРИАЦИОННЫЙ РЯД – ЭТО ...

- 1) однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления
- 2) ряд чисел, сгруппированных в таблицу и полученных при измерении изучаемого признака у единиц наблюдения в соответствии с планом и программой исследования
- 3) количественное выражение изучаемого признака

Решение: вариационный ряд – это однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления.

Правильный ответ: 1.

113. К ОБОЗНАЧЕНИЯМ ЭЛЕМЕНТОВ ВАРИАЦИОННОГО РЯДА ОТНОСЯТСЯ:

- 1) V – варианта

- 2) p – частота повторений варианты
- 3) n – общее число наблюдений
- 4) m – ошибка репрезентативности
- 5) σ – среднее квадратическое отклонение

Решение: к обозначениям элементов вариационного ряда относятся: V – варианта; p – частота повторений варианты; n – общее число наблюдений.

Правильный ответ: 1, 2, 3.

114. ЧАСТОТА ПОВТОРЕНИЙ ВАРИАНТЫ - ЭТО:

- 1) число, указывающее, сколько раз встречается отдельная варианта в вариационном ряду, обозначаемое « p »
- 2) общее число наблюдений в вариационном ряду, обозначаемое « n »
- 3) варианта, расположенная в середине вариационного ряда, упорядоченного по возрастанию или убыванию, обозначаемая « Me »

Решение: частота повторений варианты - это число, указывающее, сколько раз встречается отдельная варианта в вариационном ряду, обозначаемое « p ».

Правильный ответ: 1.

115. РАЗЛИЧАЮТ СЛЕДУЮЩИЕ ВИДЫ ВАРИАЦИОННЫХ РЯДОВ:

- 1) простой
- 2) ранжированный
- 3) сгруппированный
- 4) дискретный (прерывный)
- 5) непрерывный
- 6) интервальный
- 7) сложный
- 8) репрезентативный

Решение: различают следующие виды вариационных рядов: простой, ранжированный, сгруппированный, дискретный (прерывный), непрерывный, интервальный.

Правильный ответ: 1, 2, 3, 4, 5, 6.

116. ВАРИАЦИОННЫЙ РЯД НАЗЫВАЕТСЯ ПРОСТЫМ, ЕСЛИ В НЕМ ...

- 1) указано, сколько раз встречается каждая варианта
- 2) варианты расположены в порядке возрастания и указана Mo (мода)
- 3) варианты расположены в порядке возрастания или убывания
- 4) каждая варианта встречается только один раз

Решение: вариационный ряд называется простым, если в нем каждая варианта встречается только один раз.

Правильный ответ: 4.

117. ВАРИАЦИОННЫЙ РЯД НАЗЫВАЕТСЯ СГРУППИРОВАННЫМ, ЕСЛИ В НЕМ ...

- 1) указано, сколько раз встречается каждая варианта
- 2) имеются варианты из двух разнородных совокупностей
- 3) имеются варианты из трех и более разнородных совокупностей
- 4) каждая варианта встречается только один раз

Решение: вариационный ряд называется сгруппированным, если в нем указано, сколько раз встречается каждая варианта.

Правильный ответ: 1.

118. ВАРИАЦИОННЫЙ РЯД НАЗЫВАЕТСЯ РАНЖИРОВАННЫМ, ЕСЛИ:

- 1) указано, сколько раз встречается каждая варианта
- 2) варианты расположены в порядке возрастания
- 3) варианты расположены в порядке убывания
- 4) каждая варианта встречается только один раз

Решение: вариационный ряд называется ранжированным, если в нем варианты расположены в порядке возрастания или убывания.

Правильный ответ: 2, 3.

119. ВАРИАЦИОННЫЙ РЯД НАЗЫВАЕТСЯ ДИСКРЕТНЫМ, ЕСЛИ В НЕМ ...

- 1) указано, сколько раз встречается каждая варианта
- 2) варианты расположены в порядке возрастания
- 3) варианты представлены только целыми числами
- 4) варианты сгруппированы по величине

Решение: вариационный ряд называется дискретным, если в нем варианты представлены только целыми числами.

Правильный ответ: 3.

120. ВАРИАЦИОННЫЙ РЯД НАЗЫВАЕТСЯ НЕПРЕРЫВНЫМ, ЕСЛИ В НЕМ ...

- 1) указано, сколько раз встречается каждая варианта
- 2) варианты расположены в порядке возрастания
- 3) варианты представлены только целыми числами
- 4) варианты представлены любыми числовыми значениями
- 5) варианты сгруппированы по величине

Решение: вариационный ряд называется непрерывным, если в нем варианты представлены любыми числовыми значениями.

Правильный ответ: 4.

121. ОБЩИМИ ХАРАКТЕРИСТИКАМИ ЗНАЧЕНИЙ ВАРИАНТ ВАРИАЦИОННОГО РЯДА ЯВЛЯЮТСЯ:

- 1) средняя арифметическая
- 2) мода
- 3) медиана

- 4) амплитуда
- 5) коэффициент вариации

Решение: общими характеристиками значений вариант вариационного ряда являются: средняя арифметическая, мода и медиана.

Правильный ответ: 1, 2, 3.

122. СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЕЛИЧИНА ВАРИАЦИОННОГО РЯДА – ЭТО ...

- 1) максимальная величина признака в вариационном ряду, характеризующая наивысший уровень явления в статистической совокупности
- 2) общая количественная характеристика изучаемого признака в вариационном ряду, рассчитанная в качественно однородной статистической совокупности
- 3) общая характеристика всех признаков статистической совокупности

Решение: средняя арифметическая величина вариационного ряда – это общая количественная характеристика изучаемого признака в вариационном ряду, рассчитанная в качественно однородной статистической совокупности.

Правильный ответ: 2.

123. СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЕЛИЧИНА ВАРИАЦИОННОГО РЯДА ПОКАЗЫВАЕТ ...

- 1) частоту явления
- 2) структуру явления
- 3) обобщенную количественную характеристика изучаемого признака

Решение: средняя арифметическая величина вариационного ряда показывает обобщенную количественную характеристика изучаемого признака.

Правильный ответ: 3.

124. СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЫЧИСЛЯЕТСЯ СЛЕДУЮЩИМИ СПОСОБАМИ:

- 1) простым
- 2) взвешенным
- 3) способом моментов
- 4) дискретным
- 5) непрерывным
- 6) сгруппированным

Решение: средняя арифметическая вычисляется следующими способами: простым, взвешенным, способом моментов.

Правильный ответ: 1, 2, 3.

125. СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ОБОЗНАЧАЕТСЯ ЗНАКОМ:

- 1) σ
- 2) n
- 3) M

- 4) m
- 5) r
- 6) t

Решение: средняя арифметическая обозначается знаком « M ».

Правильный ответ: 3.

126. ВАРИАНТЫ В ВАРИАЦИОННОМ РЯДУ ОБОЗНАЧАЮТСЯ ЗНАКОМ ...

- 1) σ
- 2) n
- 3) M
- 4) m
- 5) V
- 6) t

Решение: варианты в вариационном ряду обозначаются знаком « V ».

Правильный ответ: 5.

127. ЧАСТОТА ВСТРЕЧАЕМОСТИ ВАРИАНТЫ В ВАРИАЦИОННОМ РЯДУ ОБОЗНАЧАЕТСЯ ЗНАКОМ ...

- 1) σ
- 2) n
- 3) M
- 4) p
- 5) V
- 6) t

Решение: частота встречаемости варианты в вариационном ряду обозначается знаком « p ».

Правильный ответ: 5.

128. ЧИСЛО НАБЛЮДЕНИЙ В ВАРИАЦИОННОМ РЯДУ ОБОЗНАЧАЕТСЯ ЗНАКОМ ...

- 1) σ
- 2) n
- 3) M
- 4) m
- 5) p
- 6) t

Решение: число наблюдений в вариационном ряду обозначается знаком « n ».

Правильный ответ: 2.

129. ФОРМУЛА ВЫЧИСЛЕНИЯ ПРОСТОЙ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ ...

1) $M = \frac{\sum V}{n}$

$$2) \sigma = \sqrt{\frac{\sum d^2}{n}}$$

$$3) m = \frac{\sigma}{\sqrt{n-1}}$$

$$4) t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: простая средняя арифметическая вычисляется по формуле: $M = \frac{\sum v}{n}$.

Правильный ответ: 1.

130. ФОРМУЛА ВЫЧИСЛЕНИЯ ВЗВЕШЕННОЙ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ ...

$$1) M = \frac{\sum vp}{n}$$

$$2) M = \frac{\sum v}{n}$$

$$3) M = A + \frac{\sum dp}{n}$$

Решение: взвешенная средняя арифметическая вычисляется по формуле:

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

Правильный ответ: 1.

131. ОСНОВНЫМИ СВОЙСТВАМИ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ ВЕЛИЧИНЫ ЯВЛЯЮТСЯ:

- 1) сумма всех отклонений от средней равна 0
- 2) при умножении (делении) всех вариантов на один и тот же множитель (делитель) средняя арифметическая умножается (делится) на тот же множитель (делитель)
- 3) если прибавить (вычесть) ко всем вариантам одно и то же число, средняя арифметическая увеличивается (уменьшается) на то же число
- 4) может быть только целым числом
- 5) может быть только положительным числом

Решение: основными свойствами средней арифметической величины являются: сумма всех отклонений от средней равна 0; при умножении (делении) всех вариантов на один и тот же множитель (делитель) средняя арифметическая умножается (делится) на тот же множитель (делитель); если прибавить (вычесть) ко всем вариантам одно и то же число, средняя арифметическая увеличивается (уменьшается) на то же число.

Правильный ответ: 1, 2, 3.

132. ВЕЛИЧИНА, НАИБОЛЕЕ ЧАСТО ВСТРЕЧАЮЩАЯСЯ В ВАРИАЦИОННОМ РЯДУ, НАЗЫВАЕТСЯ ...

- 1) медианой (Me)
- 2) средней арифметической (M)
- 3) модой (Mo)
- 4) средним квадратическим отклонением

Решение: величина, наиболее часто встречающаяся в вариационном ряду, называется медианой (Mo).

Правильный ответ: 3.

133. МЕДИАНА ВАРИАЦИОННОГО РЯДА (Me) – ЭТО ...

- 1) наибольшая по значению варианта вариационного ряда
- 2) варианта, встречающаяся чаще других в вариационном ряду
- 3) центральная варианта, делящая ранжированный вариационный ряд пополам

Решение: медиана вариационного ряда (Me) – это центральная варианта, делящая ранжированный вариационный ряд пополам.

Правильный ответ: 3.

134. СРЕДНЯЯ ВЕЛИЧИНА ВАРИАЦИОННОГО РЯДА, КОТОРАЯ ОПРЕДЕЛЯЕТСЯ КАК НАИБОЛЕЕ ЧАСТО ВСТРЕЧАЮЩАЯСЯ ВАРИАНТА, НАЗЫВАЕТСЯ _____ (ВПИСАТЬ).

Решение: средняя величина вариационного ряда, которая определяется как наиболее часто встречающаяся варианта, называется «Мода».

Правильный ответ: мода.

135. СРЕДНЯЯ ВЕЛИЧИНА ВАРИАЦИОННОГО РЯДА, КОТОРАЯ ОПРЕДЕЛЯЕТСЯ КАК ВАРИАНТА, ЗАНИМАЮЩАЯ СРЕДИННОЕ ПОЛОЖЕНИЕ В РАНЖИРОВАННОМ ВАРИАЦИОННОМ РЯДУ, НАЗЫВАЕТСЯ _____ (ВПИСАТЬ).

Решение: средняя величина вариационного ряда, которая определяется как варианта, занимающая срединное положение в ранжированном вариационном ряду, называется «Медиана».

Правильный ответ: медиана.

136. АМПЛИТУДА ВАРИАЦИОННОГО РЯДА ОБОЗНАЧАЕТСЯ ЗНАКОМ ...

- 1) M
- 2) m
- 3) A_m
- 4) r
- 5) t

б) σ **Решение:** амплитуда вариационного ряда обозначается знаком « A_m ».**Правильный ответ:** 3.

137. АМПЛИТУДА В ВАРИАЦИОННОМ РЯДУ ВЫЧИСЛЯЕТСЯ ПО ФОРМУЛЕ

...

1) $m = \sqrt{\frac{Pq}{n}}$

2) $A_m = V_{max} - V_{min}$

3) $C = \frac{\sigma}{M} * 100\%$

Решение: амплитуда вариационного ряда вычисляется по формуле:

$$A_m = V_{max} - V_{min} .$$

Правильный ответ: 2.

138. СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ ОБОЗНАЧАЕТСЯ ЗНАКОМ

...

1) M 2) m 3) A_m 4) r 5) t 6) σ **Решение:** среднее квадратическое отклонение обозначается знаком « σ ».**Правильный ответ:** 6.

139. СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ ХАРАКТЕРИЗУЕТ...

1) возможные отклонения параметров генеральной совокупности по сравнению с выборочной совокупностью

2) разницу между вариантой и средней арифметической

3) среднее отклонение всех вариантов вариационного ряда от средней арифметической

Решение: среднее квадратическое отклонение характеризует среднее отклонение всех вариантов вариационного ряда от средней арифметической.**Правильный ответ:** 3.

140. СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ МОЖЕТ РАССЧИТЫВАТЬСЯ СЛЕДУЮЩИМИ СПОСОБАМИ:

1) простым

2) как средневзвешенное

3) способом моментов

4) прямым

5) обратным

Решение: среднее квадратическое отклонение может рассчитываться следующими способами: простым; как средневзвешенное; способом моментов.

Правильный ответ: 1, 2, 3.

141. СТАТИСТИЧЕСКИЙ КРИТЕРИЙ, КОТОРЫМ СРАВНИВАЮТ РАЗНООБРАЗИЕ ВАРИАЦИОННЫХ РЯДОВ, ЕСЛИ ЕДИНИЦЫ ИЗМЕРЕНИЯ ВАРИАНТ В НИХ РАЗЛИЧНЫ, НАЗЫВАЕТСЯ...

- 1) ошибка репрезентативности
- 2) среднее квадратическое отклонение
- 3) коэффициент вариации
- 4) критерий Стьюдента
- 5) коэффициент корреляции

Решение: статистический критерий, которым сравнивают разнообразие вариационных рядов, если единицы измерения вариант в них различны, называется «Коэффициент вариации».

Правильный ответ: 3.

142. ФОРМУЛА ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА ВАРИАЦИИ ...

- 1) $m = \frac{\sigma}{\sqrt{n-1}}$
- 2) $C = \frac{\sigma}{M} * 100\%$
- 3) $t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$

Решение: формула вычисления коэффициента вариации: $C = \frac{\sigma}{M} * 100\%$.

Правильный ответ: 2.

143. СТЕПЕНЬ РАССЕЙНИЯ ВАРИАНТ ВОКРУГ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ, ЕСЛИ ЗНАЧЕНИЕ КОЭФФИЦИЕНТА ВАРИАЦИИ ДО 10%, ЯВЛЯЕТСЯ ...

- 1) малой
- 2) средней
- 3) сильной

Решение: степень рассеяния вариант вокруг средней арифметической, если значение коэффициента вариации до 10%, является малой.

Правильный ответ: 1.

144. СТЕПЕНЬ РАССЕЯНИЯ ВАРИАНТ ВОКРУГ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ, ЕСЛИ ЗНАЧЕНИЕ КОЭФФИЦИЕНТА ВАРИАЦИИ ОТ 10 ДО 20%, ЯВЛЯЕТСЯ ...
- 1) малой
 - 2) средней
 - 3) сильной

Решение: степень рассеяния вариант вокруг средней арифметической, если значение коэффициента вариации от 10 до 20%, является средней.

Правильный ответ: 2.

145. СТЕПЕНЬ РАССЕЯНИЯ ВАРИАНТ ВОКРУГ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ, ЕСЛИ ЗНАЧЕНИЕ КОЭФФИЦИЕНТА ВАРИАЦИИ БОЛЕЕ 20%, ЯВЛЯЕТСЯ ...
- 1) малой
 - 2) средней
 - 3) сильной

Решение: степень рассеяния вариант вокруг средней арифметической, если значение коэффициента вариации более 20%, является сильной.

Правильный ответ: 3.

146. «ПРАВИЛО ТРЕХ СИГМ» - ЭТО...
- 1) правило отбора единиц наблюдения в статистическую совокупность
 - 2) вероятностная зависимость между значением средней арифметической, средним квадратическим отклонением и вариантами
 - 3) отношение средней величины к среднему квадратическому отклонению

Решение: «Правило трех сигм» - это вероятностная зависимость между значением средней арифметической, средним квадратическим отклонением и вариантами.

Правильный ответ: 2.

147. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ – ЭТО ...
- 1) размах колебаний средней арифметической изучаемого признака, который можно оценить статистическими методами
 - 2) разница между максимальной и минимальной вариантами
 - 3) разница между средним квадратическим отклонением и ошибкой репрезентативности
 - 4) отношение средней величины к среднему квадратическому отклонению

Решение: доверительный интервал – это размах колебаний средней арифметической изучаемого признака, который можно оценить статистическими методами.

Правильный ответ: 1.

148. СВОЙСТВО РЕПРЕЗЕНТАТИВНОСТИ ХАРАКТЕРНО ДЛЯ СТАТИСТИЧЕСКОЙ СОВОКУПНОСТИ ...

- 1) генеральной
- 2) выборочной

Решение: свойство репрезентативности характерно для выборочной статистической совокупности.

Правильный ответ: 2.

149. РЕПРЕЗЕНТАТИВНОСТЬ – ЭТО ...

- 1) соответствие средней арифметической величины выборочной совокупности аналогичному параметру генеральной совокупности
- 2) понятие, характеризующее связь между признаками
- 3) характеристика методики исследования

Решение: репрезентативность – это соответствие средней арифметической величины выборочной совокупности аналогичному параметру генеральной совокупности.

Правильный ответ: 1.

150. С УВЕЛИЧЕНИЕМ ОБЪЕМА НАБЛЮДЕНИЙ ОШИБКА РЕПРЕЗЕНТАТИВНОСТИ ...

- 1) увеличивается
- 2) остается без изменений
- 3) уменьшается

Решение: с увеличением объема наблюдений ошибка репрезентативности уменьшается.

Правильный ответ: 3.

151. ОШИБКА РЕПРЕЗЕНТАТИВНОСТИ:

- 1) является мерой изменчивости значения средней арифметической, которая может быть получена при повторных исследованиях
- 2) позволяет с определенной вероятностью перенести результаты изучения признака в выборочной совокупности на генеральную совокупность
- 3) служит оценкой рассеяния вариант

Решение: ошибка репрезентативности: является мерой изменчивости значения средней арифметической, которая может быть получена при повторных исследованиях и позволяет с определенной вероятностью перенести результаты изучения признака в выборочной совокупности на генеральную совокупность.

Правильный ответ: 1, 2.

152. СТЕПЕНЬ СОВПАДЕНИЯ ВЫБОРОЧНОЙ СОВОКУПНОСТИ С ГЕНЕРАЛЬНОЙ ОЦЕНИВАЕТСЯ С ПОМОЩЬЮ ...

- 1) коэффициента корреляции
- 2) среднего квадратического отклонения

3) ошибки репрезентативности

Решение: степень совпадения выборочной совокупности с генеральной оценивается с помощью ошибки репрезентативности.

Правильный ответ: 3.

153. РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРОЧНОЙ СОВОКУПНОСТИ ОБЕСПЕЧИВАЕТСЯ:

1) соответствующим числом единиц наблюдения

2) стандартизацией

3) случайностью отбора единиц наблюдения

4) кодированием единиц наблюдения

Решение: репрезентативность выборочной совокупности обеспечивается соответствующим числом единиц наблюдения, а также случайностью отбора единиц наблюдения.

Правильный ответ: 1, 3.

154. ФОРМУЛА ВЫЧИСЛЕНИЯ СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ ...

$$1) M = \frac{\sum v}{n}$$

$$2) \sigma = \sqrt{\frac{\sum d^2}{n}}$$

$$3) m = \frac{\sigma}{\sqrt{n-1}}$$

$$4) t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: формула вычисления среднего квадратического отклонения: $\sigma = \sqrt{\frac{\sum d^2}{n}}$.

Правильный ответ: 2.

155. ФОРМУЛА ВЫЧИСЛЕНИЯ ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ В ВАРИАЦИОННОМ РЯДУ ...

$$1) m = \frac{\sigma}{\sqrt{n-1}}$$

$$2) m = \sqrt{\frac{Pq}{n}}$$

$$3) C = \frac{\sigma}{M} * 100\%$$

$$4) t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: формула вычисления ошибки репрезентативности в вариационном ряду: $m = \frac{\sigma}{\sqrt{n-1}}$.

Правильный ответ: 1.

156. ФОРМУЛА ВЫЧИСЛЕНИЯ ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ ДЛЯ ОТНОСИТЕЛЬНЫХ ВЕЛИЧИН ...

$$1) m = \frac{\sigma}{\sqrt{n-1}}$$

$$2) m = \sqrt{\frac{Pq}{n}}$$

$$3) C = \frac{\sigma}{M} * 100\%$$

$$4) t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: формула вычисления ошибки репрезентативности для относительных величин: $m = \sqrt{\frac{Pq}{n}}$.

Правильный ответ: 2.

157. ВЕЛИЧИНА ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ ЗАВИСИТ ОТ:

- 1) числа наблюдений
- 2) среднего квадратического отклонения
- 3) коэффициента корреляции
- 4) критерия Стьюдента

Решение: величина ошибки репрезентативности зависит от среднего квадратического отклонения и числа наблюдений.

Правильный ответ: 1, 2.

158. ПРИ ПРОВЕДЕНИИ МЕДИКО-СОЦИАЛЬНЫХ ИССЛЕДОВАНИЙ В ВЫБОРОЧНОЙ СОВОКУПНОСТИ НЕВОЗМОЖНО ИЗБЕЖАТЬ ...

- 1) арифметических ошибок вычислений

- 2) ошибок степени точности вычисления (округления чисел)
- 3) методических ошибок организации исследования
- 4) ошибок репрезентативности

Решение: при проведении медико-социальных исследований в выборочной совокупности невозможно избежать ошибок репрезентативности.

Правильный ответ: 4.

159. ЧТОБЫ УМЕНЬШИТЬ ОШИБКУ РЕПРЕЗЕНТАТИВНОСТИ, ЧИСЛО НАБЛЮДЕНИЙ НЕОБХОДИМО ...

- 1) увеличить
- 2) уменьшить
- 3) оставить без изменения

Решение: чтобы уменьшить ошибку репрезентативности, число наблюдений необходимо увеличить.

Правильный ответ: 1.

160. ЧЕМ МЕНЬШЕ ЧИСЛО НАБЛЮДЕНИЙ, ТЕМ ВЕЛИЧИНА ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ ...

- 1) меньше
- 2) больше

Решение: чем меньше число наблюдений, тем величина ошибки репрезентативности больше.

Правильный ответ: 2.

161. ЧЕМ МЕНЬШЕ ВАРИАБЕЛЬНОСТЬ ПРИЗНАКА, ТЕМ ВЕЛИЧИНА СРЕДНЕЙ ОШИБКИ ...

- 1) меньше
- 2) больше

Решение: чем меньше вариабельность признака, тем величина средней ошибки меньше.

Правильный ответ: 1.

162. ДОВЕРИТЕЛЬНЫЕ ГРАНИЦЫ - ЭТО ...

- 1) вероятностная оценка возможных отклонений, в пределах которых может колебаться искомая средняя величина признака при повторных исследованиях
- 2) разница между максимальной и минимальной вариантами
- 3) разница между средним квадратическим отклонением и ошибкой репрезентативности
- 4) отношение средней величины к среднему квадратическому отклонению

Решение: доверительные границы - это вероятностная оценка возможных отклонений, в пределах которых может колебаться искомая средняя величина признака при повторных исследованиях.

Правильный ответ: 1.

163. ДОВЕРИТЕЛЬНЫЕ ГРАНИЦЫ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ С ВЕРОЯТНОСТЬЮ, ДОСТАТОЧНОЙ ДЛЯ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ, ВЫЧИСЛЯЮТСЯ ПО ФОРМУЛЕ...

- 1) $M \pm 1m$, вероятность 68,3%
- 2) $M \pm 2m$, вероятность 95,5%
- 3) $M \pm 3m$, вероятность 99,9%

Решение: доверительные границы средней арифметической с вероятностью 95,5%, достаточной для медико-биологических исследований, вычисляются по формуле $M \pm 2m$.

Правильный ответ: 2.

164. ВЕРОЯТНОСТЬ БЕЗОШИБОЧНОГО ПРОГНОЗА ДЛЯ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ СОСТАВЛЯЕТ ...

- 1) 68%
- 2) 90%
- 3) 92%
- 4) 95%

Решение: вероятность безошибочного прогноза для медико-биологических исследований составляет 95%.

Правильный ответ: 4.

165. ЕСЛИ ИЗВЕСТНО, ЧТО $M = 40,0\text{КГ}$, $\sigma = 3,0\text{КГ}$, ТО КРАЙНИЕ ВАРИАНТЫ ВАРИАЦИОННОГО РЯДА, С ВЕРОЯТНОСТЬЮ ПРОГНОЗА 99,7%, БУДУТ НАХОДИТЬСЯ В ДИАПАЗОНЕ

- 1) 37 - 43кг
- 2) 31 - 49кг
- 3) 39 - 42кг

Решение: рассеяние вариант вокруг средней арифметической с вероятностью 99,7% определяется в интервале $M \pm 3\sigma$, если известно, что $M=40,0\text{кг}$, а $\sigma=3,0\text{кг}$, то крайние варианты вариационного ряда, с вероятностью прогноза 99,7%, будут находиться в диапазоне от 31 до 49кг.

Правильный ответ: 2.

166. ГРАНИЦЫ ВОЗМОЖНЫХ СЛУЧАЙНЫХ КОЛЕБАНИЙ СРЕДНЕЙ ВЕЛИЧИНЫ В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ С ВЕРОЯТНОСТЬЮ ПРОГНОЗА 95,5 % ЛЕЖАТ В ПРЕДЕЛАХ ...

- 1) $M \pm m$
- 2) $M \pm 2m$
- 3) $M \pm 3m$

Решение: границы возможных случайных колебаний средней величины в генеральной совокупности с вероятностью прогноза 95,5 % лежат в пределах $M \pm 2m$.

Правильный ответ: 2.

167. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ $M \pm 3m$ СООТВЕТСТВУЕТ ВЕРОЯТНОСТИ ПРОГНОЗА ...

- 1) 68,3%
- 2) 95,5%
- 3) 99,7%
- 4) 100%

Решение: доверительный интервал $M \pm 3m$ соответствует вероятности прогноза 99,7%.

Правильный ответ: 3.

168. ДЛЯ ПРИЗНАКОВ, ПОДЧИНЯЮЩИХСЯ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ, ДОСТОВЕРНОСТЬ РАЗЛИЧИЯ ДВУХ СРЕДНИХ ВЕЛИЧИН ОПРЕДЕЛЯЕТСЯ С ПОМОЩЬЮ ...

- 1) ошибки репрезентативности
- 2) коэффициента вариации
- 3) средней арифметической
- 4) критерия Стьюдента

Решение: для признаков, подчиняющихся нормальному закону распределения, достоверность различия двух средних величин определяется с помощью критерия Стьюдента.

Правильный ответ: 4.

169. СТАТИСТИЧЕСКАЯ ДОСТОВЕРНОСТЬ РАЗЛИЧИЯ МЕЖДУ ДВУМЯ СРЕДНИМИ ВЕЛИЧИНАМИ ОПРЕДЕЛЯЕТСЯ С ПОМОЩЬЮ ...

- 1) ошибки репрезентативности
- 2) коэффициента вариации
- 3) среднего квадратического отклонения
- 4) критерия Стьюдента

Решение: статистическая достоверность различия между двумя средними величинами определяется с помощью критерия Стьюдента.

Правильный ответ: 4.

170. КРИТЕРИЙ ДОСТОВЕРНОСТИ СТЬЮДЕНТА УКАЗЫВАЕТ ...

- 1) во сколько раз разность сравниваемых средних величин превышает их среднюю ошибку
- 2) во сколько раз среднее квадратическое отклонение меньше средней арифметической

3) на ошибку разности средних величин в генеральной и выборочной совокупности

Решение: критерий достоверности Стьюдента указывает во сколько раз разность сравниваемых средних величин превышает их среднюю ошибку.

Правильный ответ: 1.

171. ВЕЛИЧИНА КРИТЕРИЯ ДОСТОВЕРНОСТИ СТЬЮДЕНТА ПРИ СРАВНЕНИИ СРЕДНИХ ВЕЛИЧИН ОПРЕДЕЛЯЕТСЯ ПО ФОРМУЛЕ ...

$$1) m = \frac{\sigma}{\sqrt{n-1}}$$

$$2) m = \sqrt{\frac{Pq}{n}}$$

$$3) C = \frac{\sigma}{M} * 100\%$$

$$4) t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: величина критерия достоверности Стьюдента при сравнении средних

величин определяется по формуле: $t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$.

Правильный ответ: 4.

172. РАСЧЕТ КРИТЕРИЯ ДОСТОВЕРНОСТИ СТЬЮДЕНТА ПРИ СРАВНЕНИИ ОТНОСИТЕЛЬНЫХ ВЕЛИЧИН ПРОИЗВОДИТСЯ ПО ФОРМУЛЕ ...

$$1) m = \frac{\sigma}{\sqrt{n-1}}$$

$$2) m = \sqrt{\frac{Pq}{n}}$$

$$3) C = \frac{\sigma}{M} * 100\%$$

$$4) t = \left| \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: расчет критерия достоверности Стьюдента при сравнении

относительных величин производится по формуле: $t = \left| \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \right|$.

Правильный ответ: 4.

173. ФОРМУЛА ВЫЧИСЛЕНИЯ КРИТЕРИЯ ДОСТОВЕРНОСТИ СТЬЮДЕНТА ...

$$1) m = \frac{\sigma}{\sqrt{n-1}}$$

$$2) m = \sqrt{\frac{Pq}{n}}$$

$$3) C = \frac{\sigma}{M} * 100\%$$

$$4) t = \left| \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \right|$$

Решение: формула вычисления критерия достоверности Стьюдента:

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

Правильный ответ: 4.

174. ВЫВОД, КОТОРЫЙ МОЖНО СДЕЛАТЬ О НАЛИЧИИ СТАТИСТИЧЕСКОЙ ДОСТОВЕРНОСТИ РАЗЛИЧИЙ МЕЖДУ ДВУМЯ ВЫБОРОЧНЫМИ СОВОКУПНОСТЯМИ, ЕСЛИ КРИТЕРИЙ ДОСТОВЕРНОСТИ СТЬЮДЕНТА ПРЕВЫШАЕТ ЗНАЧЕНИЕ 2 (ПРИ $n > 30$) ...

- 1) выявленные различия случайны
- 2) выявленные различия не случайны
- 3) различий нет

Решение: если при сравнении двух выборочных совокупностей критерий достоверности Стьюдента превышает значение 2 (при $n > 30$), можно сделать вывод о достоверном различии между ними с вероятностью 95,5%, т.е. выявленные различия не случайны.

Правильный ответ: 2.

175. РАЗЛИЧИЯ МЕЖДУ СРЕДНИМИ ВЕЛИЧИНАМИ ДВУХ ПРИЗНАКОВ ЯВЛЯЮТСЯ СУЩЕСТВЕННЫМИ, ЕСЛИ ВЕЛИЧИНА КРИТЕРИЯ СТЬЮДЕНТА (t) БОЛЬШЕ ИЛИ РАВНА:

- 1) 0,5
- 2) 1,0
- 3) 1,5
- 4) 2,0

5) 2,5

Решение: различия между средними величинами двух признаков являются существенными, если величина критерия Стьюдента (t) равна или превышает значение 2 (при $n > 30$).

Правильный ответ: 4, 5.

176. ПРИ ЗНАЧЕНИИ t -КРИТЕРИЯ (СТЬЮДЕНТА) БОЛЬШЕ ИЛИ РАВНОМ 2 РАЗЛИЧИЯ ДВУХ СРЕДНИХ ВЕЛИЧИН ...

- 1) достоверны
- 2) не достоверны
- 3) однородны
- 4) независимы

Решение: если при сравнении двух выборочных совокупностей критерий достоверности Стьюдента превышает значение 2 (при $n > 30$), можно сделать вывод о достоверном различии между ними с вероятностью 95,5%.

Правильный ответ: 1.

177. СОБЫТИЯ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ СЧИТАЮТСЯ НЕ ДОСТОВЕРНЫМИ, ЕСЛИ ИХ ВЕРОЯТНОСТЬ:

- 1) 68,3%
- 2) 95,5%
- 3) 99,7%
- 4) 50,0%

Решение: событие в медико-биологических исследованиях считается не достоверным, если вероятность его прогноза меньше 95,5%.

Правильный ответ: 1, 4.

178. ПОНЯТИЕ «НЕОДНОРОДНОСТЬ СТАТИСТИЧЕСКИХ СОВОКУПНОСТЕЙ» ОЗНАЧАЕТ ...

- 1) отсутствие взаимосвязи между признаками
- 2) отсутствие упорядочения вариационных рядов
- 3) различие между совокупностями по характеризующим признакам, влияющим на изучаемый признак
- 4) различие между совокупностями по изучаемым признакам

Решение: понятие «Неоднородность статистических совокупностей» означает различие между совокупностями по характеризующим признакам, влияющим на изучаемый признак.

Правильный ответ: 3.

179. ДЛЯ СРАВНЕНИЯ ПОКАЗАТЕЛЕЙ, ПОЛУЧЕННЫХ НА НЕОДНОРОДНЫХ ПО СВОЕМУ СОСТАВУ СОВОКУПНОСТЯХ, ИСПОЛЬЗУЕТСЯ МЕТОД ...

- 1) корреляции

- 2) выравнивания динамических рядов
- 3) стандартизации
- 4) экстраполяции

Решение: для сравнения показателей, полученных на неоднородных по своему составу совокупностях, используется метод стандартизации.

Правильный ответ: 3.

180. ТИПЫ ВЗАИМОСВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ, КОТОРЫЕ МОЖНО УСТАНОВИТЬ МАТЕМАТИЧЕСКИМИ МЕТОДАМИ:

- 1) корреляционная
- 2) механическая
- 3) функциональная
- 4) косвенная

Решение: типами взаимосвязи между явлениями, которые можно установить математическими методами, являются функциональная и корреляционная.

Правильный ответ: 1, 3.

181. СТРОГАЯ ЗАВИСИМОСТЬ ПРОЦЕССОВ ИЛИ ЯВЛЕНИЙ, ВЫРАЖЕННАЯ МАТЕМАТИЧЕСКОЙ ФОРМУЛОЙ, НАЗЫВАЕТСЯ...

- 1) корреляционная
- 2) стандартизованная
- 3) функциональная
- 4) регрессионная

Решение: строгая зависимость процессов или явлений, выраженная математической формулой, называется «Функциональная связь».

Правильный ответ: 3.

182. СТАТИСТИЧЕСКИЙ АНАЛИЗ, КОТОРЫЙ ИСПОЛЬЗУЕТСЯ ДЛЯ ВЫЯВЛЕНИЯ ВЕРОЯТНОСТНОЙ ВЗАИМОСВЯЗИ МЕЖДУ ПРИЗНАКАМИ В СОЦИАЛЬНО-ГИГИЕНИЧЕСКИХ И МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ, НАЗЫВАЕТСЯ...

- 1) корреляция
- 2) аппроксимация
- 3) стандартизация
- 4) дисперсия

Решение: для выявления вероятностной взаимосвязи между признаками в социально-гигиенических и медико-биологических исследованиях используется корреляционный анализ.

Правильный ответ: 1.

183. КОРРЕЛЯЦИОННАЯ ВЗАИМОСВЯЗЬ МЕЖДУ ИЗУЧАЕМЫМИ ПРИЗНАКАМИ МОЖЕТ БЫТЬ ОБНАРУЖЕНА:

- 1) с помощью статистической таблицы

- 2) с помощью графика
- 3) расчетом коэффициента корреляции
- 4) применением метода стандартизации

Решение: корреляционная взаимосвязь между изучаемыми признаками может быть обнаружена с помощью статистической таблицы, графика и коэффициента корреляции.

Правильный ответ: 1, 2, 3.

184. О НАЛИЧИИ СИЛЬНОЙ ПРЯМОЙ КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТИ МОЖНО ГОВОРИТЬ ПРИ ЗНАЧЕНИИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ...

- 1) больше или равно $-0,6$
- 2) меньше или равно $0,1$
- 3) больше $0,7$
- 4) меньше или равно $0,4$

Решение: о наличии сильной прямой корреляционной зависимости можно говорить при значении коэффициента корреляции больше $0,7$.

Правильный ответ: 3.

185. О НАЛИЧИИ СИЛЬНОЙ ОБРАТНОЙ КОРРЕЛЯЦИОННОЙ СВЯЗИ МЕЖДУ ПРИЗНАКАМИ МОЖНО ГОВОРИТЬ ПРИ ЗНАЧЕНИИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ...

- 1) от 0 до $0,3$
- 2) от $0,3$ до $0,7$
- 3) от $0,7$ до $1,0$
- 4) от 0 до $-0,3$
- 5) от $-0,3$ до $-0,7$
- 6) от $-0,7$ до $-1,0$

Решение: о наличии сильной обратной корреляционной связи между признаками можно говорить при значении коэффициента корреляции от $-0,7$ до $-1,0$.

Правильный ответ: 6.

186. ЗНАЧЕНИЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ДЛЯ ПРЯМОЙ ВЗАИМОСВЯЗИ МЕЖДУ ПРИЗНАКАМИ ...

- 1) положительные
- 2) отрицательные
- 3) любые
- 4) меньше или равны -1

Решение: значения коэффициента корреляции для прямой взаимосвязи между признаками являются положительными.

Правильный ответ: 1.

187. ЗНАЧЕНИЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ДЛЯ ОБРАТНОЙ ВЗАИМОСВЯЗИ МЕЖДУ ПРИЗНАКАМИ ...

- 1) положительные
- 2) отрицательные
- 3) любые
- 4) меньше или равно 1

Решение: значения коэффициента корреляции для обратной взаимосвязи между признаками являются отрицательными.

Правильный ответ: 2.

188. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ МОЖЕТ ИЗМЕНЯТЬСЯ В ПРЕДЕЛАХ (ПО МОДУЛЮ ЧИСЛА) ...

- 1) от 1 до 10
- 2) от 0 до 1
- 3) от 0 до 100
- 4) могут встречаться любые значения

Решение: коэффициент корреляции может изменяться в пределах (по модулю числа) от 0 до 1.

Правильный ответ: 2.

189. МИНИМАЛЬНОЕ АБСОЛЮТНОЕ ЗНАЧЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ВЫРАЖАЕТСЯ ЗНАЧЕНИЕМ ____ (ВНЕСИТЕ ЦИФРУ).

Решение: коэффициент корреляции может изменяться в пределах (по модулю числа) от 0 до 1.

Правильный ответ: 0.

190. МАКСИМАЛЬНОЕ ЗНАЧЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ (ПО МОДУЛЮ) ВЫРАЖАЕТСЯ ЗНАЧЕНИЕМ ____ (ВНЕСИТЕ ЦИФРУ).

Решение: коэффициент корреляции может изменяться в пределах (по модулю числа) от 0 до 1.

Правильный ответ: 1.

191. ВЫЗОВ МОДУЛЯ ВЫЧИСЛЕНИЯ СРЕДНИХ ВЕЛИЧИН И ХАРАКТЕРИСТИК ВАРИАЦИОННОГО РЯДА В ПРОГРАММЕ EXCEL ПРОИЗВОДИТСЯ КОМАНДОЙ ...

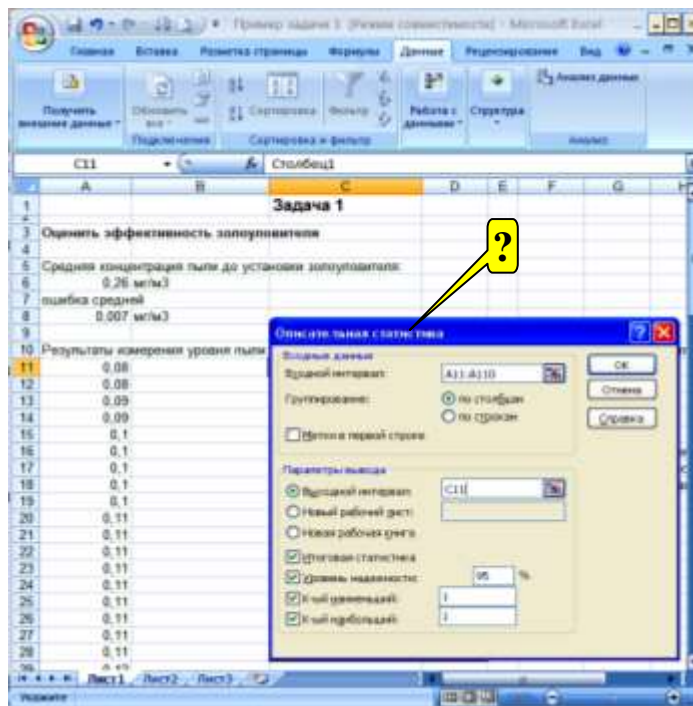
- 1) «Сервис (Данные)» → «Обработка данных» → «Поиск решения»
- 2) «Сервис (Данные)» → «Анализ данных» → «Описательная статистика»
- 3) «Сервис (Данные)» → «Обработка данных» → «Основная статистика»

Решение: вызов модуля вычисления средних величин и характеристик вариационного ряда в программе Excel производится командой «Сервис (Данные)» → «Анализ данных» → «Описательная статистика».

Правильный ответ: 2.

192. РЕЗУЛЬТАТ ОБРАБОТКИ ВАРИАЦИОННОГО РЯДА, ВЫПОЛНЕННЫЙ МОДУЛЕМ «ОПИСАТЕЛЬНАЯ СТАТИСТИКА» ПРОГРАММЫ EXCEL (СМ. РИСУНОК), БУДЕТ ПОМЕЩЕН В СЛЕДУЮЩИЕ ЯЧЕЙКИ ...

- 1) начиная с ячейки A11
- 2) начиная с ячейки C11
- 3) в диапазоне ячеек A11:A110
- 4) в ячейку A110

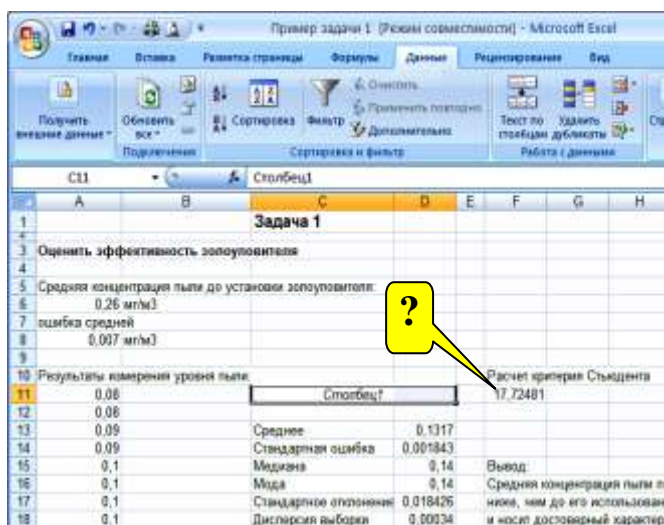


Решение: результат обработки вариационного ряда, выполненный модулем «описательная статистика»

программы Excel размещается в выходном интервале (см.рисунок) начиная с ячейки C11 и ниже.

Правильный ответ: 2.

193. ФОРМУЛОЙ ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА СТЬЮДЕНТА В ПРОГРАММЕ EXCEL (СМ. РИСУНОК) ЯВЛЯЕТСЯ ...



1) $(A6-D13)/\sqrt{A8^2+D14^2}$

2) $(A6+D13)*\sqrt{A8^2+D14^2}$

3) $(A6-D13)/\sqrt{A8^2-D14^2}$

Решение: формулой вычисления коэффициента Стьюдента в программе Excel (см.рисунок) является « $(A6-D13)/\sqrt{A8^2+D14^2}$ ».

Правильный ответ: 1.

194. ОБОЗНАЧЕНИЕ ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ В ТАБЛИЦЕ РЕЗУЛЬТАТОВ ОБРАБОТКИ ВАРИАЦИОННОГО РЯДА ПАКЕТОМ «ОПИСАТЕЛЬНАЯ СТАТИСТИКА» ПРОГРАММЫ EXCEL (СМ. РИСУНОК) СООТВЕТСТВУЕТ ПОЗИЦИИ ...

- 1) 1
- 2) 2
- 3) 3
- 4) 4

Строчка	Среднее	Стандартная ошибка	Медiana	Мода	Стандартное отклонение	Дисперсия выборки	Эксцесс	Асимметричность	Интервал	Минимум	Максимум	Сумма	Счет	Наибольший(1)	Наименьший(1)	Уровень надежности
11	0,08															
12	0,08															
13		0,1317														
14		0,001843														
15		0,14														
16		0,14														
17		0,018426														
18		0,00034														
19		0,300694														
20		-0,62102														
21		0,09														
22		0,08														
23		0,17														
24		13,17														
25		100														
26		0,17														
27		0,08														
28		0,003656														

Решение:

обозначение ошибки репрезентативности в таблице результатов обработки вариационного ряда пакетом «Описательная статистика» программы Excel (см.рисунок) соответствует позиция 2 (Стандартная ошибка).

Правильный ответ: 2.

195. ФОРМУЛОЙ ВЫЧИСЛЕНИЯ ОТНОСИТЕЛЬНОГО ПОКАЗАТЕЛЯ (P) В ПРОГРАММЕ EXCEL (СМ. РИСУНОК) ЯВЛЯЕТСЯ ...

- 1) $=A1*100/D3$
- 2) $=D3*100/A3$
- 3) $=A2*100/D2$
- 4) $=G3*100/A3$

Строчка	Количество наблюдений (n)	Наступило выздоровление	Относительный показатель (P)
2			
3	130	106	81,53846
4			
5	Величина, обратная P	Ошибка репрезентативности (m)	Доверительные границы (P±2m)
6	18,46154	3,402856	74,73275 88,34417
7	Вывод: при лечении препаратом Z могут выздоравливать от 74,7% до 88,3% больных		
8			
9			
10			
11			
12			

Решение: формулой вычисления относительного показателя (P) в программе Excel (см. рисунок) является «=D3*100/A3».

Правильный ответ: 2.

196. ФОРМУЛОЙ ВЫЧИСЛЕНИЯ ВЕЛИЧИНЫ, ОБРАТНОЙ P, В ПРОГРАММЕ EXCEL (СМ. РИСУНОК) ЯВЛЯЕТСЯ ...

1) = 100 - G2

2) = 100 - A5

3) = 100 - G3

4) = 100 - A6

	A	B	C	D	E	F	G	H	I	J	K
1		Задача 2									
2		Количество наблюдений (n)		Наступило выздоровление			Относительный показатель (P)				
3		130		106			81,53846				
4											
5		Величина, обратная P		Ошибка репрезентативности (m)			Доверительные границы (P±2m)				
6		18,46154		3,402856			74,73275		88,34417		
7		Вывод: при лечении препаратом Z									
8		могут выздоравливать от 74,7% до 88,3% больных									
9											
10											
11											
12											

Решение: формулой вычисления величины, обратной P, в программе Excel (см. рисунок) является «= 100 - G3».

Правильный ответ: 3.

197. ФОРМУЛОЙ ВЫЧИСЛЕНИЯ ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ ОТНОСИТЕЛЬНОГО ПОКАЗАТЕЛЯ В ПРОГРАММЕ EXCEL (СМ. РИСУНОК) ЯВЛЯЕТСЯ ...

1) =корень(G3*A6/A3)

2) =корень(D3*A6/A3)

3) =корень(G3*A6*A3)

4) =корень(D3*A6*A3)

	A	B	C	D	E	F	G	H	I	J	
1		Задача 2									
2		Количество наблюдений (n)		Наступило выздоровление			Относительный показатель (P)				
3		130		106			81,53846				
4											
5		Величина, обратная P		Ошибка репрезентативности (m)			Доверительные границы (P±2m)				
6		18,46154		3,402856			74,73275		88,34417		
7		Вывод: при лечении препаратом Z									
8		могут выздоравливать от 74,7% до 88,3% больных									

Решение: формулой вычисления ошибки репрезентативности относительного показателя в программе Excel (см. рисунок) является «=корень(G3*A6/A3)».

Правильный ответ: 1.

198. ФОРМУЛОЙ ВЫЧИСЛЕНИЯ НИЖНЕЙ ДОВЕРИТЕЛЬНОЙ ГРАНИЦЫ В ПРОГРАММЕ EXCEL (СМ. РИСУНОК) ЯВЛЯЕТСЯ ...

- 1) $=G2-2*D5$
- 2) $=G3-2*D6$
- 3) $=G3-2*D6$
- 4) $=G3+2*D6$

The screenshot shows an Excel spreadsheet titled "Пример задачи2 [Режим совместимости] - Microsoft Excel". The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H	I	J	K
1		Задача 2									
2		Количество наблюдений (n)		Наступило выздоровление		Относительный показатель (P)					
3		130		106		81,53846					
4											
5		Величина, обратная P		Ошибка репрезентативности (m)		Доверительные границы (P±2m)					
6		18,46154		3,402856		74,73275		88,34417			
7		Вывод: при лечении препаратом Z									
8		могут выздоравливать от 74,7% до 88,3% больных									
9											
10											
11											
12											

A yellow callout box with a question mark is positioned over cell J6, which contains the value 88,34417.

Решение: формулой вычисления нижней доверительной границы в программе Excel (см. рисунок) является « $=G3-2*D6$ ».

Правильный ответ: 3.

199. ФОРМУЛОЙ ВЫЧИСЛЕНИЯ ВЕРХНЕЙ ДОВЕРИТЕЛЬНОЙ ГРАНИЦЫ В ПРОГРАММЕ EXCEL (СМ. РИСУНОК) ЯВЛЯЕТСЯ ...

- 1) $=G2+2*D5$
- 2) $=G3-2*D6$
- 3) $=G3-2*D6$
- 4) $=G3+2*D6$

The screenshot shows an Excel spreadsheet titled "Пример задачи2 [Режим совместимости] - Microsoft Excel". The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H	I	J	K
1		Задача 2									
2		Количество наблюдений (n)		Наступило выздоровление		Относительный показатель (P)					
3		130		106		81,53846					
4											
5		Величина, обратная P		Ошибка репрезентативности (m)		Доверительные границы (P±2m)					
6		18,46154		3,402856		74,73275		88,34417			
7		Вывод: при лечении препаратом Z									
8		могут выздоравливать от 74,7% до 88,3% больных									
9											
10											
11											
12											

A yellow callout box with a question mark is positioned over cell J6, which contains the value 88,34417.

Решение: формулой вычисления верхней доверительной границы в программе Excel (см. рисунок) является « $=G3+2*D6$ ».

Правильный ответ: 4.

Приложение 1

Критические значения критерия Шапиро-Уилкса (W)

Число наблюдений, n	Критическое значение критерия W для уровня вероятности p	
	$p=0,01$	$p=0,05$
8	0,749	0,818
9	0,764	0,829
10	0,781	0,842
11	0,792	0,850
12	0,805	0,859
13	0,814	0,866
14	0,825	0,874
15	0,835	0,881
16	0,844	0,887
17	0,851	0,892
18	0,858	0,897
19	0,863	0,901
20	0,868	0,905
21	0,873	0,908
22	0,878	0,911
23	0,881	0,914
24	0,884	0,916
25	0,888	0,918
26	0,891	0,920
27	0,894	0,923
28	0,896	0,924
29	0,898	0,926
30	0,900	0,927
31	0,902	0,929
32	0,904	0,930
33	0,906	0,931
34	0,908	0,933
35	0,910	0,934
36	0,912	0,935
37	0,914	0,936
38	0,916	0,938
39	0,917	0,939
40	0,919	0,940
41	0,920	0,941
42	0,922	0,942
43	0,923	0,943
44	0,924	0,944
45	0,926	0,945
46	0,927	0,945
47	0,928	0,946
48	0,929	0,947
49	0,929	0,947
50	0,930	0,947

Приложение 2

Критические значения критерия Стьюдента

Степени свободы, df	Уровень значимости, p				
	0,1	0,05	0,02	0,01	0,001
1	6,3138	12,7062	31,8205	63,6567	636,6192
2	2,9200	4,3027	6,9646	9,9248	31,5991
3	2,3534	3,1824	4,5407	5,8409	12,924
4	2,1318	2,7764	3,7469	4,6041	8,6103
5	2,0150	2,5706	3,3649	4,0321	6,8688
6	1,9432	2,4469	3,1427	3,7074	5,9588
7	1,8946	2,3646	2,9980	3,4995	5,4079
8	1,8595	2,3060	2,8965	3,3554	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,4370
12	1,7823	2,1788	2,6810	3,0545	4,3178
13	1,7709	2,1604	2,6503	3,0123	4,2208
14	1,7613	2,1448	2,6245	2,9768	4,1405
15	1,7531	2,1314	2,6025	2,9467	4,0728
16	1,7459	2,1199	2,5835	2,9208	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,6460
35	1,6896	2,0301	2,4377	2,7238	3,5911
40	1,6839	2,0211	2,4233	2,7045	3,5510
45	1,6794	2,0141	2,4121	2,6896	3,5203
50	1,6759	2,0086	2,4033	2,6778	3,4960
55	1,6730	2,004	2,3961	2,6682	3,4764
60	1,6706	2,0003	2,3901	2,6603	3,4602
70	1,6669	1,9944	2,3808	2,6479	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,4019
100	1,6602	1,9840	2,3642	2,6259	3,3905
110	1,6588	1,9818	2,3607	2,6213	3,3812
120	1,6577	1,9799	2,3578	2,6174	3,3735
∞	1,6448	1,9600	2,3263	2,5758	3,2905

Приложение 3

Критические значения критерия χ^2

df	p				df	p				df	p			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
1	2,706	3,842	6,635	10,829	31	41,422	44,993	52,203	61,118	61	75,514	80,232	89,591	100,887
2	4,605	5,992	9,211	13,817	32	42,585	46,202	53,498	62,508	62	76,630	81,381	90,802	102,165
3	6,251	7,815	11,346	16,269	33	43,745	47,408	54,789	63,891	63	77,745	82,529	92,010	103,442
4	7,779	9,488	13,278	18,470	34	44,903	48,610	56,074	65,269	64	78,860	83,675	93,217	104,717
5	9,236	11,071	15,088	20,519	35	46,059	49,810	57,356	66,641	65	79,973	84,821	94,422	105,988
6	10,645	12,593	16,814	22,462	36	47,212	51,007	58,634	68,008	66	81,085	85,965	95,626	107,257
7	12,017	14,067	18,478	24,327	37	48,363	52,201	59,907	69,370	67	82,197	87,108	96,828	108,525
8	13,362	15,509	20,093	26,130	38	49,513	53,393	61,177	70,728	68	83,308	88,250	98,028	109,793
9	14,684	16,921	21,669	27,883	39	50,660	54,582	62,444	72,080	69	84,418	89,391	99,227	111,055
10	15,987	18,309	23,213	29,594	40	51,805	55,768	63,707	73,428	70	85,527	90,531	100,425	112,317
11	17,275	19,677	24,729	31,271	41	52,494	56,953	64,967	74,772	71	86,635	91,670	101,621	113,577
12	18,549	21,028	26,221	32,917	42	54,090	58,135	66,224	76,111	72	87,743	92,808	102,816	114,834
13	19,812	22,365	27,693	34,536	43	55,230	59,314	67,477	77,447	73	88,850	93,945	104,010	116,092
14	21,064	23,688	29,146	36,132	44	56,369	60,492	68,728	78,779	74	89,956	95,081	105,202	117,347
15	22,307	24,999	30,583	37,706	45	57,505	61,668	69,976	80,107	75	91,061	96,217	106,393	118,599
16	23,542	26,299	32,006	39,262	46	58,641	62,841	71,221	81,431	76	92,166	97,351	107,582	119,850
17	24,769	27,591	33,415	40,801	47	59,774	64,013	72,463	82,752	78	94,374	99,617	109,958	122,347
18	25,989	28,873	34,812	42,323	48	60,907	65,183	73,703	84,069	79	95,476	100,749	111,144	123,595
19	27,204	30,147	36,198	43,832	49	62,038	66,351	74,940	85,384	80	96,578	101,879	112,329	124,839
20	28,412	31,415	37,574	45,327	50	63,167	67,518	76,175	86,694	90	107,565	113,145	124,116	137,208
21	29,615	32,675	38,940	46,810	51	64,295	68,683	77,408	88,003	100	118,498	124,342	135,807	149,449
22	30,813	33,929	40,298	48,281	52	65,422	69,846	78,638	89,308	110	129,385	135,480	147,414	161,582
23	32,007	35,177	41,647	49,742	53	66,548	71,008	79,866	90,609	120	140,233	146,567	158,950	173,618
24	33,196	36,420	42,989	51,194	54	67,673	72,168	81,092	91,909	130	151,045	157,610	170,423	185,573
25	34,382	37,658	44,324	52,635	55	68,796	73,326	82,316	93,205	140	161,827	138,613	181,841	197,450
26	35,563	38,891	45,652	54,068	56	69,919	74,484	83,538	94,499	150	172,581	179,581	193,207	209,265
27	36,741	40,119	46,973	55,493	57	71,040	75,639	84,758	95,790	200	226,021	233,994	249,445	267,539
28	37,916	41,343	48,289	56,910	58	72,160	76,794	85,976	97,078	250	279,050	287,882	304,939	324,831
29	39,087	42,564	49,599	58,320	59	73,279	77,947	87,192	98,365	300	331,788	341,395	359,906	381,424
30	40,256	43,780	50,904	59,722	60	74,397	79,099	88,406	99,649	350	384,306	394,626	414,474	

Приложение 4

Критические значения критерия F-Фишера при $p=0,05$

df_2	Степени свободы df_1											
	1	2	3	4	5	6	7	8	10	12	24	25
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,785	8,745	8,638	8,634
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,735	4,678	4,527	4,521
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,637	3,575	3,410	3,404
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	2,978	2,913	2,737	2,730
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,854	2,788	2,609	2,601
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,753	2,687	2,505	2,498
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,671	2,604	2,420	2,412
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,602	2,534	2,349	2,341
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,544	2,475	2,288	2,280
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,494	2,425	2,235	2,227
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,412	2,342	2,150	2,141
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,348	2,278	2,082	2,074
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,165	2,092	1,887	1,878
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,077	2,003	1,793	1,783
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,026	1,952	1,737	1,727
70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	1,969	1,893	1,674	1,664
100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,927	1,850	1,627	1,616
200	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,878	1,801	1,572	1,561
∞	3,843	2,998	2,607	2,374	2,216	2,100	2,011	1,940	1,833	1,754	1,519	1,528

Список сокращений

ANOVA – дисперсионный анализ (от англ. ANalysis Of VAriance);

cov_{xy} – ковариация признаков x и y ;

SS – сумма квадратов (от англ. Sum of Squares);

ЛПУ – лечебно-профилактическое учреждение;

МСЭК – медико-социальная экспертная комиссия;

ЧСС – частота сердечных сокращений.

Основная литература

1. Гельман В.Я. Медицинская информатика. Практикум. СПб: Питер, 2001. – 420 с.
2. Информатика. Книга 2. Основы медицинской информатики : учебник / В.И. Чернов, И. Э. Есауленко, М.В. Фролов и др. – М. : Дрофа, 2009. – 205, [3] с. : ил.
3. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: учебное пособие для практических занятий / под ред. В.З. Кучеренко. - 4-е изд., перераб. и доп. - М. : ГЭОТАР-Медиа, 2011. – 256 с.
4. Сабанов В.И., Голубев А.Н., Комина Е.Р. Медицинская информатика и автоматизированные системы управления в здравоохранении: Учебно-методическое пособие к практическим занятиям. - Волгоград: Изд-во ВолГМУ, 2006. – 144с.
5. Симонович С.В. Информатика. Базовый курс: Учебник для вузов. 3-е изд. Стандарт третьего поколения. – СПб.: Питер, 2011. – 640 с.: ил.

Дополнительная литература

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
2. ГОСТ Р ИСО 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. - М.: Изд-во стандартов. 2002. - 30 с.
3. Зайцев В.М., Лифляндский В.Г., Маринкин В.И. Прикладная медицинская статистика: Учебное пособие. СПб, 2003.
4. Информатика для медиков [Электронный ресурс]: учебное пособие / Г. А. Хай. - СПб. : СпецЛит, 2009. - 223 с. – Режим доступа: <http://www.studmedlib.ru>
5. Ланг Т.А. Как описывать статистику в медицине. Аннотированное руководство для авторов, редакторов и рецензентов / Т.А. Ланг, М Сесик; пер. с англ. Под ред. В. П. Леонова. – М.: Прикладная медицина, 2011. – 480 с.: ил.
6. Основы высшей математики и математической статистики [Электронный ресурс]: учебник / И.В. Павлушков и др.: 2-е изд., испр. - М. : ГЭОТАР-Медиа, 2009. - 432 с.: ил. – Режим доступа: <http://www.studmedlib.ru>
7. Статистические методы анализа в здравоохранении. Краткий курс лекций. [Электронный ресурс]: Подготовлены авторским коллективом в составе: д.м.н., проф. Леонов С.А., при участии к.м.н. Вайсман Д.Ш., Моравская С.В, Мирсков Ю.А. - М.: ИД "Менеджер здравоохранения", 2011. - 172 с. – Режим доступа: <http://www.studmedlib.ru>
8. Степанов А.Н. «Информатика. Учебник для вузов. 4-е изд.». СПб., Питер, 2006, 684 с.

Учебное издание

Голубев Алексей Николаевич
Грибина Лариса Николаевна
Дьяченко Тамара Сергеевна

**Ситуационные задачи по медицинской статистике с примерами
решений в программе Microsoft Excel.
Под общей редакцией профессора В.И. Сабанова**

Редакторы _____
Компьютерная верстка _____
Директор Издательства ВолгГМУ *Л.К. Кожевников*

Подписано в печать __. __. 20__ г. Формат 60x84/16
Усл. печ. л. __. Уч.-изд. л. __. Бумага офсетная.
Гарнитура Times. Тираж 500 экз. Заказ № __.

Волгоградский государственный медицинский университет
400131, Волгоград, пл. Павших борцов, 1
Издательство ВолгГМУ
400006б, Волгоград, ул. Дзержинского, 45