МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ «ВИТЕБСКИЙ ГОСУДАРСТВЕННЫЙ ОРДЕНА ДРУЖБЫ НАРОДОВ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»

Кафедра общественного здоровья и здравоохранения

В.С. ГЛУШАНКО, А.П. ГРУЗНЕВИЧ, С.Л. ГАРАНИЧЕВА
Н.С. АЛЯХНОВИЧ, Л.П. КОЛБАСИЧ

ОСНОВЫ МЕДИЦИНСКОЙ СТАТИСТИКИ

Учебно-методическое пособие

Рекомендовано Учебно-методическим объединением Республики Беларусь по медицинскому образованию по специальности высшего образования 1-79 80 23 «общественное здоровье и здравоохранение» в качестве учебно-методического пособия для студентов высших учебных заведений.

Витебск
Издательство ВГМУ
2012

Библиотека ВГМУ
0 0 0 3 4 9 9 2

УДК 61:31(072)
ББК 51.1(4Беи),02р30
Г 55

Рекомендовано Учебно-методическим объединением Республики Беларусь по медицинскому образованию по специальности высшего образования 1-79 80 23 «общественное здоровье и здравоохранение» в качестве пособия для студентов высших учебных заведений.

Рецензент:

Тищенко Е.М. – декан медико-диагностического факультета УО «Гродненский государственный медицинский университет», профессор кафедры общественного здоровья и здравоохранения;

Бурак И.И. зав. каф. общей гигиены и экологии ВГМУ, профессор

Глушанко, В.С.

Г 55 Основы медицинской статистики: учеб.-метод. пособие/ В.С. Глушанко, А.П. Грузневич, С.Л. Гараничева, Н.С. Аляхнович, Л.П. Колбасич. – Витебск: ВГМУ, 2012. – 155 с.

ISBN 978-985-466-488-0

2870/2

Учебно-методическое пособие «Основы медицинской статистики» по дисциплине «Общественное здоровье и здравоохранение» соответствует типовому учебному плану и типовой учебной программе по предмету, утвержденным Министерством здравоохранения Республики Беларусь.

Пособие предназначено для студентов лечебного, фармацевтического и стоматологического факультетов, факультета подготовки иностранных граждан, слушателей ФПК и ПК, магистрантов, аспирантов и соискателей высших медицинских учреждений образования.

УДК 61:31(072)
ББК 51.1(4Беи),02р30

ВИТЕБСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ
БИБЛИОТЕКА

ISBN 978-985-466-488-0

© Глушанко В.С., Грузневич А.П., С.Л. Гараничева, Н.С. Аляхнович, Л.П. Колбасич. 2012
© УО «Витебский государственный медицинский университет», 2012

ОСНОВЫ МЕДИЦИНСКОЙ СТАТИСТИКИ

Главным требованием к учебному процессу на кафедре общественного здоровья и здравоохранения является подготовка врачей-специалистов, отвечающих современным запросам общества в условиях построения и развития белорусской модели социально-ориентированной рыночной экономики и вхождения в мировое рыночное хозяйство.

Для общественного здоровья и здравоохранения статистический метод является основным, поскольку позволяет обосновать тактику и приоритеты использования лечебно-профилактических мероприятий, выявить динамику состояния здоровья населения и его особенности, а также успешно решать проблемы менеджмента и маркетинга в пределах инновационного поля деятельности.

В этой связи возрастает значимость обучения будущих врачей теоретическим основам и практическим навыкам применения классических и современных методов медицинской статистики.

ТЕМА 1

СТАТИСТИЧЕСКИЕ СОВОКУПНОСТИ. СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ

I. ЦЕЛЬ И ЗАДАЧИ ЗАНЯТИЯ

Цель: Изучить основные статистические методы оценки общественного здоровья и здравоохранения путем расчета и анализа средних величин, построения и преобразования динамических рядов, адекватного выбора и построения графических изображений полученных результатов.

Задачи

1. Изучить особенности статистического метода, применяемого в медицине и организациях здравоохранения.
2. Усвоить методику определения объектов наблюдения и единиц измерения.
3. Ознакомиться с сущностью относительных величин, таких как интенсивные показатели, экстенсивные показатели, показатели соотношения и показатели наглядности.
4. Овладеть методиками расчета показателей интенсивности и показателей экстенсивности, показателей соотношения и показателей наглядности.
5. Определить области применения относительных величин в медицине и здравоохранении.

6. Освоить методику построения динамических рядов и их преобразование.

7. Овладеть способами наглядного представления статистических данных.

8. Усвоить методику построения графических изображений.

9. Рассмотреть области применения графических изображений в медицине и здравоохранении.

II. СТУДЕНТ ДОЛЖЕН ЗНАТЬ

1. Определение статистики как науки, основные категории и понятия.

2. Область применения, цели, задачи медицинской статистики и ее значение в работе врача.

3. Определение и содержание статистического метода исследований как основного в общественном здоровье и здравоохранении.

4. Определение объекта и единицы статистического исследования.

5. Понятие статистической совокупности, признаки сходства, признаки различия единиц статистической совокупности.

6. Виды статистических совокупностей, учетные признаки; их характер и роль в статистической совокупности; виды частотных распределений количественных и качественных признаков.

7. Величины, используемые для характеристики статистической совокупности.

8. Определение динамического ряда, уровни, виды, основные показатели, преобразование.

9. Графические изображения, применяемые в медицинской статистике, их виды, требования к выбору.

10. Способы обработки статистической совокупности в среде электронных таблиц Microsoft Excel.

III. СТУДЕНТ ДОЛЖЕН УМЕТЬ

1. Устанавливать объект, единицу и учетные признаки при выполнении статистического исследования.

2. Рассчитывать относительные статистические величины: интенсивные показатели, экстенсивные показатели, показатели соотношения и наглядности.

3. Вычислять показатели динамического ряда: абсолютный прирост (убыль), темп прироста (убыли), значение 1% прироста (убыли), темп роста (убыли).

4. Наглядно предстaвлять статистические величины.

5. Преобразовывать динамический ряд методом укрупнения интервала и сглаживанием ряда при помощи групповой и скользящей средней.

6. Вычислять показатели динамического ряда с использованием персональных компьютеров (ПК).

7. Применять различные виды графических изображений, используя возможности электронных таблиц Microsoft Excel.

IV. МОТИВАЦИЯ ТЕМЫ ЗАНЯТИЯ

Расчет и применение относительных величин, построение и преобразование динамических рядов, наглядное изображение статистической информации широко применяются в современной медицинской статистике. Названные методы необходимы при статистическом анализе заболеваемости, рождаемости, смертности, инвалидности, а также при изучении их структуры и динамики.

V. ОСНОВНЫЕ ВОПРОСЫ ТЕМЫ

1. Статистика. Медицинская статистика, ее роль в деятельности врача. Группы показателей оценки здоровья населения: медико-демографические, заболеваемости и травматизма, физического развития.

2. Объект статистического исследования. Виды статистических совокупностей. Единица статистического наблюдения, учетные признаки.

3. Статистические величины (абсолютные, относительные, средние). Виды относительных величин, значение, применение в медицине, методика вычисления.

4. Динамический ряд, уровни, виды. Показатели динамического ряда. Выравнивание динамического ряда.

5. Графические методы отображения статистических данных: виды диаграмм и графиков, требования к выбору и построению.

VI. МАТЕРИАЛЬНОЕ ОСНАЩЕНИЕ

1. Плакаты и мультимедийные презентации с иллюстрациями показателей оценки состояния здоровья населения и групп факторов, влияющих на состояние здоровья населения.

2. Логическая структура изучаемой темы (рис.1.1).

3. Наглядные пособия, отражающие медико-демографические показатели, показатели заболеваемости и физического развития, для оценки которых используются относительные величины.

4. Примеры графических изображений статистических величин и их видов на ПК.

5. Технические средства обучения: ПК, мультимедийный проектор, операционная система семейства Windows, табличный процессор Microsoft Excel.

VII. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ

На лечебном факультете тема изучается в течение 6 академических часов и состоит из трех частей:

1. **Вступительное слово преподавателя.**

Во вступительном слове преподаватель обращает внимание студентов на значение статистического метода для медицины и здравоохранения; знакомит с основными понятиями медицинской статистики.

2. **Диагностика знаний студентов по теме: «Статистическая совокупность. Статистические величины».**

Преподаватель выясняет исходный уровень знаний студентов.

3. **Инструктаж по выполнению работы и контроль над ходом ее выполнения.**

На примере типового задания преподаватель разбирает методику вычисления относительных величин, составление динамического ряда и способы построения графических изображений относительных величин. Рассматривает особенности оформления выводов.

Далее студенты самостоятельно под контролем преподавателя выполняют варианты индивидуальных заданий по составлению плана и программы статистического исследования, определению объема выборки (количества наблюдений) для конкретного исследования. Проверяют результаты выполнения индивидуальных заданий на ПК.

VIII. ИНФОРМАЦИОННЫЙ БЛОК ТЕМЫ

1. Статистика. Медицинская статистика, ее роль в деятельности врача. Группы показателей оценки здоровья населения: медико-демографические, заболеваемости и травматизма, физического развития

Статистика (нем. statistik, от итал. Stato-государство) - это наука, изучающая количественную сторону массовых социально-экономических явлений и процессов, в неразрывном единстве с их качественной стороной в конкретных условиях места и времени.

В естественных науках понятие «статистика» означает анализ массовых явлений, основанных на применении методов теории вероятности.

Статистика разрабатывает специальную методологию исследования и обработки материалов: массовые статистические наблюдения, метод группировок, средних величин, индексов, балансовый метод, метод графических изображений.

Методологическими особенностями статистики являются изучение: массовости явления; качественно однородных признаков того или иного явления в динамике.

К статистической информации предъявляется *ряд требований:* качественная однородность, достоверность, полнота сведений, своевременность поступления материалов.

Статистика включает *ряд разделов,* среди которых: общая теория статистики, экономическая статистика, отраслевые статистики - промышленная, сельского хозяйства, транспорта, медицинская и др.

Медицинская статистика - это раздел статистики, рассматривающий количественные и качественные аспекты лечебно-профилактической и оздоровительной деятельности.

Медицинская статистика дает возможность врачу обосновать оптимальную тактику лечения и профилактики заболеваний.

Медицинская статистика рассматривает статистику здоровья населения и статистику здравоохранения.

Статистика здоровья населения используется для оценки и анализа здоровья людей, а статистика здравоохранения – для оценки и анализа деятельности медицинских учреждений и их подразделений.

К задачам медицинской статистики относятся:

- выявление особенностей состояния здоровья населения и определяющих его факторов;
- изучение данных о сети, деятельности и кадрах учреждений здравоохранения (УЗ);
- совершенствование применения методов медицинской статистики в экспериментальных, клинических, гигиенических и лабораторных исследованиях.

Здоровье населения характеризуется тремя группами основных показателей.

а) Медико-демографические – отражают состояние и динамику демографических процессов:

– статика населения (плотность, размещение, социальный состав, состав по полу и возрасту, грамотность, образование, национальность, язык, культура);

– динамика населения (механическая эмиграция и иммиграция, естественная рождаемость, смертность, естественный прирост);

– семейное состояние (коэффициент брачности, разводов, средняя продолжительность брака);

– процессы воспроизводства (суммарная плодовитость, брутто-коэффициент и нетто-коэффициент);

– средняя ожидаемая продолжительность жизни;

– смертность (структура смертности, показатели смертности в зависимости от причины, характера заболеваемости и возраста).

б) Показатели заболеваемости и травматизма (первичная заболеваемость, распространенность, накопленная заболеваемость, патологическая пораженность, индекс здоровья, летальность, травматизм, инвалидность).

в) Показатели физического развития:

– антропометрические (рост, масса тела, окружность грудной клетки, головы, плеча, предплечья, голени, бедра);

– физиометрические (жизненная емкость легких, мышечная сила кистей рук, становая сила);

– соматоскопические (телосложение, развитие мускулатуры, степень упитанности, форма грудной клетки, форма голеней и стоп, выраженность вторичных половых признаков).

2. Объект статистического исследования. Виды статистических совокупностей. Единица статистического наблюдения, учетные признаки.

Объектом статистического исследования является ***статистическая совокупность*** – группа или множество относительно однородных элементов, взятых вместе в конкретных границах времени и пространства и обладающих признаками сходства и различия *(учетные признаки)*. Статистическая совокупность бывает *генеральной* (состоит из всех единиц наблюдения) и *выборочной* (представлена частью генеральной совокупности).

Единица статистического наблюдения является носителем основного признака изучаемого явления и представляет собой первичный элемент статистической совокупности.

Учетные признаки единицы наблюдения бывают *качественные (атрибутивные)* (выражены словесно - пол, профессия, и т.д.), *количественные* (выражены числом – рост, масса тела и др.), *факторные* (под влиянием которых изменяются другие, зависящие от них признаки) и *результативные* (изменяющиеся под влиянием факторного признака) (рис. 1.1). Например: возраст - факторный признак, а рост, по отношению к возрасту - результативный.

3. Статистические величины (абсолютные, относительные, средние). Виды относительных величин, значение, применение в медицине, методика вычисления.

В статистических исследованиях применяются абсолютные, относительные и средние величины.

Абсолютные величины, сгруппированные в таблицах, хотя и не характеризуют явление или процесс в полной мере, но уже сами по себе несут важную количественную информацию: численность населения города, число коек в стационаре и т.д.

Для повышения информативности абсолютных величин, последние преобразуются в *относительные* путем деления одной абсолютной величины на другую и умножением на 100 (или 1000, 10000 и т.д.).

Относительные величины используются для характеристики изучаемой совокупности по качественным признакам, сравнения и сопоставления одной совокупности с другой, они могут быть выражены в процентах (%), промилле ($^0/_{00}$), продецимилле ($^0/_{000}$).

Средние величины способны характеризовать целую группу однородных единиц одним (средним) числом: средний рост, средняя продолжи-

тельность жизни и др. Значение средних величин - в их свойстве уравновешивать все индивидуальные отклонения, в результате проявляется то наиболее устойчивое и типичное, что позволяет отличать один групповой объект от другого (например, сравнительная оценка физического развития нескольких коллективов детей).

Виды величин и другие объекты, рассматриваемые в данной теме представлены на рис. 1.1.

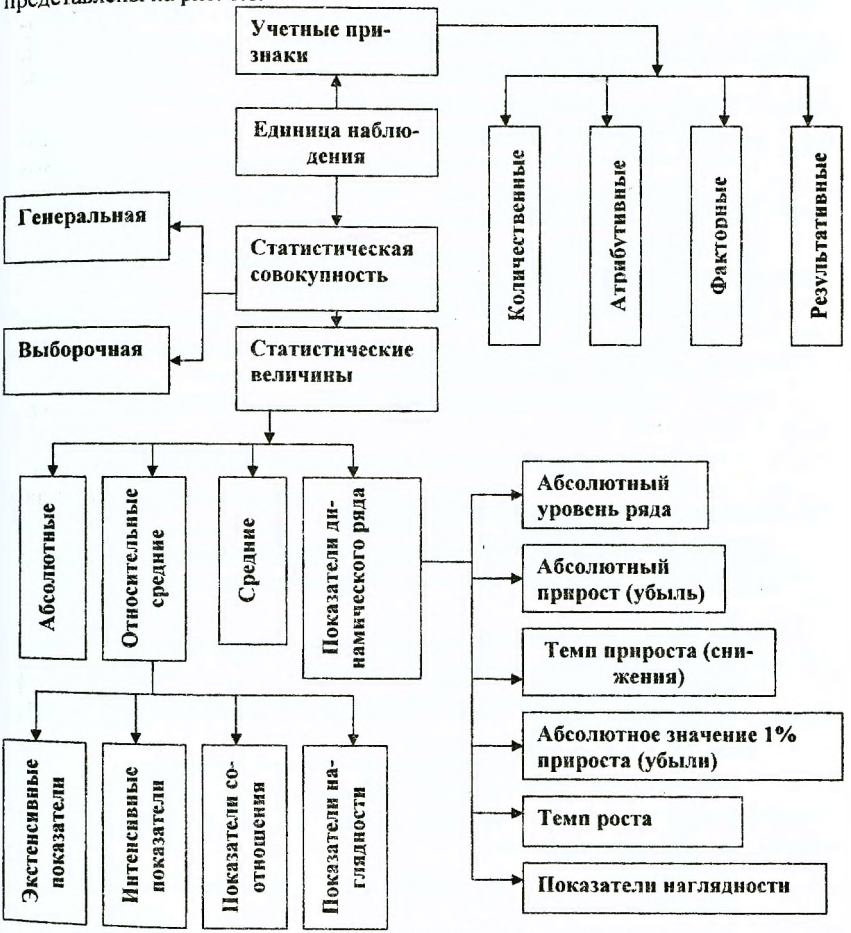


Рис. 1.1 Логическая структура изучаемой темы

Различают следующие виды относительных величин:

– ***интенсивные показатели*** – отражают уровень, частоту или распространенность явления и позволяют сравнить различные совокупности по указанным признакам (заболеваемость, рождаемость, смертность, летальность);

– ***экстенсивные показатели*** – демонстрируют отношение части к целому или долю части в целом, структуру или распределение явления внутри одной совокупности (структура заболеваемости, смертности, инвалидности);

– ***показатели наглядности*** – показывают, на сколько процентов, или во сколько раз произошло увеличение или уменьшение сравниваемых величин. Одну из сравниваемых однородных величин принимают за 100, а все остальные выражают в виде отношения к данной, принятой за 100 величине (например, динамика обеспеченности населения койками);

– ***показатели соотношения*** – отражают отношение двух самостоятельных независимых совокупностей, т.е. отношение изучаемого явления к какой-то величине, принятой за соизмеритель (обеспеченность населения врачами). Обычно рассчитываются на 1000 населения;

4. Динамический ряд, уровни, виды. Показатели динамического ряда. Выравнивание динамического ряда

Динамический ряд – это ряд статистических однородных величин, показывающих изменение какого-то явления в последовательные периоды времени. Его называют также хронологическим.

Числа, составляющие динамический ряд, являются *уровнями динамического ряда.*

Ряды могут быть *простыми и сложными*. В ***простых рядах*** уровни ряда представлены абсолютными величинами. Различают два вида простых рядов: интервальный и моментный.

Интервальный ряд состоит из последовательного ряда уровней, характеризующих изменение явления за определенный интервал времени.

Моментный ряд представлен уровнями, определяющими размеры явления на определенную дату (момент).

Сложный динамический ряд формируется из относительных или средних величин (уровней).

Динамический ряд характеризуют следующие показатели:

– ***абсолютный уровень ряда*** - величины (уровни), из которых состоит динамический ряд (отражают размер явления на определенный момент или интервал времени);

– ***абсолютный прирост (убыль) ряда*** (показывает разность между последующим и предыдущим уровнем динамического ряда);

– ***темп прироста или снижения*** (выражает отношение абсолютного прироста к предыдущему уровню динамического ряда в % и показывает, на сколько процентов увеличился (снизился) последующий уровень, по сравнению с предыдущим);

– ***абсолютное значение 1 % прироста или убыли*** (выражает отношение абсолютного прироста к темпу прироста);

– ***темп роста*** (представляет процентное отношение последующего уровня к предыдущему уровню);

– ***показатель наглядности*** (показывает отношение каждого уровня ряда к одному из них, чаще начальному, принятому за 100%).

С целью выявления основных закономерностей, свойственных явлению за наблюдаемый период, применяют следующие *методы выравнивания динамического ряда:*

– ***укрупнение ряда*** (суммирование данных соседних периодов);

– ***расчет групповой средней*** – определение средней величины каждого укрупненного периода (интервала);

– ***расчет скользящей средней*** (каждый уровень ряда заменяется на среднюю из данного уровня и двух соседних с ним);

Рассмотрим пример выравнивания динамического ряда по методу укрупнения интервалов (столбец 3) и сглаживания ряда методом групповой скользящей средней (столбцы 4, 5 и 6):

1. Проводим группировку по трем периодам наблюдения ($\sum_{i=1}^{3} Xi$).

2. Рассчитываем среднюю по укрупненным интервалам ($\bar{X}$).

3. Проводим группировку с отступом на один показатель (1,2,3; 2,3,4; 3,4,5 и т.д.).

Вычисляем Y_{i+1} по следующей формуле: $Y_{i+1} = (\sum_{i=1}^{3} Xi)$.

Например: $Y_2 = 3,5+2,8+3,6$; $Y_3 = 2,8+3,6+3,7$ и т.д.

4. Рассчитываем скользящую среднюю ($\bar{y}$), как ($\bar{y} = \sum_{i=1}^{3} Y_{i+1}/3$)

Таблица 1.1 Исходные данные и результаты выравнивания динамического ряда

Период наблюдения	Xi	По укрупненным интервалам	Групповая средняя	Скользящая средняя	
		$\sum_{i=1}^{3} Xi$	$\bar{X}$	Y_{i+1}	$\bar{Y}$
1	2	3	4	5	6
1	3,5	9,9	3,3	-	-
2	2,8			9,9	3,3
3	3,6			10,1	3,4
4	3,7	12,2	4,07	11,8	3,9
5	4,5			12,2	4,1
6	4,0			13,2	4,4
7	4,7	13,6	4,5	13,3	4,4
8	4,6			13,6	4,5
9	4,3			-	-

5. Графические методы отображения статистических данных. Виды диаграмм и графиков, требования к выбору и построению

Для наглядного представления статистических данных используется их графическое изображение в виде графиков, диаграмм, картограмм и картодиаграмм. При применении графического метода следует учитывать, что содержание каждого показателя должно соответствовать виду графического отображения.

Интенсивные показатели, показатели *соотношения* и показатели *наглядности*, могут быть графически представлены в виде четырех типов диаграмм: столбиковой, линейной, картограммы и картодиаграммы. Для изображения *экстенсивных* показателей применяют секторные и внутристолбиковые диаграммы.

При использовании графических методов, наряду с правильным выбором типа диаграммы, необходимо соблюдение ряда дополнительных требований:

- наличие полного названия, которое отражает смысл графического изображения, его содержание, место и время;
- соблюдение определенного масштаба;
- наличие условных обозначений.

Диаграммы – это наглядные средства представления и анализа статистических данных посредством точек, линий и геометрических фигур. Различают плоскостные (двухмерные) и объемные (трехмерные) диаграммы, а по форме - в виде линий (лент), секторов, столбиков и геометрических фигур.

Линейные (ленточные) диаграммы преимущественно используются для характеристики явления или процесса в динамике.

Секторные диаграммы наиболее удачно отображают структуру и структурные сдвиги процессов и явлений.

Столбиковые диаграммы применяются для представления частоты явления однородных, но не связанных между собой процессов и явлений.

Объемные диаграммы, как правило, представляют собой линейные в трехмерном пространстве (пространственные и фигурные диаграммы, стереограммы).

Картограммы – это особые географические (контурные) карты с различными штриховками или оттенками цветов определенных территорий и регионов. Разновидностью картограмм могут быть ***картодиаграммы***, которые представлены сочетанием картограммы с нанесенными на ней диаграммами (чаще всего столбиковыми).

IX. МЕТОДИКА ВЫЧИСЛЕНИЯ ОТНОСИТЕЛЬНЫХ ВЕЛИЧИН

Приведем формулы для вычисления относительных величин:

$$\text{Показатель соотношения} = \frac{\text{Абсолютный размер явления * 100 (1000 и т.д)}}{\text{Абсолютный размер среды}}$$

$$\text{Экстенсивный показатель} = \frac{\text{Абсолютный размер части явления}}{\text{Абсолютный размер явления в целом}} * 100$$

$$\text{Интенсивный показатель} = \frac{\text{Абсолютный размер явления * 100 (1000 и т.д)}}{\text{Абсолютный размер среды}}$$

$$\text{Показатель наглядности} = \frac{\text{Явление}}{\text{Аналогичное из ряда сравниваемых}} *100$$

Примечание: *например, при вычислении показателя наглядности по данным обращаемости лиц (таблица 1.3), в качестве «явления» берутся данные обращаемости из строки «интенсивный» в соответствующей группе (20-59, 60-69, 70 и более) - числитель, а в качестве знаменателя берется показатель обращаемости в возрастной группе 15-19 – его значение принимается за 100%.*

X. МЕТОДИКА ВЫЧИСЛЕНИЯ ПОКАЗАТЕЛЕЙ ДИНАМИЧЕСКОГО РЯДА

Основные формулы, применяемые для выравнивания динамического ряда

$$\text{Абсолютный прирост (убыль)} = \text{последующий показатель} - \text{предыдущий показатель}$$

$$\text{Темп прироста (убыли)} = \frac{\text{Абсолютный прирост (убыль)}}{\text{Предыдущий уровень}} * 100\ (\%)$$

или

$$\text{Темп прироста (убыли)} = \text{Темп роста (убыли)} - 100\ (\%)$$

$$\text{Значение 1\% прироста} = \frac{\text{Абсолютный прирост (убыль)}}{\text{Темп прироста (убыль)}}$$

$$\text{Темп роста (убыль)} = \frac{\text{Последующий уровень}}{\text{Предыдущий уровень}} * 100\ (\%)$$

IX. ПРИМЕРЫ ДИАГРАММ И ГРАФИКОВ

Рис. 1.2 Столбиковая диаграмма

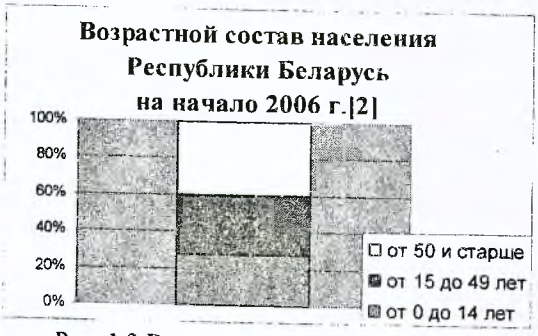


Рис. 1.3 Внутристолбиковая диаграмма

Рис. 1.4 Линейная диаграмма

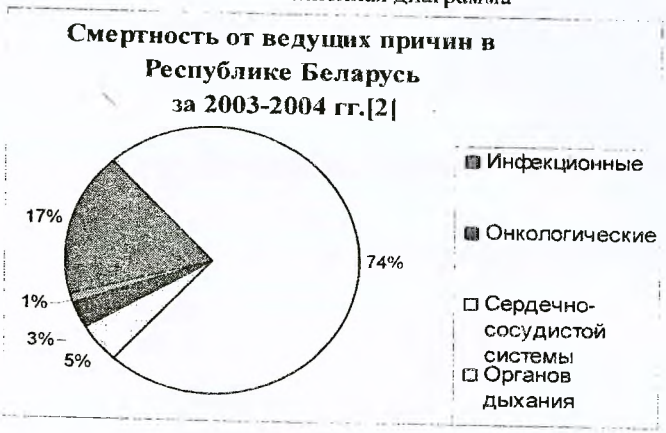


Рис. 1.5 Секторная диаграмма

Территориальное распределение заболеваемостью корью на 100000 населения в 1988-2000 гг. [3]

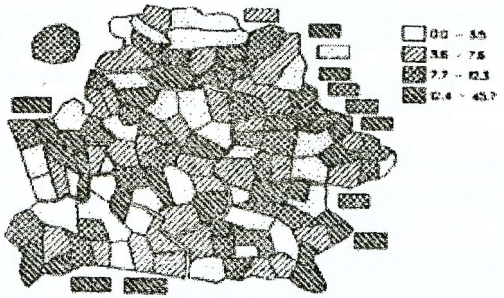


Рис. 1.6. Картограмма

XII. УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА СТУДЕНТОВ

ЗАДАНИЕ № 1. ВЫЧИСЛЕНИЕ ОТНОСИТЕЛЬНЫХ ВЕЛИЧИН

Типовое задание (*Вариант 1*)

Вычислить интенсивные и экстенсивные показатели, а также показатели соотношения и наглядности по исходным данным:

Таблица 1.2 Численность населения разных возрастов в районе деятельности городской поликлиники № 1.

Возрастная группа	Численность населения	Число обращений
15 – 19	3000	2000
20 – 59	28000	36000
60 – 69	6000	10000
70 и более	3000	2000
Итого:	***40000***	***50000***

Примечание: *Общая численность посещений составляет - 400000, а число врачей - 30.*

Образец выполнения типового задания

1. Расчет интенсивных показателей (уровень посещаемости, общий уровень обращаемости и уровень обращаемости по различным возрастным группам):

$$\text{Уровень посещаемости} = \frac{\text{Число посещений}}{\text{Численность населения}} * 1000 = \frac{400000}{40000} * 1000 = 10000‰$$

Уровень обращаемости (общий) = Число обращений / Численность населения * 1000 = 50000 / 40000 *1000=1250‰

Уровень обращаемости (лица 15-19 лет) = Число обращений / Численность населения * 1000 = 2000 / 3000 *1000=667‰

Уровень обращаемости (лица 20-59 лет) = Число обращений / Численность населения * 1000 = 36000 / 28000 *1000=1286‰

Уровень обращаемости (лица 60-69 лет) = Число обращений / Численность населения * 1000 = 10000 / 6000 * 1000=1667‰

Уровень обращаемости (лица 70 и более лет) = Число обращений / Численность населения * 1000 = 2000 / 3000 * 1000=667‰

2. Расчет экстенсивных показателей (по возрастным группам):

Доля обращения лиц 15-19 лет = Число обращений лиц 15-19 лет / Все обращения * 100= 2000 / 50000 * 100=4%

Доля обращения лиц 20-59 лет = Число обращений лиц 20-59 лет / Все обращения * 100= 36000 / 50000 * 100=72%

Доля обращения лиц 60-69 лет = Число обращений лиц 60-69 лет / Все обращения * 100= 10000 / 50000 * 100=20%

Доля обращения лиц 70 и более лет = Число обращений лиц 70 и более лет / Численность населения * 100= 2000 / 50000 * 100=4%

3. Расчет показателей соотношения:

Число врачей на 10000 населения = Число врачей / Численность населения * 10000= 30 / 40000 * 10000=7,5‰

4. Расчет показателя наглядности при условии, что показатель обращаемости лиц 15 - 19 лет принимается за 100%:

Таблица 1.3 Интенсивные показатели и показатели наглядности

Показатели	**Возрастные группы**			
	15 - 19	20 - 59	60 - 69	70 и более
Интенсивный	667	1285	1667	667
Наглядности	100	193	250	100

Таблица 1.4 Сводная таблица данных

Возрастная группа	**Структура обращаемости в %**	**Возрастная обращ. на 1000 населения**	**Показатели наглядности**
15 – 19	4,0	667	100
20 – 59	72,0	1285	193
60 – 69	20,0	1667	250
70 и более	4,0	667	100
Итого:	**100**	**1250**	-
Посещаемость на 1000 населения	10000 ‰		
Число врачей на 10000 населения	7,5 ‰		

Выводы:

При расчете интенсивных показателей уровень посещаемости поликлиники составил 10000‰, уровень общей обращаемости был равен 1250‰, а по возрастной группе 60-69 лет был наибольшим (1667‰).

Вычисление экстенсивного показателя - доля обращения лиц по возрастной группе 20-59 лет была максимальной и составила 72%, а минимальной у лиц 15-19 и 70 и более лет.

По показателю соотношения - число врачей на 10000 населения равнялся 7,5‰, что приблизительно соответствует врачебной нагрузке 1300 человек на 1 врача-терапевта участкового (ВТУ).

Показатель наглядности при условии, что показатель обращаемости лиц 15 - 19 лет принимается за 100% составил для возрастной группы 20-59 лет - 193%, 60-69 - 250% и 70 и более лет - 100%.

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Алгоритм обработки данных

1. Составить на первом листе книги документа Excel макет для обработки данных на примере задания варианта 1 (таблица 1.2). Макет вычислений представлен на рис. 1.7.

2. Ввести исходные данные своего варианта в заранее подготовленный макет (рис.1.7).

3. Проанализировать результаты.

Результаты вычислений для варианта 1 в среде электронных таблиц Excel приведены на рис. 1.8.

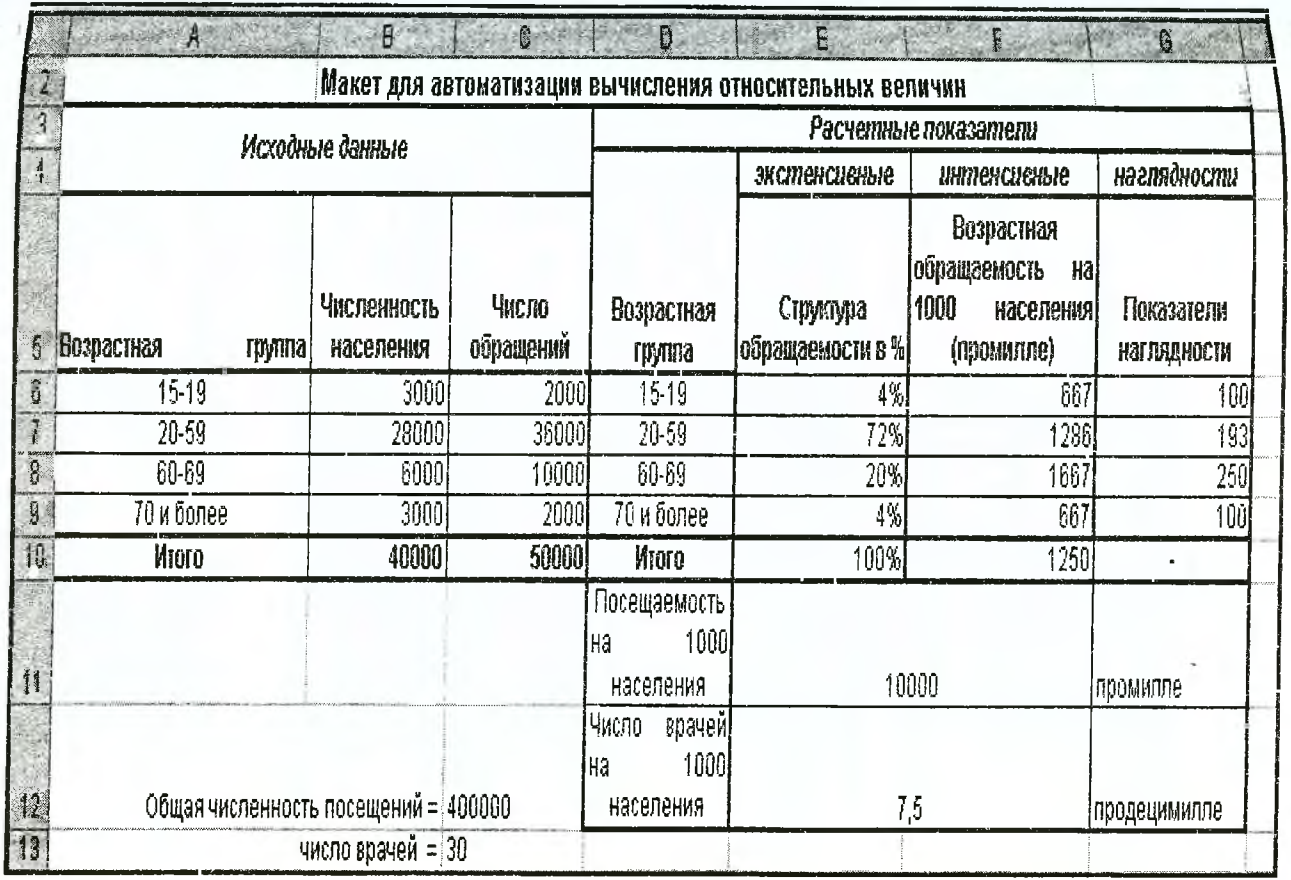
Примечание:

1. *Формулы начинаются со знака «равно» (=).*
2. *Диапазон ячеек (прямоугольный блок) обозначается адресами его диагонали, разделенными символом «двоеточие» (:).*
3. *При вводе формул следует учитывать, что диапазон размещения массивов исходных данных в различных вариантах отличается. В соответствии с этим, формулы для вычислений следует вводить исходя из их смысловой нагрузки.*

ВИТЕБСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ
БИБЛИОТЕКА

	A	B	C	D	E	F	G
2	**Макет для автоматизации вычисления относительных величин**						
3	*Исходные данные*			*Расчетные показатели*			
4					*экстенсивные*	*интенсивные*	*наглядности*
5	**Возрастная группа**	**Численность населения**	**Число обращений**	**Возрастная группа**	**Структура обращаемости в %**	**Возрастная обращаемость на 1000 населения (промилле)**	**Показатели наглядности**
6	15-19	3000	2000	15-19	=C6/$C$10	=C6/B6*1000	100
7	20-59	28000	36000	20-59	=C7/$C$10	=C7/B7*1000	=F7/$F$6*100
8	60-69	6000	10000	60-69	=C8/$C$10	=C8/B8*1000	=F8/$F$6*100
9	70 и более	3000	2000	70 и более	=C9/$C$10	=C9/B9*1000	=F9/$F$6*100
10	**Итого**	**=СУММ(B6:B9)**	**=СУММ(C6:C9)**	**Итого**	=СУММ(E6:E9)	= C10/B10*1000	
11				Посещаемость на 1000 населения	=C12/B10*1000		промилле
12	Общая численность посещений =		400000	Число врачей на 1000 населения	=C13/B10*10000		продецимилле
13	число врачей =		30				

Рис. 1.7 Макет для автоматизации вычисления относительных величин



	A	B	C	D	E	F	G
2	Макет для автоматизации вычисления относительных величин						
3	*Исходные данные*			*Расчетные показатели*			
4					*экстенсивные*	*интенсивные*	*наглядности*
5	Возрастная группа	Численность населения	Число обращений	Возрастная группа	Структура обращаемости в %	Возрастная обращаемость на 1000 населения (промилле)	Показатели наглядности
6	15-19	3000	2000	15-19	4%	667	100
7	20-59	28000	36000	20-59	72%	1286	193
8	60-69	6000	10000	60-69	20%	1667	250
9	70 и более	3000	2000	70 и более	4%	667	100
10	**Итого**	**40000**	**50000**	**Итого**	100%	1250	-
11				Посещаемость на 1000 населения	10000		промилле
12	Общая численность посещений =		400000	Число врачей на 1000 населения	7,5		продецимилле
13	число врачей =		30				

Рис. 1.8 Результаты расчета относительных величин

Варианты заданий на вычисление относительных величин для самостоятельного выполнения

Вариант 2. Численность населения разных возрастов в районе деятельности городской поликлиники № 2.

Возрастная группа	Численность населения	Число обращений
15 – 19	3000	1000
20 – 59	25000	30000
60 – 69	5000	8000
70 и более	2000	1000
Итого:	**35000**	**40000**

Примечание: *Общая численность посещений составляет - 350000, а число врачей - 25.*

Вариант 3. Численность населения разных возрастов в районе деятельности городской поликлиники № 3.

Возрастная группа	Численность населения	Число обращений
15 – 19	3500	1500
20 – 59	35000	30000
60 – 69	7000	9000
70 и более	3000	1500
Итого:	**48500**	**42000**

Примечание: *Общая численность посещений составляет - 400000, а число врачей - 35.*

Вариант 4. Численность населения разных возрастов в районе деятельности городской поликлиники № 4.

Возрастная группа	Численность населения	Число обращений
15 – 19	4000	1500
20 – 59	40000	30000
60 – 69	10000	12000
70 и более	2000	1300
Итого:	**56000**	**44800**

Примечание: *Общая численность посещений составляет - 450000, а число врачей - 40.*

Вариант 5. Численность населения разных возрастов в районе деятельности городской поликлиники № 5.

Возрастная группа	Численность населения	Число обращений
15 – 19	3000	2500
20 – 59	40000	30000
60 – 69	10000	18000
70 и более	2000	1500
Итого:	**55000**	**52000**

Примечание: *Общая численность посещений составляет - 300000, а число врачей - 25.*

Вариант 6. Численность детского населения по возрастным группам в районе деятельности детской поликлиники № 1.

Возрастная группа	Численность детей	Число обращений
0 – 1	3000	2500
1 – 3	2500	2500
4 – 14	6500	5000
Итого:	**12000**	**10000**

Примечание: *Общая численность посещений составляет - 50000, а число врачей-педиатров - 15.*

Вариант 7. Численность детского населения по возрастным группам в районе деятельности детской поликлиники № 2.

Возрастная группа	Численность детей	Число обращений
0 – 1	3500	3000
1 – 3	4000	3700
4 – 14	7500	5300
Итого:	**15000**	**12000**

Примечание: *Общая численность посещений составляет - 55000, а число врачей-педиатров - 17.*

Вариант 8. Численность детского населения по возрастным группам в районе деятельности детской поликлиники № 3.

Возрастная группа	Численность детей	Число обращений
0 – 1	2500	2000
1 – 3	3000	2500
4 – 14	4500	3500
Итого:	**10000**	**8000**

Примечание. Общая численность посещений составляет - 40000, а число врачей-педиатров - 12.

Вариант 9. Численность детского населения по возрастным группам в районном центре и число профилактических посещений поликлиник (в абсолютных числах).

Возрастная группа	Численность детей	Число проф. посещений
0 – 1	100	1000
1 – 3	300	1300
4 – 14	4600	4700
Итого:	**5000**	**7000**

Примечание: *Число педиатрических коек- 3, а число врачей-педиатров - 4.*

Вариант 10. Численность детского населения по возрастным группам в городе и число профилактических посещений амбулаторно-поликлинических организаций (в абсолютных числах).

Возрастная группа	Численность детей	Число проф. посещений
0 – 1	500	5000
1 – 3	1000	4000
4 – 14	18500	16000
Итого:	**20000**	**25000**

Примечание: *Число педиатрических коек - 10, а число врачей-педиатров - 30.*

Вариант 11. Численность детского населения по различным возрастным группам в областном центре и число профилактических посещений амбулаторно-поликлинических организаций (в абсолютных числах).

Возрастная группа	Численность детей	Число проф. посещений
0 – 1	5000	50000
1 – 3	10000	30000
4 – 14	85000	70000
Итого:	**100000**	**150000**

Примечание: *Число педиатрических коек - 400, а число врачей-педиатров - 150.*

Вариант 12. Число учащихся учебных заведений, прошедших медицинские профилактические осмотры, и количество выявленных с нарушениями функции желудочно-кишечного тракта.

Группы учащихся	Число прошедших проф. осмотр	Число выявленных больных
Школьники	23000	350
Учащиеся ССУЗ-ов	17000	250
Учащиеся ВУЗ-ов	10000	100
Итого:	**50000**	**700**

Примечание: *В лечебно-профилактических организациях города работает 200 врачей, численность населения - 250000 человек.*

Вариант 13. Число детей, прошедших медицинские профилактические осмотры, и количество выявленных с понижением остроты зрения.

Возрастная группа	Число прошедших проф. осмотр	Число выявленных с пониженем остроты зрения
Дошкольного возраста	15000	800
Учащихся 1 класса	12000	650
Учащихся 2 - 8 классов	13000	750
Итого:	**40000**	**2200**

Примечание: *В лечебно-профилактических организациях города работает 250 врачей, численность населения - 300000 человек.*

Вариант 14. Численность женского населения детородного (фертильного) возраста и число обращений и посещений женской консультации № 1.

Возрастная группа	Численность женского населения	Число обращений
15 - 24	7000	1200
25 - 34	6000	3000
35 - 49	17000	5800
Итого:	**30000**	**10000**

Примечание: *Общая численность посещений составляет - 50000, а число врачей-гинекологов - 20.*

ЗАДАНИЕ № 2. ВЫЧИСЛЕНИЕ ПОКАЗАТЕЛЕЙ ДИНАМИЧЕСКОГО РЯДА

Типовое задание (*Вариант 1*)

Произвести расчет показателей динамического ряда: абсолютный прирост, темп прироста, значение 1% прироста и темп роста.

Объем миграции населения Республики Беларусь (2001-2007 гг.)

1. 2001 г. – 424521
2. 2002 г. – 429101
3. 2003 г. – 446732
4. 2004 г. – 460224
5. 2005 г. – 474685
6. 2006 г. – 474004
7. 2007 г. – 472334

Образец выполнения типового задания

1. Абсолютный прирост (убыль) - разность между последующим и предыдущим уровнем объема миграции населения:

429101 (2002 г.) – 424521(2001 г.) = 4580;
446732 (2003 г.) – 429101(2002 г.) = 17631;
460224 (2004 г.) – 446732(2003 г.) = 13492;
474685 (2005 г.) – 460224(2004 г.) = 14461;
474004 (2006 г.) – 474685(2005 г.) = -681;
472334 (2007 г.) – 474004(2006 г.) = -1670
За 7 лет: 472334 (2007 г.)-424521 (2001 г.) = 47813.

2. Темп прироста (убыли) - отношение абсолютного прироста к предыдущему уровню в %:

$$\frac{4580 * 100}{424521} = 1,1\ \%$$

$$\frac{17631 * 100}{429101} = 4,1\ \%$$

$$\frac{13492 * 100}{446732} = 3\ \%$$

$$\frac{14641 * 100}{460224} = 3{,}1\ \%$$

$$\frac{-681 * 100}{474685} = -0{,}1\ \%$$

$$\frac{-1666 * 100}{474004} = -0{,}4\ \%$$

За 7 лет: $\frac{\mathbf{47813}}{\mathbf{424521}} * 100 = 11{,}3\ \%$

3. Значение 1% прироста (убыли) - отношение абсолютного прироста к темпу прироста (убыли):

4580/1,1 = 4163,6
17631/4,1 = 4300,2
13492/3 = 4497,3
14461/3,1 = 4664,8
-681/(-0,1) = 681,1
-1670/(-0,4) = 1670,4
За 7 лет: 47813/11.3=4231,2

4. Темп роста (убыли) - отношение последующего объема миграции к предыдущему в % :

$$\frac{429101}{424521} * 100 = 101{,}1\ \%$$

$$\frac{446732}{429101} * 100 = 104{,}1\ \%$$

$$\frac{460224}{446732} * 100 = 103\ \%$$

$$\frac{474685}{460224} * 100 = 103{,}1\ \%$$

$$\frac{474004}{474685} * 100 = 99{,}9\ \%$$

$$\frac{472334}{474004} * 100 = 99{,}6\ \%$$

За 7 лет: $\frac{\mathbf{472334}}{\mathbf{424521}} * \mathbf{100} = \mathbf{111{,}3\ \%}$

5. Оформление результатов:

Таблица 1.5 Динамика объема миграции населения Республики Беларусь (2001-2007 гг.)

Показатели	2001	2002	2003	2004	2005	2006	2007	Итого за 7 лет
Объем миграции	424521	429101	446732	460224	474685	474004	472334	-

Абсолютный прирост (убыль)	-	4580	17631	13492	14461	-681	-1670	47813
Темп прироста (убыли)	-	1,1%	4,1%	3%	3,1%	-0,1%	-0,4%	11,3%
Значение 1% прироста (убыли)	-	4163,6	4300,2	449,3	4664,8	-681,1	-1670,4	4231,2
Темп роста (убыли)	-	101,1%	104,1%	103%	103,1%	99,9%	99,9%	111,3%

6. Выводы:

1. За семь лет абсолютный прирост лиц, вовлеченных в миграционный процесс, увеличился на 47813 человек.

2. Максимальный темп прироста объема миграции наблюдался в 2003 году и составил 4,1%, затем, отмечалось постепенное его снижение, и с 2006 года началась убыль, достигшая в 2007 году - 0,4%.

3. Значение 1% прироста объема миграции увеличивалось до 2003 года включительно и составило 4300,2; а с 2006 года наблюдалась убыль, составившая в 2007 году - 1670,4 человек.

4. Темп роста (убыли) был относительно стабильным и варьировал в пределах 4,5%.

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Алгоритм обработки данных

1.Составить на первом листе книги документа Excel макет для обработки данных на примере задания варианта №1 (см. рис.1.9).

2. Ввести исходные данные своего варианта в заранее подготовленный макет.

3. Проанализировать результаты.

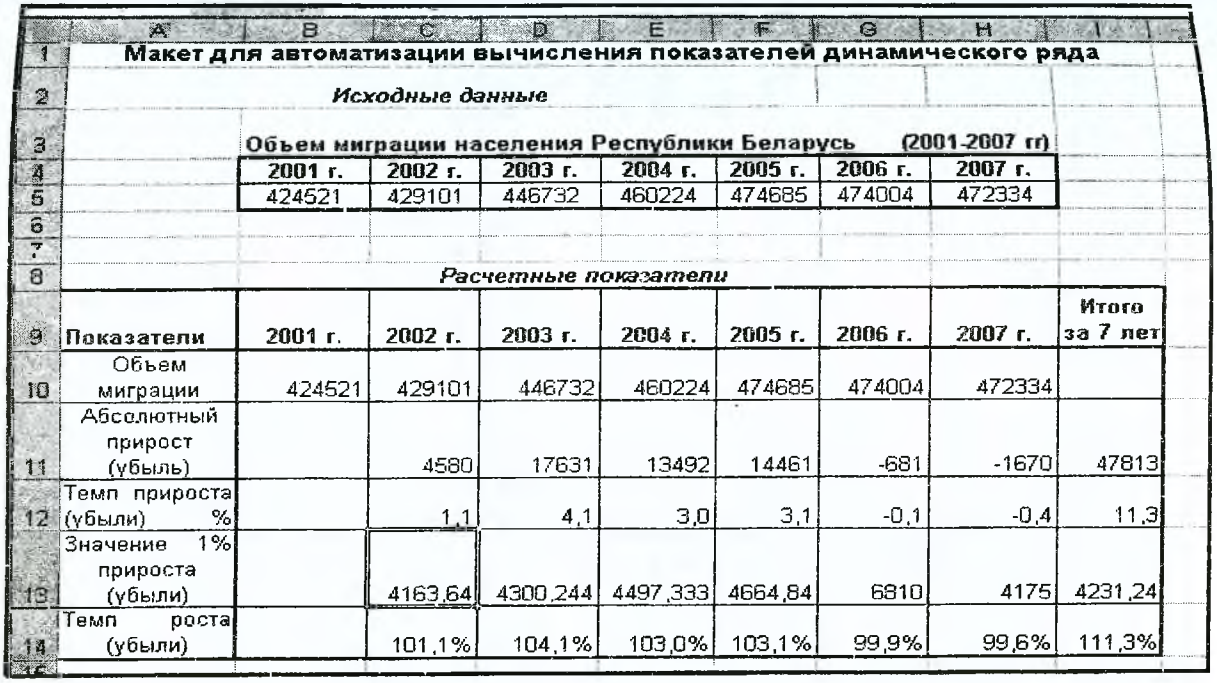
Результаты вычислений для варианта №1 приведены на рис.1.10.

Примечание:

1. *При вводе формул следует учитывать, что диапазон размещения массивов исходных данных в различных вариантах отличается. В соответствии с этим, формулы для вычислений следует вводить исходя из их смысловой нагрузки.*
2. *Для обеспечения полного соответствия результатов расчета на ПК данным, полученным путем выполнения расчетов на калькуляторе с точностью один десятичный знак после запятой в формулах Excel можно применять функцию округления: ОКРУГЛ (число;n), где число – адрес ячейки округляемого значения, n – количество знаков после запятой (принимаем n = 1). Для округления результатов вычисления (=C9/C10), следует применять формулу = ОКРУГЛ (C9/C10).*

	A	B	C	D	E	F	G	H	I
1	**Макет для автоматизации вычисления показателей динамического ряда**								
2			*Исходные данные*						
3		Объем м							
4		2001 г.	2002 г.	2003 г.	2004 г.	2005 г.	2006 г.	2007 г.	
5		424521	429101	446732	460224	474685	474004	472334	
6									
7									
8			*Расчетные показатели*						
9	Показатели	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.	2006 г.	2007 г.	Итого за 7 лет
10	Объем миграции	424521	429101	446732	460224	474685	474004	472334	
11	Абсолютный прирост (убыль)		=C5-B5	=D5-C5	=E5-D5	=F5-E5	=G5-F5	=H5-G5	=H5-B5
12	Темп прироста (убыли) %		=ОКРУГЛ((C5/B5-1)*100;1)	=ОКРУГЛ((D5/C5-1)*100;1)	=ОКРУГЛ((E5/D5-1)*100;1)	=ОКРУГЛ((F5/E5-1)*100;1)	=ОКРУГЛ((G5/F5-1)*100;1)	=ОКРУГЛ((H5/G5-1)*100;1)	=ОКРУГЛ((H5/B5-1)*100;1)
13	Значение 1% прироста		=C11/C12	=D11/D12	=E11/E12	=F11/F12	=G11/G12	=H11/H12	=I11/I12
14	Темп роста (убыли)		=C10/B10	=D10/C10	=E10/D10	=F10/E10	=G10/F10	=H10/G10	=H5/B5

Рис. 1.9 Макет для автоматизации вычисления показателей динамического ряда



Макет для автоматизации вычисления показателей динамического ряда

Исходные данные

Объем миграции населения Республики Беларусь (2001-2007 гг)

2001 г.	2002 г.	2003 г.	2004 г.	2005 г.	2006 г.	2007 г.
424521	429101	446732	460224	474685	474004	472334

Расчетные показатели

Показатели	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.	2006 г.	2007 г.	Итого за 7 лет
Объем миграции	424521	429101	446732	460224	474685	474004	472334	
Абсолютный прирост (убыль)		4580	17631	13492	14461	-681	-1670	47813
Темп прироста (убыли) %		1,1	4,1	3,0	3,1	-0,1	-0,4	11,3
Значение 1% прироста (убыли)		4163,64	4300,244	4497,333	4664,84	6810	4175	4231,24
Темп роста (убыли)		101,1%	104,1%	103,0%	103,1%	99,9%	99,6%	111,3%

Рис. 1.10 Результаты расчета показателей динамического ряда

Варианты заданий по вычислению показателей динамического ряда для самостоятельного выполнения

Вариант 2.

Численность постоянного населения Республики Беларусь на начало года (тыс.)

2004 – 9849,1
2005 – 9800,1
2006 – 9750,5
2007 – 9714,5
2008 – 9689,8

Вариант 3.

Численность постоянного городского населения Республики Беларусь на начало года (тыс.)

2004 – 7045,5
2005 – 7055,9
2006 – 7059,0
2007 – 7074,8
2008 – 7108,1

Вариант 4.

Численность постоянного сельского населения Республики Беларусь на начало года (тыс.)

2004 – 2803,6
2005 – 2744,2
2006 – 2691,5
2007 – 2639,7
2008 – 2581,7

Вариант 5.

Рождаемость населения в Республике Беларусь (на 1000 населения)

2003 – 9,0
2004 – 9,1
2005 – 9,3
2006 – 9,9
2007 – 10,7

Вариант 6.

Смертность населения Республики Беларусь (на 1000 населения)

2003 – 14,5
2004 – 14,3
2005 – 14,5
2006 – 14,2
2007 – 13,7

Вариант 7.

Естественный прирост населения Республики Беларусь (на 1000 населения)

2003 – -5,5
2004 – -5,2
2005 – -5,2
2006 – -4,3
2007 – -3,0

Вариант 8.

Младенческая смертность в Республике Беларусь (на 100 родившихся живыми)

2003 – 7,7
2004 – 6,9
2005 – 7,1
2006 – 6,1
2007 – 5,2

Вариант 9.

Обеспеченность населения Республики Беларусь больничными койками по основным профилям (на 10000 населения-все ведомости).

2003 – 113,7
2004 – 107,4
2005 – 111,6
2006 – 112,0
2007 – 112,4

Вариант 10.

Обеспеченность населения Республики Беларусь терапевтическими больничными койками (на 10000 населения-все ведомости)

2003 – 35,8
2004 – 32,3
2005 – 31,9
2006 – 31,9
2007 – 32,4

Вариант 11

Число кожно-венерологических диспансеров в Республике Беларусь.

2003 – 30
2004 – 28
2005 – 23
2006 – 21
2007 – 21

Вариант 12.

Число амбулаторных посещений на одного жителя Республики Беларусь

2003 – 11,7
2004 – 12,4
2005 – 12,5
2006 – 12,6
2007 – 12,7

Вариант 13.

Число госпитализированных пациентов (включая рожениц и родильниц) на 100 городских и сельских жителей Республики Беларусь

2003 – 27,5
2004 – 27,0
2005 – 26,4
2006 – 26,4
2007 – 27,4

Вариант 14.

Заболеваемость населения Республики Беларусь активным туберкулезом всех форм с впервые в жизни установленным диагнозом (на 1000 населения)

2003 – 48,8
2004 – 51,0
2005 – 47,9
2006 – 46,2
2007 – 42,9

ЗАДАНИЕ № 3. ПРИМЕНЕНИЕ ДИАГРАММ И ГРАФИКОВ НАГЛЯДНОГО ПРЕДСТАВЛЕНИЯ СТАТИСТИЧЕСКИХ ВЕЛИЧИН

Типовое задание (*Вариант 1*)

Представить структуру обращаемости за медицинской помощью по возрастным группам в виде секторной диаграммы (рассчитанные показатели задания №2 варианта 1).

Структура обращаемости за медицинской помощью:

15-19 лет – 4%
20-59 лет – 72%
60-69 лет – 20%
70 лет и более – 4%

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Алгоритм обработки данных

1. Составить на первом листе книги документа Excel макет таблицы с результатами, полученными в задании №2 варианта №1 (см. рис. 1.11).

2. Ввести исходные данные своего варианта в заранее подготовленный макет.

3. Построить предложенный в варианте тип диаграммы.

Примечание: *Алгоритм построения диаграммы*

1. *Выделить на рабочем листе Excel ячейки с данными для построения диаграммы;*
2. *Вызвать «Мастер диаграмм», щелкнув мышью по соответствующей пиктограмме на панели инструментов «Стандартная»;*
3. *Выбрать тип и задать параметры диаграммы (при построении внутристолбиковой диаграммы выбрать тип – «гистограмма», вид – «нормированная гистограмма», в разделе диапазон данных выбрать – «ряды в строках»).*

Пример вычислений и построения диаграммы приведены на рис. 1.11

Возрастная группа, лет	Структура обращаемости за медицинской помощью, %
15-19	4
20-59	72
60-69	20
70 и более	4

Рис. 1.11 Результаты графической обработки статистических величин

Варианты заданий по наглядному представлению статистических величин для самостоятельного выполнения

Вариант 2. Представить структуру обращаемости за медицинской помощью населения разных возрастов в районе деятельности городской поликлиники № 2 в виде секторной диаграммы (по заданию №1 своего варианта).

Вариант 3. Представить структуру обращаемости за медицинской помощью населения разных возрастов в районе деятельности городской поликлиники № 3 в виде внутристолбиковой диаграммы (по заданию №1 своего варианта).

Вариант 4. Представить уровни обращаемости за медицинской помощью населения разных возрастов в районе деятельности городской поликлиники № 4 в виде столбиковой диаграммы (по заданию №1 своего варианта).

Вариант 5. Представить численность населения разных возрастов в районе деятельности городской поликлиники № 5 в виде линейной диаграммы (по заданию №1 своего варианта).

Вариант 6. Представить структуру обращаемости за медицинской помощью детей разных возрастных групп в районе деятельности детской поликлиники № 1 в виде секторной диаграммы (по заданию №1 своего варианта).

Вариант 7. Представить структуру обращаемости детей разных возрастов за медицинской помощью населения в районе деятельности детской поликлиники № 2 в виде внутристолбиковой диаграммы (по заданию №1 своего варианта).

Вариант 8. Представить уровни обращаемости детей разных возрастных групп за медицинской помощью в районе деятельности детской поликлиники № 3 в виде столбиковой диаграммы (по заданию №1 своего варианта).

Вариант 9. Представить число профилактических посещений амбулаторно-поликлинических организаций детьми различных возрастных групп в районном центре в виде столбиковой диаграммы (по заданию №1 своего варианта).

Вариант 10. Представить в виде линейной диаграммы число профилактических посещений амбулаторно-поликлинических организаций в районном центре детьми различных возрастных групп (в промиллях) (по заданию №1 своего варианта).

Вариант 11. Представить в виде столбиковой диаграммы число профилактических посещений амбулаторно-поликлинических организаций в областном центре детьми различных возрастных групп (в промиллях) (по заданию №1 своего варианта).

Вариант 12. Представить структуру выявляемости нарушений функции желудочно-кишечного тракта среди прошедших профилактический осмотр учащихся различных учреждений образования на 1000 населения (промилле) в виде секторной диаграммы (по заданию №1 своего варианта).

Вариант 13. Представить структуру выявляемости понижения остроты зрения среди прошедших профилактический осмотр детей дошкольного возраста и учащихся различных учреждений образования на 1000 населения (промилле) в виде внутристолбиковой диаграммы (по заданию №1 своего варианта).

Вариант 14. Представить структуру обращаемости за медицинской помощью женского населения детородного (фертильного) возраста в районе деятельности женской консультации № 1 в виде секторной диаграммы (по заданию №1 своего варианта).

Ответы к заданиям по темам

Тема 1. СТАТИСТИЧЕСКИЕ СОВОКУПНОСТИ. СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ.

Ответы к заданию №1 «Вычисление относительных величин»

Показатели	№ варианта			
	2	3	4	5
Посещаемость на 1000 населения	10000	7216	8036	5455
Уровень обращаемости (общий)	1143	866	800	945
Доля обращений лиц 70 и более лет	2,5	4	3	3
Число врачей на 1000 населения	7,1	5,2	7,1	4,5

Показатели	№ варианта			Показатели	№ варианта		
	6	7	8		9	10	11
Посещаемость на 1000 детей	4167	3667	4000	Число педиатрических коек на 1000 населения	1	1	4
Уровень обращаемости (общий)	833	800	800	Уровень проф.обращаемости (общий)	1400	1250	1500
Доля обращений детей 4-14 лет	50	44	44	Доля проф.обращений детей 4-14 лет	67	64	47
Число врачей	3	3,1	3	Показатель наглядно-	24	22	27

на 1000 детского населения				сти (дети 4-14 лет)			

Показатели	№ варианта	Показатели	№ варианта	Показатели	№ варианта
	12		13		14
Число прошедших проф.осмотр на 1000 населения	200	Число прошедших проф.осмотр на 1000 населения	133	Посещаемость на 1000 женского населения	1667
Уровень проф. выявляемости нарушений желудочно-кишечного тракта (общий)	14	Уровень выявляемости снижения остроты зрения среди детей (общий)	55	Уровень обращаемости (общий)	333
Доля проф. выявлений пациентов учащихся ВУЗов среди осмотренных учащихся	14	Доля детей дошкольного возраста со снижением остроты зрения среди осмотренных	36	Доля обращений женщин (35-49 лет)	58
Показатель наглядности (учащиеся ВУЗов)	66	Показатель наглядности (учащиеся 2-8 классов)	108	Показатель наглядности (35-49 лет)	199

Ответы к заданию №2
«Вычисление показателей динамического ряда»

№ варианта / Итого за все годы	2	3	4	5	6	7	8
Абсолютный прирост (убыль)	-159,3	62,6	-221,9	1,7	-0,8	2,5	-2,5
Темп прироста (убыли) %	-1,60	0,90	-7,9	18,9	-5,5	-45,5	-32,5
Значение 1% прироста (убыли)	99,56	69,56	28,1	0,09	0,15	-0,05	0,08
Темп роста (убыли)	98,4%	100,9%	92,1%	118,9%	94,5%	54,5%	67,5%

№ варианта / Итого за все годы	9	10	11	12	13	14
Абсолютный прирост (убыль)	-1,3	-3,4	-9	1,0	-0,1	-5,9
Темп прироста (убыли) %	-1,1	-9,5	-30,0	8,5	-0,4	-12,1
Значение 1% прироста (убыли)	1,18	0,36	0,30	0,12	0,25	0,49
Темп роста (убыли)	98,9%	90,5%	70,0%	108,5%	99,6%	87,9%

ТЕМА 2

ВАРИАЦИОННАЯ СТАТИСТИКА. СРЕДНИЕ ВЕЛИЧИНЫ

I. ЦЕЛЬ И ЗАДАЧИ ЗАНЯТИЯ

Цель: Изучить обобщающие характеристики количественных признаков путем вычисления средних величин и их параметров применительно к социально-гигиеническим, экспериментальным и клиническим исследованиям.

Задачи

1. Изучить основные положения теории средних величин.
2. Освоить методику расчета средних величин при малом и большом числе наблюдений.
3. Приобрести навыки составления различных вариационных рядов (простого и сгруппированного).
4. Провести вычисление средних величин при малом и большом числе наблюдений.
5. Усвоить критерии, характеризующие разнообразие признака в статистической совокупности.
6. Научиться практическому применению показателей средних величин в медицине и здравоохранении.

II. СТУДЕНТ ДОЛЖЕН ЗНАТЬ

1. Определение вариационного ряда.
2. Виды вариационных рядов и их характеристики.
3. Порядок составления простого и сгруппированного вариационных рядов.
4. Критерии разнообразия признака в вариационном ряду.
5. Определение средней величины (среднего арифметического), понятия: простое и взвешенное среднее арифметическое, мода и медиана.
6. Свойства средних величин.
7. Методику определения средних величин при большом и малом числе наблюдений.
8. Области применения средних величин в медицине и здравоохранении.

III. СТУДЕНТ ДОЛЖЕН УМЕТЬ

1. Составлять простой и сгруппированный вариационные ряды.
2. Вычислять простое среднее арифметическое (встречающееся с одинаковой частотой и при числе наблюдения $N \leq 30$).
3. Определять моду и медиану.
4. Вычислять взвешенное среднее арифметическое (встречающееся с неодинаковой частотой и при числе наблюдения $N > 30$).
5. Вычислять среднее арифметическое по способу моментов.
6. Определять среднее квадратическое отклонение при малом и большом числе наблюдений.
7. Оценивать результаты вычисления данных показателей.

IV. МОТИВАЦИЯ ТЕМЫ ЗАНЯТИЯ

Кроме относительных величин-коэфициентов, характеризующих частоту (интенсивность), либо состав (структуру) изучаемого явления, статистические совокупности варьирующего объекта могут быть охарактеризованы при помощи средних величин.

Средние величины используют для выявления наиболее устойчивых и типичных свойств, характеризующих качественное своеобразие варьирующего объекта и позволяющих отличать один групповой объект от другого.

Вариационные ряды и их графики дают наглядное представление о распределении признаков, но не достаточное для полного описания варьирующих объектов. Для этой цели служат логически и теоретически обоснованные статистические характеристики – средние величины и показатели вариации.

V. ОСНОВНЫЕ ВОПРОСЫ ТЕМЫ

1. Вариационный ряд, варьирующий признак, варианта, ранжирование. Характеристики и виды вариационных рядов. Методика составления простого и сгруппированного вариационного ряда.
2. Средние величины, понятие, свойства. Среднее арифметическое (простое, взвешенное), мода, медиана, способы вычисления.
3. Понятие критериев разнообразия. Методика их вычисления.

VI. МАТЕРИАЛЬНОЕ ОСНАЩЕНИЕ

1. Плакаты и мультимедийные презентации, иллюстрирующие методики расчета средних величин.
2. Технические средства обучения: ПК, видео двойка MPEG 4 – DVD, мультимедийный проектор, операционная система семейства Windows, табличный процессор Microsoft Excel, программа презентаций Microsoft Power Point.

VII. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ

На лечебном факультете тема изучается в течение 6 часов и состоит из трех частей:

1. Определение исходного уровня знаний студентов, разбор и обсуждение основных вопросов темы.
2. Разбор типовых и индивидуальных заданий по расчету средних величин.
3. Самостоятельное, под контролем преподавателя, выполнение варианта индивидуальных заданий по расчету, анализу и оценке средних величин.

VIII. ИНФОРМАЦИОННЫЙ БЛОК ТЕМЫ

1. Вариационный ряд, варьирующий признак, варианта, ранжирование. Характеристики и виды вариационных рядов. Методика составления простого и сгруппированного вариационного ряда

Вариационный ряд (ряд распределения) – это ряд числовых изменений определенного признака, отличающихся друг от друга по своей величине и расположенных в определенном порядке. В таком ряду количественно изменяющийся признак носит название *варьирующего*, а отдельные его количественные выражения называют *вариантами*.

Числа, показывающие, как часто встречается та или иная варианта в составе данного ряда, называются ***частотами.***

Цели распределения вариант в вариационный ряд:

- Упрощение обработки данных при вычислении обобщающих числовых характеристик и показателей вариации;
- Выявление закономерности варьирования учитываемого признака.

Чтобы вариационный ряд полностью удовлетворял этим целям, значения изменяющегося признака следует ранжировать.

Под ***ранжированием*** понимают расположение вариант ряда в возрастающем (убывающем) порядке. Процесс ранжирования данных, выполняемый на ПК, называют ***сортировкой*** данных.

Вариационный ряд характеризуется средней величиной и различными критериями разнообразия (вариации) – *лимитом, амплитудой ря-*

да, средним квадратическим отклонением (стандартным отклонени-ем) и коэффициентом вариации.

Характеристики вариационного ряда представлены: числовым значением признака или вариантой; частотой, с которой встречается каждая варианта, общим числом наблюдений.

Различают *несколько видов* вариационных рядов в зависимости от частоты встречаемости, группировки, ранжирования и характеристики вариант.

Если варианта встречается в вариационном ряду 1 раз, то такой ряд называется *простым*, а если несколько раз – *взвешенным*.

Если каждая варианта вариационного ряда обозначается отдельно - вариационный ряд будет *несгруппированным*, если же варианты объединяются в группы с указанием частоты их встречаемости в каждой группе, то вариационный ряд является *сгруппированным*.

Как отмечалось выше, вариационный ряд может быть *ранжированным* (при расположении вариант по степени возрастания или убывания их числового значения) и *неранжированным*, когда варианты располагаются хаотично.

Если же варианта принимает только дискретное, т.е. фиксированное значение, выраженное целыми или дробными значениями - вариационный ряд считается *простым*, а если варианта способна принимать любые числовые значения, представленные относительными или средними величинами, – *сложным*.

МЕТОДОЛОГИЯ ПОСТРОЕНИЯ ВАРИАЦИОННОГО РЯДА

Вариационный ряд, отражая один из типов распределения признака в изучаемой совокупности, можно охарактеризовать средними величинами и следующими характеристиками вариационного ряда:

варианта (x_i) – числовое значение изучаемого признака;

частота (n_i) – количество встречаемости каждой (i) варианты;

общее число наблюдений - $N=\sum_{i=1}^{k} n_i$, где k – количество неповторяющихся вариант.

Простой вариационный ряд строится при $N \leq 30$ (малое количество наблюдений).

Сгруппированный - при $N > 30$ (большое количество наблюдений).

При построении сгруппированного вариационного ряда количество групп в ряду зависит от числа вариант и определяется по таблице (таблица 2.1).

Таблица 2.1 Количество групп в вариационном ряду в зависимости от числа вариант

Число вариант (N)	31 - 45	46 - 100	101 - 200	201 - 500
Число групп (g)	6 - 7	8 - 10	11 - 12	13 - 17

Определение интервала значения признака (Δx) между группами:

$$\Delta x = \frac{x_{max} - x_{min}}{g}$$

где x_{max} - максимальное значение варианты, x_{min} - минимальное значение варианты, g - число групп.

Полученный интервал округлить до целого числа. Далее определяются *границы и середина каждой группы.* После распределения вариант по группам строится графическое изображение вариационного ряда (частотное распределение).

2. Средние величины, понятие, свойства. Среднее арифметическое (простое, взвешенное), мода, медиана, способы вычисления

После составления вариационного ряда (простого или сгруппированного) определяется средний уровень признака - *средняя величина* или *среднее арифметическое.*

Средняя величина $(\bar{x})$ – это обобщающая величина изучаемого признака исследуемой совокупности, которая отражает его типичный уровень в расчете на единицу совокупности в конкретных условиях места и времени.

Средняя, рассчитанная по совокупности в целом, называется *общей средней.* Она отражает общие черты явления.

Средняя величина, вычисленная для каждой группы, - это *групповая средняя.* Она характеризует размер явления, наблюдаемый в конкретных условиях данной группы.

Способность средних величин сохранять свойства статистических совокупностей называют *определяющим свойством.*

Средние величины делятся на 2 больших класса:

1. ***Степенные средние*** (среднее гармоническое, среднее геометрическое, среднее арифметическое, среднее квадратическое, среднее кубическое);

2. ***Структурные средние*** (мода и медиана). Мода и медиана определяются структурой распределения. Их часто используют как среднюю характеристику в тех совокупностях, где расчет средней величины невозможен или нецелесообразен (когда используют качественные признаки или при количественных признаках закон частотного распределения не подчиняется распределению Гаусса).

Наиболее часто употребляются три вида средних величин: *среднее арифметическое (простое и взвешенное), мода и медиана.*

Под ***средним арифметическим*** понимают такое значение признака, которое имела бы каждая единица совокупности, если бы общий итог всех значений признака был распределен равномерно между всеми единицами совокупности. Практически, ***простое среднее арифметическое*** вычисляется в случаях, когда варианты встречаются с одинаковой частотой и в совокупности, где $N \leq 30$.

Взвешенное среднее арифметическое ($X_в$) определяется в случаях, когда варианты встречаются с неодинаковой частотой и в совокупности при $N > 30$.

Если же варианты представлены большими числами (например, масса тела новорожденных в граммах) и имеется число наблюдений, выраженное сотнями или тысячами случаев, взвешенное среднее арифметическое может быть вычислено *по способу моментов*.

Среднее арифметическое имеет следующие *свойства*:

1. Занимает срединное положение;
2. Имеет абстрактный характер;
3. Сумма отклонений всех вариант от средней величины равна нулю;
4. Позволяет охарактеризовать исследуемую совокупность одним числом;
5. Дает возможность сравнить отдельные величины со средним арифметическим;
6. Способствует определению тенденции развития какого-либо явления;
7. Позволяет сравнить разные совокупности;
8. Используется для вычисления других статистических показателей.

Следует помнить, что среднее арифметическое используется только для *количественных признаков, подчиняющихся нормальному закону частотного распределения.*

Являясь одной из основных характеристик вариационного ряда, среднее арифметическое не лишено недостатков, т.к. весьма *чувствительно к увеличению или уменьшению числа наблюдений* за счет вариант, резко отличающихся по своей величине от основной массы. Поэтому на величину среднего арифметического могут значительно влиять крайние варианты ранжированного вариационного ряда, которые как раз и наименее характерны для данной совокупности. Устранить указанный выше недостаток среднего арифметического можно путем определения *моды и медианы*, т.к. на их величины не оказывают влияние числовые значения крайних вариант.

Мода (Mo) – соответствует величине признака, которая чаще других встречается в данной совокупности. За моду принимают варианту, которой соответствует наибольшее количество частот (n) вариационного ряда.

Медиана (Me) - величина признака, занимающая срединное положение в вариационном ряду. Она делит ряд на две равные части по числу наблюдений. Для определения медианы надо найти середину ряда. При четном числе наблюдений за медиану принимают среднюю величину из двух центральных вариант. При нечетном числе наблюдений медианой будет срединная (центральная) варианта.

Медиану и моду используют для описания *качественных* признаков и *количественных с законом частотного распределения, не соответствующего «нормальному».*

МЕТОДИКА РАСЧЕТА СРЕДНИХ ВЕЛИЧИН

Как отмечалось выше, обобщенной характеристикой признака в статистической совокупности является средняя величина. Виды средних величин: мода (**Мо**), медиана (**Ме**), среднее арифметическое ($\bar{X}$).

1. Простое среднее арифметическое:

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

где N – количество вариант в вариационном ряду,

x_i – i-тое числовое значение признака (варианта).

2. Взвешенное среднее арифметическое:

$$X_в = \frac{\sum_{i=1}^{k} x_i * n_i}{N}$$

где n_i – количество вариант в i-той группе,

к - количество неповторяющихся вариант в вариационном ряду.

3. Если число наблюдений достигает сотен (тысяч) случаев **взвешенное среднее арифметическое вычисляется по способу моментов:**

$$X_в = A + \frac{\sum_{i=1}^{k} a_i * n_i}{N} * \Delta x$$

где А – условное среднее (чаще всего в качестве условного среднего берется мода (Мо));

Δx – интервал значений исследуемого признака;

a_i – отклонение i-той варианты (в интервалах) от условного среднего: $a_i = x_i - A$.

3. Понятие критериев разнообразия. Методика их вычисления

Величина того или иного признака неодинакова у всех членов статистической совокупности, несмотря на ее относительную однородность. В этом проявляется *разнообразие, вариабельность признака* изучаемого явления. Статистика позволяет охарактеризовать это специальными критериями, определяющими уровень разнообразия каждого признака в этой или иной группе. $\bar{X}$

Выделяют следующие ***критерии разнообразия признака***:

1. Характеризующие *границы совокупности*: ***лимит и амплитуда***;

2. Характеризующие *внутреннюю структуру* совокупности: ***мода и медиана.***

МЕТОДИКА ВЫЧИСЛЕНИЯ КРИТЕРИЕВ РАЗНООБРАЗИЯ

Границы статистической совокупности характеризуются ***лимитом (Lim)*** и ***амплитудой (Am)***, вычисляются по формулам:

$Lim = x_{max} \div x_{min}$, где ÷ - символ диапазона;

$Am = x_{max} - x_{min}$

Внутреннюю структуру статистической совокупности характеризует ***квадратическое отклонение (стандартное отклонение)*** (***σ*** – для генеральной совокупности, ***S*** – для выборочной совокупности). Стандартное отклонение характеризует степень разброса вариант от среднего в вариационном ряду.

$$\text{при } N \leq 30, \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{N-1}}$$

где d_i – является разностью между i-той вариантой ряда (x_i) и средним арифметическим (X□): $d_i = x_i - X$.

При использовании этой формулы имеется в виду, что в вариационном ряду все частоты (n) равны единице.

При n > 1 используют формулу такого вида:

$$\text{при } N \leq 30, \quad \sigma = \sqrt{\frac{\sum_{i=1}^{k} d_i^2 n_i}{N-1}}$$

Можно рассчитать среднее квадратическое отклонение *способом моментов* по формуле:

$$\sigma = \Delta x * \sqrt{\frac{\sum_{i=1}^{k} \alpha_i^2 n_i}{N} - \left(\frac{\sum_{i=1}^{k} \alpha_i n_i}{N}\right)^2},$$

где $\frac{\sum_{i=1}^{k} \alpha_i^2 n_i}{N}$ – момент второй степени, $\left(\frac{\sum_{i=1}^{k} \alpha_i n_i}{N}\right)^2$ – момент первой степени, возведенный в квадрат.

Коэффициент вариации (C_v), вычисляется по формуле:

$$C_v = \frac{\sigma}{\bar{x}} * 100\%,$$

где σ - среднее квадратическое отклонение;

$\bar{x}$ - среднее арифметическое.

Для ориентировочной оценки степени разнообразия признака пользуются следующими *градациями коэффициента вариации:*

$C_v \leq 10\%$ характеризует слабое разнообразие признака,

$C_v = 10 - 20\%$ – среднее разнообразие признака,

$C_v > 20\%$ – сильное разнообразие признака.

Следует помнить, что выборка (вариационный ряд), состоящий из количественных признаков, подчиняющихся «нормальному» закону частотного распределения Гаусса, однозначно может быть описана средним арифметическим и средним квадратическим отклонением (стандартным отклонением).

Обобщающие характеристики количественных признаков: средние величины и их параметры используются в социально-гигиенических, экспериментальных и клинических исследованиях. В практическом здравоохранении средние величины применяют для характеристики работы врачей, организаций здравоохранения, для оценки медико-физиологических показателей организма, физического развития и т.д.

IX. УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА СТУДЕНТОВ

ЗАДАНИЕ № 1. СОСТАВЛЕНИЕ ПРОСТОГО ВАРИАЦИОННОГО РЯДА И РАСЧЕТ ПРОСТОГО СРЕДНЕГО АРИФМЕТИЧЕСКОГО ПРИ МАЛОМ ЧИСЛЕ НАБЛЮДЕНИЙ (N < 30)

Типовое задание *(Вариант 1)*

Составить простой вариационный ряд и вычислить простое среднее арифметическое по частоте дыхания (дыхательных движений) у мужчин среднего возраста (45-55 лет) не занимающихся физической культурой и спортом: частота дыхания в минуту – **16, 22, 20, 24, 19.**

Образец выполнения типового задания

1. Составляем возрастающий вариационный ряд **(16, 19, 20, 22, 24).**
2. Рассчитываем простое среднее арифметическое:

$$X = \frac{\sum_{i=1}^{N} x_i}{N} =$$

$$= (16+19+20+22+24)/5=20,2$$

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории

1. Для вычисления среднего арифметического в среде электронных таблиц Excel можно использовать встроенную статистическую функ-

цию **СРЗНАЧ(A5:A10)**, где **(A5:A10)** – адреса ячеек, в которых находятся данные вариационного ряда.

2. Для ранжирования (сортировки) данных используется команда Excel

Данные ⇒ Сортировка.

Алгоритм обработки данных

1. Введите *текстовые* значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных, который представлен на рис. 2.1, на примере задания варианта №1.

2. Введите исходные данные в блок ячеек с адресами **(A4:A8)**.

3. Скопируйте данные блока ячеек **(A4:A8)** в ячейку с адресом **C4**.

Используйте команды: ***Правка ⇒ Копировать,***

Правка ⇒ Вставить.

4. Выполните сортировку данных по возрастанию в блоке ячеек **(C4:C8)**. (Выделите ячейки **(C4:C8)** и инициируйте команду ***Данные ⇒ Сортировка***).

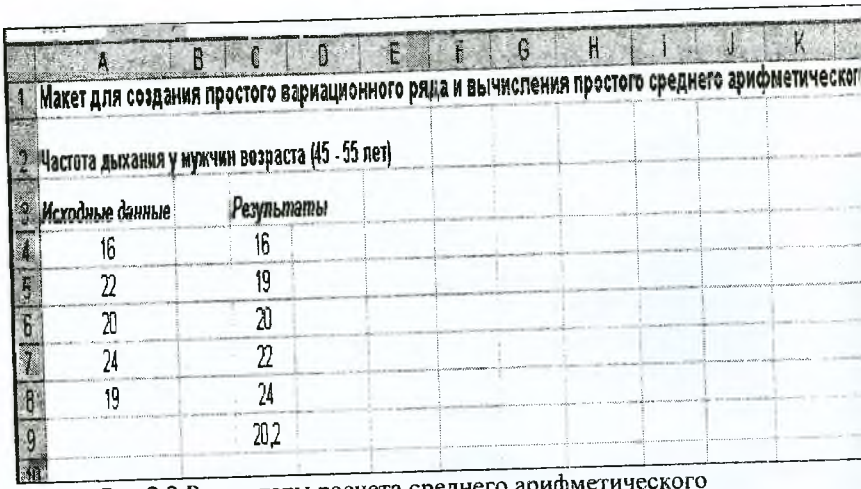
5. Рассчитайте среднее арифметическое, применив встроенную функцию **СРЗНАЧ(C4:C8)**.

6. Проанализируйте полученные результаты (результаты вычислений для варианта №1 приведены на рис. 2.2).

Примечание: *При вводе формул следует учитывать, что диапазон размещения массивов исходных данных в различных вариантах отличается. В соответствии с этим, формулы для вычислений следует вводить исходя из их смысловой нагрузки.*

	A	B	C	D	E	F	G
1	Макет для создания простого вариационного ряда и вычисления простого среднего арифметического						
2	Частота дыхания у мужчин возраста (45 - 55 лет)						
3	*Исходные данные*		*Результаты*				
4	16		16				
5	22		19				
6	20		20				
7	24		22				
8	19		24				
9			=СРЗНАЧ(C4:C8)				
10							

Рис. 2.1 Макет для создания простого вариационного ряда и расчета среднего арифметического



	A	B	C	D	E	F	G	H	I	J	K
1	Макет для создания простого вариационного ряда и вычисления простого среднего арифметическог										
2	Частота дыхания у мужчин возраста (45 - 55 лет)										
3	Исходные данные		Результаты								
4	16		16								
5	22		19								
6	20		20								
7	24		22								
8	19		24								
9			20,2								
10											

Рис.2.2 Результаты расчета среднего арифметического

Варианты заданий на составление простого вариационного ряда и вычисление среднего арифметического для самостоятельного выполнения

Вариант 2. Частота дыхания (дыхательных движений в минуту) у 5 мужчин 30-летнего возраста после 20 приседаний: 27, 23, 24, 28, 25.

Вариант 3. Частота пульса (ударов в минуту) в состоянии покоя у 7 спортсменок занимающихся легкой атлетикой: 68, 70, 67, 75, 72, 69, 65.

Вариант 4. Численность населения в 5 районах области: 32150, 15000, 18542, 12385, 15795.

Вариант 5. Численность населения в четырех районах областного центра: 120000, 85000, 110000, 78000.

Вариант 6. Рост 10 школьников 9 класса средней школы (в см): 180, 175, 177, 174, 181, 167, 172, 182, 164, 173.

Вариант 7. Масса тела (в кг) 6 мужчин проходивших профосмотр при приеме на работу: 65, 80, 74, 93, 70, 105.

Вариант 8. Результат измерения вечерней температуры (в °С) у 5 стационарных пациентов терапевтического отделения: 37,1; 36,8; 37,5; 38,4; 36,6.

Вариант 9. Число пациентов состоящих на диспансерном учете у 7 врачей общей практики: 122, 134, 131, 127, 143, 148, 129.

Вариант 10. Число семей, обслуживаемых 5 врачами общей практики в амбулаториях врача общей практики: 354, 371, 402, 368, 387.

Вариант 11. Число детей до года, состоящих под наблюдением у 9 педиатров участковых детской поликлиники: 62, 54, 57, 52, 60, 53, 55, 61, 63.

Вариант 12. Число обращений с травмами в травматологический пункт районного центра в первые 7 дней месяца: 127, 145, 142, 134, 128, 132, 128.

Вариант 13. Число студентов, прошедших медицинский осмотр за каждый день первой недели учебного года: 12, 15, 10, 14, 22, 25, 18

Вариант 14. Уровень систолического артериального давления (мм рт. ст.) у 5 мужчин 45-летнего возраста: 130, 110, 135, 120, 140.

ЗАДАНИЕ № 2. СОСТАВЛЕНИЕ ПРОСТОГО ВАРИАЦИОННОГО РЯДА, ОПРЕДЕЛЕНИЕ МОДЫ И МЕДИАНЫ, РАСЧЕТ ВЗВЕШЕННОГО СРЕДНЕГО АРИФМЕТИЧЕСКОГО ПРИ БОЛЬШОМ ЧИСЛЕ НАБЛЮДЕНИЙ (N > 30)

Типовое задание *(Вариант 1)*

Составить вариационный ряд при большом числе наблюдений (N > 30), определить моду и медиану, вычислить взвешенное среднее арифметическое по данным длительности амбулаторного лечения 34 пациентов с гриппом: 7, 3, 8, 11, 9, 8, 4, 7, 10, 5, 11, 6, 7, 8, 10, 8, 9, 10, 4, 9, 11, 7, 12, 8, 10, 9, 6, 5, 13, 9, 12, 8, 14, 6.

Образец выполнения типового задания

1. Составляем возрастающий вариационный ряд с учетом числа повторяющихся величин (таблица 2.2).

Таблица 2.2 Возрастающий вариационный ряд

xi	3	4	5	6	7	8	9	10	11	12	13	14
ni	1	2	2	3	4	6	5	4	3	2	1	1

2. Определяем моду (Мо): наиболее часто встречающаяся длительность лечения равна 8 дням.

3. Определяем медиану (Ме):

- порядковый номер медианы:

$$№\ Me = \frac{34}{2} = 17,$$ или Ме = 8 (значение варианты с номером 17)

4. Рассчитываем взвешенную среднюю арифметическую ($X_в$):

$$X_в = \frac{3\cdot1+4\cdot2+5\cdot2+6\cdot3+7\cdot4+8\cdot6+9\cdot5+10\cdot4+11\cdot3+12\cdot2+13\cdot1+14\cdot1}{34} = 8{,}35 \text{ дней.}$$

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории

1. Для вычисления моды в среде электронных таблиц Excel можно использовать встроенную статистическую функцию **МОДА(A3:A40)**, где **(A3:A40)** - адреса блока ячеек; медиана определяется с помощью статистической функции **МЕДИАНА(A3:A40).**

2. Упорядочение данных (сортировки) осуществляет команда Excel

Данные ⇒ Сортировка.

3. Для вычисления количества повторений в массиве заданного значения можно использовать встроенную математическую функцию Excel **СЧЕТЕСЛИ(A3:A40;K)**, где первый параметр **(A3:A40)** обозначает массив ячеек, в которых осуществляется поиск; второй параметр **(K)** - критерий, число повторений которого в заданном массиве вычисляет функция.

Например, **СЧЕТЕСЛИ(A3:A40;5)** - функция вычисляет число повторений значения 5 в массиве данных, расположенного в ячейках **(A3:A40)**.

4. Сумму произведений соответствующих элементов двух массивов ячеек вычисляет функция Excel **СУММПРОИЗВ(D6:D17;E6:E17)**, параметрами которой являются адреса соответствующих массивов.

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта №1 (см. рис. 2.3).

2. Введите исходные данные в блок ячеек с адресами **(A6:A39)**.

3. Скопируйте данные блока ячеек **(A6:A39)** в блок данных ячейки с адресом B6.

4. Выполните сортировку данных по возрастанию в блоке ячеек **(B6:B39)**.

5. Заполните вручную на основании отсортированных данных столбец Xi.

6. Вычислите частоту повторений каждого значения столбца Xi в заданном массиве данных, применив функцию **СЧЕТЕСЛИ(B6:B39;Xi)**.

7. Определите медиану и моду массива данных с помощью соответствующих функций Excel.

8. Вычислите сумму произведений соответствующих элементов массивов Xi и n, применив встроенную функцию **СУММПРОИЗВ(D6:D17;E6:E17)**.

9. Вычислите взвешенное среднее арифметическое.

10. Вычислите с помощью функции **СРЗНАЧ(B6:B39)** среднее значение заданного массива.

11. Сравните два последних значения (макет для выполнения указанных расчетов приведен на рис. 2.3).

12. Проанализируйте результаты расчетов (фрагмент листа книги Excel с результатами приведен на рис. 2.4).

13. Сделайте вывод, о сущности отличия значений среднего арифметического и взвешенного среднего арифметического.

G7

	A	B	C	D	E
1	**Макет для создания простого вариационного ряда**				
2	**определения моды и медианы, взвешенного среднего арифметического**				
3			**Длительность амбулаторного лечения больных гриппом**		
4	***Исходные***			***Результаты***	
5	***данные***	***сортировки***		***Xi***	***ni***
6	7	3		3	=СЧЁТЕСЛИ(B6:B39;D6)
7	3	4		4	=СЧЁТЕСЛИ(B7:B40;D7)
8	8	4		5	=СЧЁТЕСЛИ(B8:B41;D8)
9	11	5		6	=СЧЁТЕСЛИ(B9:B42;D9)
10	9	5		7	=СЧЁТЕСЛИ(B10:B43;D10)
11	8	6		8	=СЧЁТЕСЛИ(B11:B44;D11)
12	4	6		9	=СЧЁТЕСЛИ(B12:B45;D12)
13	7	6		10	=СЧЁТЕСЛИ(B13:B46;D13)
14	10	7		11	=СЧЁТЕСЛИ(B14:B47;D14)
15	5	7		12	=СЧЁТЕСЛИ(B15:B48;D15)
16	11	7		13	=СЧЁТЕСЛИ(B16:B49;D16)
17	6	7		14	=СЧЁТЕСЛИ(B17:B50;D17)
18	7	8	**Мода**	=МОДА(B6:B39)	
19	8	8	**Медиана**	=МЕДИАНА(B6:B39)	
20	10	8	**Сумма произведений**		
21	8	8	**массивов**	=СУММПРОИЗВ(D6:D17;E6:E17)	
22	9	8	**Xв**	=D21/34	
23	10	8	**Среднее арифметическое**	=СРЗНАЧ(B6:B39)	
24	4	9			
25	9	9			
26	11	9			
27	7	9			
28	12	9			
29	8	10			
30	10	10			
31	9	10			
32	6	10			
33	5	11			
34	13	11			
35	9	11			

Рис. 2.3 Фрагмент макета для создания простого вариационного ряда и расчета медианы, моды, взвешенного среднего арифметического

	A	B	C	D	E
1	**Макет для создания простого вариационного ряда**				
2	**определения моды и медианы, взвешенного среднего арифметического**				
3	**Длительность амбулаторного лечения больных гриппом**				
4	*Исходные*	*Результаты*			
5	*данные*	*сортировки*		*Xi*	*ni*
6	7	3		3	1
7	3	4		4	2
8	8	4		5	2
9	11	5		6	3
10	9	5		7	4
11	8	6		8	6
12	4	6		9	5
13	7	6		10	4
14	10	7		11	3
15	5	7		12	2
16	11	7		13	1
17	6	7		14	1
18	7	8	**Мода**	8	
19	8	8	**Медиана**	8	
20	10	8	**Сумма произведений**		
21	8	8	**массивов**	284	
22	9	8	**Xв**	8.35	
23	10	8	**Среднее арифметическое**	8.35	
24	4	9			
25	9	9			
26	11	9			
27	7	9			
28	12	9			
29	8	10			
30	10	10			
31	9	10			
32	6	10			
33	5	11			
34	13	11			
35	9	11			

Рис. 2.4 Фрагмент листа Excel с результатами создания простого вариационного ряда, расчета медианы, моды, взвешенного среднего арифметического

Варианты заданий на составление простого вариационного ряда и расчета медианы, моды, взвешенного среднего арифметического для самостоятельного выполнения

Вариант 2. Длительность пребывания в стационаре 40 пациентов с хроническим панкреатитом (в койко-днях): 11, 13, 25, 12, 21, 21, 24, 23, 18, 15, 17, 24, 22, 23, 18, 17, 15, 12, 19, 22, 19, 21, 26, 23, 22, 11, 12, 14, 21, 24, 22, 23, 21, 10, 11, 13, 14, 20, 23, 18.

Вариант 3. Длительность временной утраты трудоспособности у 32 пациентов с диагнозом острое респираторное заболевание: 3, 6, 7, 6, 3, 4, 3, 11, 4, 5, 6, 7, 10, 8, 3, 4, 5, 3, 5, 9, 7, 3, 4, 6, 6, 3, 6, 7, 6, 3, 7, 8.

Вариант 4. Частота пульса (ударов в минуту) у 35 мужчин в возрасте 50 лет: 65, 78, 62, 64, 71, 75, 80, 64, 73, 72, 69, 78, 65, 67, 66, 65, 73, 68, 65, 72, 75, 82, 67, 64, 61, 67, 77, 79, 68, 63, 64, 67, 70, 73, 75.

Вариант 5. Число лиц, состоящих на диспансерном учете у 35 врачей общей практики: 80, 76, 85, 75, 79, 74, 78, 87, 88, 79, 76, 73, 82, 81, 80, 85, 76, 74, 81, 83, 78, 82, 83, 74, 79, 78, 77, 81, 83, 78, 72, 84, 81, 73, 75.

Вариант 6. Число лиц, прошедших обязательный медицинский осмотр перед устройством на работу у 32 врачей терапевтов за первую неделю месяца: 45, 37, 38, 39, 34, 37, 39, 43, 40, 36, 33, 43, 45, 39, 41, 44, 34, 32, 36, 37, 45, 43, 36, 35, 33, 42, 41, 38, 40, 32, 34, 41.

Вариант 7. Число лиц, впервые взятых на диспансерный учет 33 терапевтами участковыми в течение года: 30, 32, 31, 24, 28, 25, 26, 27, 29, 33, 32, 25, 28, 21, 23, 35, 34, 31, 30, 28, 20, 25, 26, 22, 23, 25, 30, 32, 32, 33, 34, 27, 25.

Вариант 8. Число лиц, состоящих на диспансерном учете у 36 невропатологов центральной поликлиники: 32, 24, 28, 26, 27, 29, 32, 28, 23, 34, 30, 28, 20, 22, 25, 32, 32, 34, 25, 32, 37, 24, 29, 30, 31, 28, 22, 27, 33, 34, 29, 27, 26, 27, 32, 34.

Вариант 9. Длительность лечения (в днях) 32 стационарных пациентов с диагнозом острая пневмония: 14, 21, 24, 27, 25, 22, 21, 20, 16, 14, 12, 13, 15, 18, 21, 14, 24, 18, 17, 22, 14, 16, 13, 12, 15, 12, 12, 14, 16, 19, 22, 17.

Вариант 10. Масса тела 35 новорожденных мальчиков (в кг): 2,5; 2,3; 2,2; 2,5; 2,2; 3,1; 3,5; 3,2; 2,3; 3,7; 3,1; 2,7; 2,3; 2,6; 2,4; 2,5; 3,2; 3,1; 2,5; 2,2; 2,5; 3,6; 3,5; 3,8; 2,7; 2,4; 2,8; 2,9; 2,5; 3,4; 3,1; 2,6; 3,6; 3,7; 2,7.

Вариант 11. Длина тела 33 новорожденных девочек (в см): 53, 48, 49, 54, 53, 51, 47, 49, 50, 54, 50, 53, 49, 49, 50, 49, 54, 55, 49, 52, 49,54, 48, 56, 55, 48, 47, 50, 51, 47, 52, 54, 49.

Вариант 12. Рост 40 студентов первого курса медицинского университета (в см): 185, 176, 190, 184, 175, 178, 181, 179, 177, 184, 188, 177, 178, 182, 188, 174, 176, 175, 184, 182, 191, 189, 185, 178, 180, 168, 180, 179, 164, 172, 182, 173, 185, 176, 174, 179, 181, 180, 177, 168.

Вариант 13. Масса тела 34 абитуриенток медицинского колледжа (в кг): 58,2; 54,6; 60,2; 71,4; 70,5; 66,8; 65,5; 61,5; 56,5; 57,3; 60,8; 72,4; 76,3; 72,1; 69,4; 67,2; 69,7; 72,2; 74,1; 66,8; 66,3; 69,2; 59,7; 55,4; 69,5; 71,4; 58,3; 55,9; 60,3; 58,9; 57,4; 55,7; 58,2; 63,7.

Вариант 14. Температура тела у 32 стационарных пациентов терапевтического профиля: 37,4; 36,6; 36,9; 37,1; 38,2; 37,7; 37,4; 37,8; 38,1; 38,3; 37,9; 36,6; 36,5; 36,9; 37,7; 37,8; 36,4; 36,6; 36,8; 36,6; 38,5; 37,8; 37,4; 38,2; 38,4; 39,2; 36,6; 37,8; 36,4; 37,4; 37,5; 36,2.

ЗАДАНИЕ № 3. СОСТАВЛЕНИЕ СГРУППИРОВАННОГО ВАРИАЦИОННОГО РЯДА, РАСЧЕТ СРЕДНЕГО АРИФМЕТИЧЕСКОГО ПО СПОСОБУ МОМЕНТОВ И ОПРЕДЕЛЕНИЕ СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ

Типовое задание *(Вариант 1)*

Длительность лечения 37 амбулаторных пациентов (в днях):

4;	5;	1;	13;	17;	8;	7;	9;	16;	10;
12;	2;	12;	10;	11;	14;	18;	4;	13;	3;
6;	13;	6;	9;	11;	16;	15;	7;	15;	8;
14;	10;	11;	17;	12;	9;	12.			

Образец выполнения типового задания

Расчеты следует выполнять в соответствии со следующим алгоритмом:

1. Построение сгруппированного вариационного ряда:

1.1. Определите число групп (g) по таблице 2.1.

В связи с тем, что количество вариант равно 37 необходимо выделить 6 групп.

1.2. Найдите интервал (классовый промежуток) по формуле:

$$\Delta x = \frac{X_{max} - X_{min}}{g} = \frac{18 - 1}{6} \approx 3$$

2. Вычисление средней арифметической по способу моментов:

2.1. Составьте сводную таблицу (таблица 2.3):

Таблица 2.3 Сводные данные по определению среднего арифметического по способу моментов

Δx	**Середина группы (c)**	**Частота (n)**	**Условное отклонение (a)**	$a_i n_i$	$a_i^2 n_i$
1 - 3	2	3	+3	+9	27
4 - 6	5	5	+2	+10	20
7 - 9	8	7	+1	+7	7
10 - 12	11	10	0	0	0
13 - 15	14	7	- 1	- 7	7
16 - 18	17	5	- 2	- 10	20
		N = 37		$\sum_{i=1}^{k} a_i n_i = 9$	$\sum_{i=1}^{k} a_i^2 n_i = 81$

2.2. Подставьте полученные данные в формулу
(в данном примере A = Mo = 12, при величине интервала Δx= 3):

$$X_в = A + \frac{\sum_{i=1}^{k} a_i n_i}{N} * \Delta x = 12 + \frac{9}{37} * 3 = 12{,}7 \text{ дня}$$

3. Определение среднего квадратического отклонения (по способу моментов):

$$\sigma = \Delta x * \sqrt{\frac{\sum_{i=1}^{N} a_i^2 n_i}{N} - \left(\frac{\sum_{i=1}^{N} a_i n_i}{N}\right)^2} = 3 * \sqrt{2{,}18 - 0{,}06} = 4{,}39$$

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории

1. Для вычисления среднего квадратического отклонения (стандартного отклонения) в электронных таблицах Excel используется встроенная статистическая функция **СТАНДОТКЛОН(A5:I8)**, параметрами которой могут быть массивы данных, отдельные ячейки, отдельные данные.

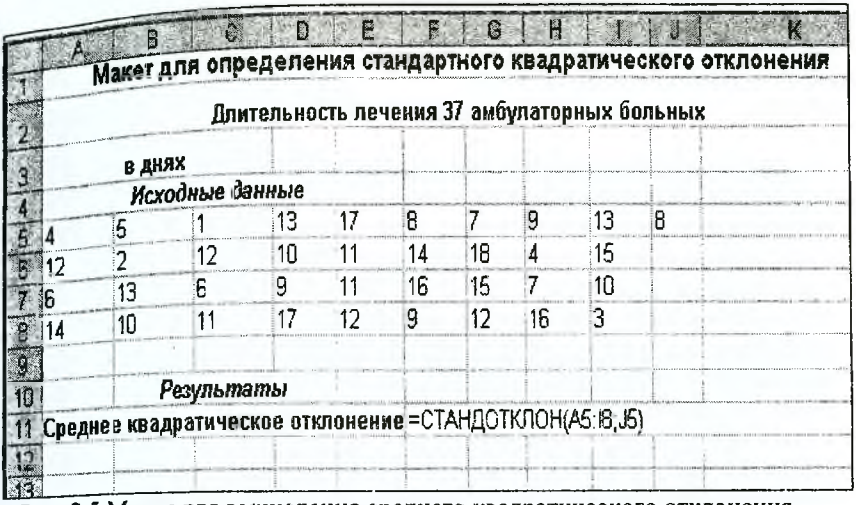
Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel в соответствии с макетом для обработки данных на примере задания варианта №1 (рис. 2.5).

2. Введите исходные данные в блок ячеек с адресами **(A5:I8, J5)**.

3. Вычислите среднее квадратическое отклонение, применив встроенную функцию **СТАНДОТКЛОН(A5:I8;J5)**.

4. Сравните результаты вычисления значений вручную, с результатами, полученными программным путем (рис. 2.6).



	A	B	C	D	E	F	G	H	I	J	K
1	Макет для определения стандартного квадратического отклонения										
2	Длительность лечения 37 амбулаторных больных										
3	в днях										
4	Исходные данные										
5	4	5	1	13	17	8	7	9	13	8	
6	12	2	12	10	11	14	18	4	15		
7	6	13	6	9	11	16	15	7	10		
8	14	10	11	17	12	9	12	16	3		
9											
10	Результаты										
11	Среднее квадратическое отклонение	=СТАНДОТКЛОН(A5:I8,J5)									
12											
13											

Рис. 2.5 Макет для вычисления среднего квадратического отклонения

	A	B	C	D	E	F	G	H	I	J
1	Макет для определения стандартного квадратического отклонения									
2	Длительность лечения 37 амбулаторных больных									
3	в днях									
4	Исходные данные									
5	4	5	1	13	17	8	7	9	13	8
6	12	2	12	10	11	14	18	4	15	
7	6	13	6	9	11	16	15	7	10	
8	14	10	11	17	12	9	12	16	3	
9										
10	Результаты									
11	Среднее квадратическое отклонение					4,4				
12										
13										
14										

Рис. 2.6 Результаты вычисления среднего квадратического отклонения

Варианты заданий на составление сгруппированного вариационного ряда, расчет среднего арифметического по способу моментов и определение среднего квадратического отклонения для самостоятельного выполнения возьмите из задания №2.

ТЕМА 3

ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

I. ЦЕЛЬ И ЗАДАЧИ ЗАНЯТИЯ

Цель: Изучить теоретические основы и практические методы оценки достоверности различий.

Задачи

1. Изучить теоретические основы выборочного метода.
2. Усвоить понятия «репрезентативность», «достоверность», «доверительный интервал», «доверительный коэффициент», «вероятность безошибочного прогноза».
3. Овладеть методикой расчета ошибок средних арифметических и относительных величин при малом и большом числе наблюдений.
4. Освоить методику определения доверительного интервала средних и относительных величин при малом и большом числе наблюдений.
5. Научиться определять степень вероятности безошибочного прогноза.
6. Освоить методику оценки достоверности различий между средними и относительными величинами выборок.

II. СТУДЕНТ ДОЛЖЕН ЗНАТЬ

1. Теоретические основы выборочного метода.
2. Определение понятий «репрезентативность», «достоверность», «доверительный интервал», «доверительный коэффициент», «вероятность безошибочного прогноза».

III. СТУДЕНТ ДОЛЖЕН УМЕТЬ

1. Рассчитывать ошибку средних величин (стандартную ошибку) при $N \leq 30$ и при $N > 30$.
2. Рассчитывать ошибку относительных величин при $N \leq 30$; $N > 30$.
3. Вычислять доверительный интервал средних величин при $N \leq 30$; $N > 30$.
4. Вычислять доверительный интервал относительных величин при $N \leq 30$; $N > 30$.
5. Определять достоверность различия между средними и относительными величинами в исследуемых выборках.

6. Подбирать по виду частотного распределения выборок применяемый статистический метод (параметрический или непараметрический).

IV. МОТИВАЦИЯ ТЕМЫ ЗАНЯТИЯ

Выборочный метод статистического наблюдения получил широкое распространение. Этот метод обеспечивает получение научно-обоснованных, точных результатов при использовании несплошного статистического наблюдения и является основой доказательной медицины. Особенно велика роль выборочного метода в медицинской статистике, так как медицинские работники имеют дело обычно только с частью изучаемого ими объекта (группа пациентов с тем или иным заболеванием; работой той или иной поликлиники или больницы; изменениями клинических показателей двух групп пациентов, лечившихся различными методами и т.д.).

Умение обоснованно подобрать объемы выборок, статистические методы для их обработки, сохранив при этом высокую степень объективности в оценке полученных результатов, - необходимое условие формирования не только клинического, но и аналитического образа мышления молодого специалиста.

V. ОСНОВНЫЕ ВОПРОСЫ ТЕМЫ

1. Выборочный метод, теоретические основы. Понятие о вероятности, репрезентативности, ошибке репрезентативности.
2. Понятие об оценке достоверности различия выборок.
3. Методика расчета показателей достоверности результатов исследования при малом и большом числе наблюдений параметрическими и непараметрическими методами (Хи-квадрат).
4. Популярные непараметрические методы оценки достоверности различий.

VI. МАТЕРИАЛЬНОЕ ОСНАЩЕНИЕ

1. Плакат с логической структурой темы - «Оценка достоверности результатов исследования».
2. Презентации с иллюстрациями методики вычисления ошибки средних и относительных величин, методики определения доверительных границ средних и относительных величин, методики определения достоверности различий между средними и относительными величинами.
3. Технические средства обучения: видео двойка MPEG 4 – DVD, персональный компьютер, мультимедийный проектор.

VII. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ

На лечебном факультете тема изучается в течение 6 академических часов, на стоматологическом факультете в течение 3 часов - и состоит из трех частей:

1. Вступительное слово преподавателя.

Во вступительном слове преподаватель обращает внимание студентов на роль статистических характеристик выборки в общественном здоровье и здравоохранении, знакомит студентов с основными группами статистических показателей. Рассматриваются методики расчета этих показателей.

2. Диагностика знаний студентов по теме: «Оценка достоверности результатов исследования».

Преподаватель выясняет исходный уровень знаний студентов по изучаемой теме.

3. Инструктаж по выполнению работы и контроль над ходом ее выполнения.

На примере типового задания преподаватель рассматривает методику оценки достоверности результатов исследования, особенности оформления выводов.

Далее студенты самостоятельно под контролем преподавателя выполняют варианты индивидуальных заданий по оценке достоверности результатов для конкретного исследования. Проверяют результаты выполнения индивидуальных заданий на ПК.

VIII. ИНФОРМАЦИОННЫЙ БЛОК ТЕМЫ

1. Выборочный метод, теоретические основы. Понятие о вероятности, репрезентативности, ошибке репрезентативности

Статистическим наблюдением можно охватить все члены изучаемой совокупности или ограничиться обследованием лишь некоторой их части. В первом случае наблюдение будет ***сплошным***, во втором – ***частичным*** или ***выборочным***. Несмотря на получение исчерпывающей информации в результате проведения сплошных исследований, к ним прибегают редко, так как эта работа может быть сопряжена с *большими материальными, временными и трудовыми затратами*, а также с практической *невозможностью и нецелесообразностью*.

Любое подмножество объектов генеральной совокупности называют ***выборочной совокупностью*** или ***выборкой - вариационным рядом,*** *если все элементы упорядочены по возрастанию. Элементы вариационного ряда называют вариантами.* Суть выборки состоит в том, что она, являясь частью генеральной совокупности, в определенной мере может характеризовать саму генеральную совокупность, т.е., обследуя часть объектов, можно сделать выводы обо всем их множестве.

Как правило, генеральная совокупность имеет достаточно большое, а в идеале и бесконечное количество элементов. *Чем больше объем выборки, тем более полно* она представляет генеральную совокупность.

В статистике под ***репрезентативностью*** понимают *главное свойство выборочной совокупности*, состоящее в близости ее характеристик к соответствующим характеристикам генеральной совокупности.

Различают *качественную* и *количественную* репрезентативность.

Качественная репрезентативность - обозначает *структурное соответствие* выборочной и генеральной совокупностей.

Количественная репрезентативность определяется *числом наблюдений, гарантирующих получение статистически достоверных данных*. Чем больше наблюдений, тем результаты достоверней.

Элементы выборки обычно *распределяются* в совокупности в соответствии с каким-то частотным законом (биноминальное, пуассоновское, нормальное распределение и др.).

Нормальное распределение (распределение Гаусса) – используется для *приближённого* описания явлений, которые носят *вероятностный, случайный характер,* когда на исследуемый процесс оказывает влияние *большое количество независимых факторов,* среди которых *нет доминирующего.* В связи с тем, что это справедливо для большинства медико-биологических явлений, они часто характеризуются нормальным частотным распределением признака, которое имеет вид кривой Гаусса (рис. 3.1).

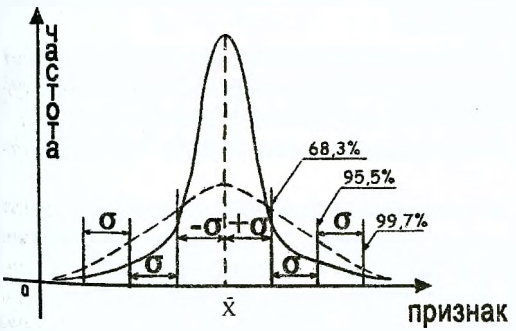


Рис. 3.1 Распределение Гаусса

Теоретическое обоснование выборочному методу дает ***математическая теория вероятности.***

Вероятностью (p) называют меру возможности возникновения каких-либо случайных событий в данных конкретных условиях.

Вероятность наступления в выборочной совокупности какого-либо события (p) определяется отношением наступивших событий (m) к числу всех возможных случаев (N):

$$p = \frac{m}{N}$$

В противоположность вероятности наступившего события различают ***вероятность отсутствия события (q)***:

$$q = \frac{N - m}{N} = 1 - \frac{m}{N} = 1 - p \quad \text{или} \quad q = 1 - p;\ p + q = 1$$

Таким образом, вероятность наступления события находится *в границах от 0 до 1*. Чем ближе вероятность события к единице, тем событие вероятнее, и, наоборот, чем ближе p к нулю, тем наступление события менее вероятно.

Теория вероятности обосновывает ***закон больших чисел***, который имеет следующие *важнейшие положения* для выборочного исследования:

1. По мере увеличения числа наблюдений результаты исследования, полученные на выборочной совокупности, *стремятся воспроизвести данные генеральной совокупности*;

2. При достижении определенного числа наблюдений в выборочной совокупности результаты исследования будут *максимально приближаться к данным генеральной совокупности*;

3. При большом количестве наблюдений законы частотного распределения количественных признаков *стремятся к нормальному частотному распределению Гаусса.*

Последнее положение называют ***предельной теоремой вселенной***.

Основные характеристики выборки, подчиняющейся нормальному закону частотного распределения*: среднее арифметическое (* $\bar{x}_{выб}$*), дисперсия (D), среднее квадратическое отклонение – стандартное отклонение (S).*

Величину отклонения исследуемых свойств выборки от их фактических значений в генеральной совокупности называют ***статистической ошибкой***, ***стандартной ошибкой*** или ***ошибкой репрезентативности (μ)***. Для измерения ошибки репрезентативности используют значение ***стандартного отклонения (среднего квадратического отклонения выборки) (S)*** – которое характеризует степень разброса признака от среднего в изучаемой выборочной совокупности. Стандартное отклонение для генеральной совокупности обозначают символом **σ**, а для выборки – символом **S**, и вычисляют по формуле:

$$S = \sqrt{D},$$

где D – дисперсия – параметр, так же как и стандартное отклонение, характеризующий степень разброса элементов выборки относительно среднего значения.

$$\text{при } N \leq 30, \quad D = \frac{\sum_{i=1}^{N} d_i^2}{N-1}$$

где d_i – является разностью между i-той вариантой ряда (x_i) и средним арифметическим ($\bar{x}$): $d_i = x_i - \bar{x}$,

N – общее количество вариант.

Геометрическая интерпретация **σ(*S – для выборки*)** представлена на рисунке 3.1. Точка $\bar{x}$ – это среднее. Интервал ($\bar{x}$ -σ; $\bar{x}$ +σ) включает варианты, отклоняющиеся от среднего не более чем на **σ.** В этом интервале при нормальном распределении оказывается 68,3% значений признака.

Таким образом, 68,3% всех вариант отклоняется от среднего значения не более чем на величину **σ.** В пределах **от -2σ до +2σ** лежат 95,5% всех вариант. В пределах от -3σ до +3σ находится **99,7%** - всех вариант («правило трёх сигм»). Один из *методов определения нормального частотного распределения* элементов выборки (метод Колмогорова-Смирнова) базируется на описанных выше соотношениях.

Из ***«правила трех сигм»*** ясно, что нормально распределенная случайная величина ***x*** в единичном испытании попадает в интервал ($\bar{x}$ -3σ; $\bar{x}$ +3σ). Вероятность этого события приближенно равна 0,9973:

$$p(\bar{x} - 3\sigma < \bar{x} < \bar{x} + 3\sigma) \approx 0{,}9973.$$

Другим способом определения соответствия частотного распределения выборки нормальному закону является следующий приблизительный критерий: *значения среднего, медианы и моды выборки должны быть примерно равными.*

2. Понятие об оценке достоверности результатов исследования

Под ***достоверностью*** статистических показателей следует понимать *степень их соответствия* отображаемой ими действительности. В теории вероятности это понятие, отражает *уверенность субъекта в правильности* наступления того или иного события.

При проведении научных исследований обычно используют *контрольную и исследуемую* группы выборок.

Контрольная группа – это выборка элементов из генеральной совокупности, которая отражает исходные свойства этой совокупности.

Исследуемая группа – это совокупность объектов, на которую осуществлялось воздействие каким-то фактором (новое лекарство, новая методика лечения и т.д.).

При выявлении достоверности различия этих выборок выдвигают ***статистическую гипотезу***, под которой понимают всякое высказывание о генеральной совокупности (случайной величине), проверяемое по выборке (результатам наблюдений).

Процедуру сопоставления высказанной гипотезы с выборочными данными называют ***проверкой*** статистической гипотезы.

Проверяемую статистическую гипотезу принято называть ***основной (нулевой) гипотезой (H_0)***, а противоречащую ей гипотезу – ***альтернативной (или конкурирующей) гипотезой (H_1)***.

По прикладному содержанию различают следующие основные виды высказываемых в ходе статистической обработки данных гипотез:

- о типе закона распределения исследуемой величины;
- об однородности двух или нескольких обрабатываемых выборок или некоторых характеристик анализируемых совокупностей;
- о числовых значениях (параметре) исследуемой генеральной совокупности;
- о типе зависимости между компонентами исследуемого многомерного признака;
- о независимости и стационарности обрабатываемого ряда наблюдений;
- другие.

Нулевая гипотеза (H_0) утверждает, что между выборками (например, между средними выборок или видами их частотных распределений и другими характеристиками) *нет статистически значимого различия* (это различие не достоверно), ***альтернативная (конкурирующая) гипотеза (H_1)*** утверждает, что это *различие статистически значимо* (достоверно).

При принятии решения справедлива ли нулевая гипотеза используют понятие ***уровень значимости (α)***, которое *характеризует допустимую ошибку* при отклонении нулевой гипотезы и принятии альтернативной. ***Уровень значимости*** – максимальное значение вероятности наступления события ***(H_0)***, при котором событие ещё считается практически невозможным.

В медицине уровень значимости принимают равным 0,05, что допускает довольно большую вероятность ошибки (5%) отклонения H_0. При более точных исследованиях (в фармакологии, токсикологии) уровни значимости принимают равными 0,01; 0,001.

Таким образом, чем *выше уровень значимости*, тем *меньше можно доверять утверждению, что различие между выборками существуют.*

Если вероятность (p) истинности H_0 меньше уровня значимости: $p_{H0} \leq \alpha$, то нулевую гипотезу отвергают и принимают H_1.

Например, принятие H_0 при уровне значимости α=0,01 означает, что эта гипотеза не противоречит наблюдаемым опытным выборочным данным. Для других выборок того же объема из той же генеральной совокупности принятая гипотеза в среднем будет справедлива в 99 случаях из 100. При этом, принимая H_0, мы рискуем ошибиться в среднем не чаще, чем в 1% случаев.

Уровень значимости *представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю генеральную совокупность.*

Уровень значимости находится в убывающей зависимости от ***надёжности результата*** **(γ): $\gamma = 1 - \alpha$.**

Более высокая статистическая значимость (α) соответствует более низкому уровню доверия (γ) к найденной в выборке характеристике.

Оценка достоверности - это ***определение вероятности безошибочного прогноза***, с которой результаты, полученные на основании выборочной совокупности, правомерно перенести на генеральную совокупность.

3. Методика расчета показателей достоверности результатов исследования при малом и большом числе наблюдений параметрическими и непараметрическими методами (Хи-квадрат)

Оценка достоверности результатов исследования предусматривает определение:

1). ***Ошибок репрезентативности*** (ошибок средних арифметических и относительных величин).

2). ***Доверительных границ*** средних (или относительных) величин (предполагается, что в пределах доверительных границ в генеральной совокупности будет находиться искомая величина – средняя или относительная).

3). ***Достоверности различия*** *средних (или относительных) величин выборок* по параметрическому критерию Стьюдента T.

4). ***Достоверности различия*** *видов частотных распределений сравниваемых выборок* по непараметрическому критерию χ^2.

5). ***Применение других непараметрических методов*** оценки достоверности различий.

3.1 Определение ошибок репрезентативности - ошибки средней величины (μ_x), относительной величины (μ_y) при малом и большом числе наблюдений

Как отмечалось выше, *ошибка репрезентативности* возникает в тех случаях, когда по выборке необходимо охарактеризовать генеральную совокупность в целом.

По величине ошибки репрезентативности определяют, насколько результаты, полученные при выборочном наблюдении, отлича-

ются от результатов, которые могли бы быть получены при проведении сплошного исследования всех без исключения элементов генеральной совокупности.

Ошибку репрезентативности можно свести к достаточно малой величине путем привлечения в выборку *достаточного количества* наблюдений (N).

Чтобы установить степень достоверности средней величины X (средняя длительность лечения, средняя масса тела и т.д.) или относительной величины Y (уровень заболеваемости, уровень летальности и т.д.) выборки соответствующим величинам генеральной совокупности, необходимо определить ***ошибку среднего*** (μ_x), или ***ошибку относительной величины*** (μ_y). Для этого надо знать *степень разнообразия признака (стандартное отклонение) в выборке (S) и количество объектов выборки (N):*

при $N \leq 30$: $\mu_{\bar{x}} = \frac{S}{\sqrt{N-1}}$; $\mu_y = \sqrt{\frac{p*q}{N-1}}$,

где p – вероятность признака, для которого определяется μ_y, q = 100 – p.

при $N > 30$: $\mu_{\bar{x}} = \frac{S}{\sqrt{N}}$; $\mu_y = \sqrt{\frac{p*q}{N}}$.

3.2 Определение доверительных границ средних и относительных величин

Для каждого исследования выбирается ***вероятность*** γ ($\gamma = 1 - \alpha$), называемая ***доверительной вероятностью*** (***надежностью***, ***вероятностью безошибочного прогноза).*** Она обычно близка к 1, при которой событие можно считать практически достоверным, например, 0,95; 0,99.

Интервал ($\bar{x} - \Delta$; $\bar{x} + \Delta$), отвечающий выбранной доверительной вероятности γ, называют ***доверительным интервалом***. Граничные значения ($\bar{x} - \Delta$ и $\bar{x} + \Delta$) – ***доверительными границами***.

Чем *больше число* γ, тем выше ***надежность***, т.е. полученная оценка является более достоверной. В то же время, для более точной оценки мы заинтересованы в *малой величине доверительного интервала.*

В случае нормального распределения доверительные интервалы оказываются *симметричными* относительно соответствующей точечной оценки параметра (среднего).

Истинное значение параметра (с надежностью γ) в генеральной совокупности находится где-то внутри доверительного интервала. Разность между точной оценкой среднего (медианы, моды) в генеральной совокупности и значением этого параметра в выборке не превышает длины интервала и называется ***ошибкой выборки*** (μ). Наибольшую из возможных ошибок называют ***предельной ошибкой выборки*** (Δ), она ха-

рактеризует *точность оценки*, равна половине доверительного интервала и находится по формуле:

$$\Delta = T^* \mu,$$

где μ - ошибка среднего выборки,

T – коэффициент Стьюдента, зависящий от надежности γ, равный ε_γ для больших выборок (его значения приведены ниже) или t_γ для малых выборок (представлен в таблице 3.1).

Таким образом, пользуясь таблицами для ε_γ и t_γ по известному значению надежности γ можно найти доверительные интервалы для средней или относительной величины, используя формулы:

Для **средней величины** - $\bar{x}_{ген} = \bar{x}_{выб} \pm T \cdot \mu_{\bar{x}}$,

Для **относительной величины** – $Y_{ген} = Y_{выб} \pm T \cdot \mu_y$,

где $\bar{x}_{ген}$ – средняя величина признака в генеральной совокупности,

$\bar{x}_{выб}$ – средняя величина, полученная при анализе выборочной совокупности,

$Y_{ген}$ – относительная величина в генеральной совокупности,

$Y_{выб}$ – относительная величина, полученная в результате анализа выборочной совокупности,

μ_y – ошибка относительной величины, $\mu_{\bar{x}}$ – ошибка средней величины,

T – доверительный коэффициент, зависящий от надежности и объема выборки.

При $N \leq 30$: $\alpha = 0{,}05$ (5%) $\Rightarrow$ $\gamma = 95\%$

$\alpha = 0{,}01$ (1%) $\Rightarrow$ $\gamma = 99\%$ $\quad T = t_\gamma$

$\alpha = 0{,}001$ (0,1%) $\Rightarrow$ $\gamma = 99{,}9\%$

Значение критерия t_γ находится по таблице значений коэффициента Стьюдента, в зависимости от числа наблюдений (таблица 3.1).

При N >30:

$\alpha = 0{,}05$ (5%) $\Rightarrow$ $\gamma = 95\%$ $\quad T = \varepsilon_\gamma = 1{,}96$

$\alpha = 0{,}01$ (1%) $\Rightarrow$ $\gamma = 99\%$ $\quad T = \varepsilon_\gamma = 2{,}7$

$\alpha = 0{,}001$ (0,1%) $\Rightarrow$ $\gamma = 99{,}9\%$ $\quad T = \varepsilon_\gamma = 3{,}6$

Таблица 3.1. Значения коэффициента Стьюдента ($T = t_\gamma$) при малом числе наблюдений (при $N \leq 30$)

N - 1	Уровень вероятности безошибочного прогноза		
	$\gamma = 95\%$	$\gamma = 99\%$	$\gamma = 99{,}9\%$
1	12,7	63,6	636,6
2	4,3	9,9	31,6
3	3,1	5,8	12,9
4	2,7	4,6	8,6

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

3.3 Параметрические методы определения достоверности различий средних (или относительных) величин (по критерию Т - Стьюдента)

При сопоставлении двух сравниваемых величин, полученных в результате исследования, возникает необходимость не только определить различие между ними, но и оценить достоверность этого различия.

Все методы определения достоверности различий делят на ***параметрические*** и ***непараметрические***.

Если выборки содержат количественные признаки, подчиняющиеся нормальному закону частотного распределения, либо в выборках используются количественные признаки с неизвестным распределением, но объемы выборок велики (по закону больших чисел их частотное распределение приближается к закону Гаусса), то используются *параметрические* методы ***(критерий Стьюдента)***. В противном случае, если признак количественный, но закон частотного распределения не соответствует распределению Гаусса или выборки содержат качественные признаки, применяют *непараметрические* методы.

Сравнение средних и относительных величин двух нормально распределенных совокупностей проводится ***по статистическому критерию Стьюдента***. При этом выдвигается H_0: «Средние двух выборок принадлежат одной и той же генеральной совокупности».

Оценка достоверности различий между средними (X_1 и X_2) или относительными (Y_1 и Y_2) величинами двух выборок определяется по формулам:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\mu^2_{\bar{x}_1} + \mu^2_{\bar{x}_2}}} \; ; \; T = \frac{Y_1 - Y_2}{\sqrt{\mu^2_{Y_1} + \mu^2_{Y_2}}}$$

где T – коэффициент Стьюдента,

$\bar{x}_1$ – среднее первой выборки, $\bar{x}_2$ - среднее второй выборки,

Y_1 – относительная величина первой выборки,

Y_2 – относительная величина второй выборки,

μ_y – ошибка относительной величины,

μ_X – ошибка среднего.

Различие сравниваемых средних или относительных величин ***статистически достоверно*** при коэффициенте $T \geq 2$ (при $\alpha = 0,05$ (5%), вероятности безошибочного прогноза $\gamma = 95\%$) – ***принимается альтернативная гипотеза.***

Если различие оказывается статистически незначимым **(T<2)**, делаем вывод о *совпадении сравниваемых средних или относительных величин*, т.е. ***принимается нулевая гипотеза*** о том, что обе *выборки извлечены из одной и той же генеральной совокупности.*

Выделяют несколько алгоритмов реализации критерия Стьюдента:

- Двухвыборочный Т-тест с одинаковыми дисперсиями;
- Двухвыборочный Т-тест с различными дисперсиями;
- Двухвыборочный парный тест.

3.4 Непараметрическая оценка достоверности различий сравниваемых групп по критерию согласия (хи-квадрат)

Методы проверки согласованности эмпирических данных с теоретическими о предполагаемом законе распределения называют ***критериями согласия.***

Оценка достоверности различия законов частотных распределений сравниваемых совокупностей может быть выполнена ***по критерию согласия (соответствия) Пирсона***, называемого также ***критерием хи-квадрат*** (χ^2).

Оценка достоверности различия между выборочными совокупностями с помощью χ^2 сводится к определению *соответствия эмпирического распределения теоретическому* (нулевая гипотеза).

Величина критерия χ^2 всегда *положительна*, так как отклонения эмпирических параметров от ожидаемых или вычисленных возведены в квадрат. Критерий χ^2 целесообразно применять в случаях:

1. Когда выборки содержат *количественные признаки, не подчиняющиеся нормальному закону* частотного распределения;

2. Когда *нет необходимости знать величину* того или иного параметра (среднюю величину или относительный показатель), не известны законы частотных распределений выборок, но требуется оценить достоверность различия *как двух, так и большего числа групп*.

Критерий χ^2 определяется по формуле:

$$\chi^2 = \frac{\sum_{i=1}^{N}(\varphi - \varphi_1)^2}{\varphi_1},$$

где χ^2 - критерий хи-квадрат,

φ - фактические (эмпирические) данные,

φ_1 - «ожидаемые» (теоретические) данные, вычисленные на основании нулевой гипотезы (предположения о том, что в сравниваемых группах отсутствует различие в распределении частот).

Определение χ^2 основано на вычислении *разницы между фактическими и «ожидаемыми» данными*. Чем она выше, тем с большей вероятностью можно считать, что существуют различия в частотных распределениях признака в сравниваемых выборочных совокупностях.

4. Популярные непараметрические методы оценки достоверности различий

Применение ***параметрических*** методов (Т-критерий Стьюдента) правомерно при *нормальном распределении количественного признака* или *большом объеме выборки*. Если же имеются *значительные отличия* распределений признака от нормального или исследуемый признак является *качественным*, или варьирующие признаки выражаются не числами, а *условными знаками* (порядковые номера, индексы), то необходимо использовать ***непараметрические*** критерии, в основе которых лежит упорядочивание (ранжирование) значений по отношению друг к другу, типа «больше-меньше» или «лучше-хуже». Это разграничение значений не предполагает точных количественных соотношений, ограничений на параметры и вид распределения, т.е. для использования непараметрических критериев нужно *меньше информации*, чем для использования параметрических.

При этом сравниваются не члены ранжированных рядов, а их ***ранги***, т.е. порядковые номера в ранжированных рядах. Наиболее популярными ранговыми критериями, применяемыми в статистических исследованиях, являются критерий Манна-Уитни (критерий однородности – непараметрический аналог непарного критерия Стьюдента), критерий Уилкоксона (непараметрический аналог парного критерия Стьюдента), F- критерий Фишера, Х –критерий Ван-дер-Вардена, критерий знаков Z.

В качестве примера рассмотрим алгоритмы реализации критерия знаков Z и критерия Уилкоксона.

Когда результаты исследования выражены не числами, а знаками «плюс» или «минус», для оценки различий между попарно связанными членами сравниваемых выборок можно применить **критерий знаков Z.** Смысл критерия Z сводится к следующему: если попарно связанные значения двух зависимых выборок не отличаются друг от друга, то количество положительных и отрицательных отличий окажется одинаковым. Если обнаружится значимое преобладание плюсовых, или минусовых разностей, то имеется положительное или отрицательное действие изучаемого фактора на результативный признак.

Аналогичную задачу сравнение двух зависимых выборок (при связи их членов общими условиями) можно решить с помощью **рангового критерия Уилкоксона**. По мощности этот критерий превосходит критерий знаков Z. Методика критерия Уилкокосона заключается в следующем: попарные разности, как положительные, так и отрицательные ранжируют в один общий ряд. Нулевые разности в расчет не принимаются, а все остальные независимо от знака распределяют так, чтобы наименьшая разность получила первый ранг. Затем находят сумму положительных и сумму отрицательных разностей. Меньшую сумму (без учета ее знака) используют в качестве фактически установленной величины критерия.

IX. УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА СТУДЕНТОВ

ЗАДАНИЕ №1. ОПРЕДЕЛЕНИЕ СТАНДАРТНОЙ ОШИБКИ СРЕДНЕГО И ГРАНИЦ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ПРИ МАЛОМ ЧИСЛЕ НАБЛЮДЕНИЙ

Типовое задание *(Вариант 1)*

Используя данные о частоте дыхательных движений 7 взрослых пациентов, проходящих профилактический осмотр в районной поликлинике, вычисляем ошибку среднего арифметического.

Частота дыхания: 12, 18, 16, 13, 15, 14, 17.

Образец выполнения типового задания

1. Строим простой возрастающий вариационный ряд:
   12, 13, 14, 15, 16, 17, 18.
2. Рассчитываем простое среднее арифметическое:

$$X = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{12+13+14+15+16+17+18}{7} = \frac{105}{7} = 15$$

3. Вычисляем среднее квадратическое отклонение:

$$\text{при } N \leq 30, \quad S = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N-1}}$$

где d_i – является разностью между i-той вариантой ряда (x_i) и средним арифметическим (X): $d_i = x_i - X$□, N – общее количество вариант.

$$S = \sqrt{\frac{\sum_{i=1}^{7}(x_i - X)^2}{6}} =$$

$$= \sqrt{\frac{(12-15)^2 + (13-15)^2 + (14-15)^2 + (15-15)^2 + (16-15)^2 + (17-15)^2 + (18-15)^2}{6}} = 2,16$$

4. Рассчитываем ошибку среднего арифметического:

при $N \leq 30$ $\mu x = \frac{S}{\sqrt{N-1}} = \frac{2,16}{\sqrt{6}} = \frac{2,16}{2,45} = 0,88$

5. Вычисляем доверительный интервал с использованием таблицы значений критерия Стьюдента (таблица 3.1).

5.1. При $\alpha = 0,05$ (5%) $\Rightarrow \gamma = 95\%$ и числе наблюдений N = 7, критерий $T = t_\gamma = 2,4$, тогда: $\Delta = T^*\mu = 2,4 \cdot 0,88 = 2,11$

Таким образом, доверительные интервалы средней величины (X-Δ и X+Δ): $X = 15 \pm 2,11$ или средняя частота дыханий в минуту колеблется от 12,89 до 17,11 $\Rightarrow$ X = [12,89; 17,11].

5.2. При $\alpha = 0,01$ (5%) $\Rightarrow \gamma = 99\%$ и числе наблюдений N = 7, критерий $T = t_\gamma = 3,7$, тогда: $\Delta = T^*\mu = = 3,7 \cdot 0,88 = 3,26$

Следовательно, доверительные интервалы средней величины - $X = 15 \pm 3,26$ или средняя частота дыханий в минуту колеблется от 11,74 до 18,26 $\Rightarrow$ X = [11,74; 18,26].

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel.

Сведения из теории

1. Корень числа X позволяет вычислить математическая функция **КОРЕНЬ(X);**

2. Для точного вычисления коэффициента Стьюдента T можно использовать встроенную статистическую функцию Excel **СТЬЮДРАСПРОБР(α;n)**, где α – уровень значимости (в долях), соответствующий двустороннему критерию Стьюдента; n – число степеней свободы (количество вариант в выборочной совокупности минус единица).

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта 1 (рис. 3.2).

2. Введите исходные данные в блок ячеек с адресами **(A4:A10)**.

3. Скопируйте данные блока ячеек **(A4:A10)** в блок ячеек с начальным адресом **C4**.

4. Выполните сортировку данных по возрастанию в блоке ячеек **(C4:C10)**.

5. Рассчитайте среднее арифметическое ($\bar{x}$), применив встроенную функцию **СРЗНАЧ(C4:C10)**.

6. Рассчитайте стандартное отклонение – S с помощью функции **СТАНДОТКЛОН(C4:C10)**.

7. Вычислите стандартную ошибку среднего арифметического ($\mu_{\bar{x}}$) по формуле **S/КОРЕНЬ(6)**.

8. Определите коэффициент Стьюдента T, применив функцию **СТЬЮДРАСПРОБР(0,05;6)** или **СТЬЮДРАСПРОБР(0,01;6)**, где первый аргумент **0,05** или **0,01** – уровень значимости, соответствующий двустороннему критерию Стьюдента; **6** – число степеней свободы (равно количество вариант минус единица).

9. Вычислите половину доверительного интервала Δ (предельную ошибку среднего), как $T*\mu_{\bar{x}}$.

10. Определите границы доверительного интервала для среднего: ($\bar{x}$ -Δ; $\bar{x}$ +Δ).

	A	B	C	D
1	Макет для расчета стандартной ошибки среднего и границ доверительного интервала среднего			
2		Частота дыхания пациентов		
3	*Исходные данные*		*Результаты*	
4	12		12	
5	18		13	
6	16		14	
7	13		15	
8	15		16	
9	14		17	
10	17		18	
11		Среднее	=СРЗНАЧ(C4:C10)	
12		Стандартное отклонение	=СТАНДОТКЛОН(C4:C10)	
13		Стандартная ошибка среднего	=C12/КОРЕНЬ(6)	
14		Коэффициент Стьюдента (T)	=СТЬЮДРАСПОБР(0,05;6)	
15		Коэффициент Стьюдента (T)	=СТЬЮДРАСПОБР(0,01;6)	
16	Доверительный интервал	Предельная ошибка среднего (Δ)	=C14*C13	α=0,05
17		Нижняя граница доверительного интервала	=C11-C16	
18		Верхняя граница доверительного интервала	=C11+C16	
19		Предельная ошибка среднего (Δ)	=C15*C13	α=0,01
20		Нижняя граница доверительного интервала	=C11-C19	
21		Верхняя граница доверительного интервала	=C11+C19	
22				

11. Проанализируйте результаты (результаты вычислений для варианта 1 приведены на рис. 3.3).

Примечание: *Некоторые расхождения с расчетными значениями объясняются более точными расчетами на ПК.*

Рис. 3.2 Макет для определения границ доверительного интервала среднего при малом числе наблюдений

	A	B	C	D	E	G
1	Макет для расчета стандартной ошибки среднего и границ доверительного интервала среднего					
2		Частота дыхания пациентов				
3	*Исходные данные*		*Результаты*			
4	12		12			
5	18		13			
6	16		14			
7	13		15			
8	15		16			
9	14		17			
10	17		18			
11		Среднее	15			
12		Стандартное отклонение	2.16			
13		Стандартная ошибка среднего	0.88			
14		Коэффициент Стьюдента (T)	2.45	α=0,05		
15		Коэффициент Стьюдента (T)	3.71	α=0,01		
16	Доверительный	Предельная ошибка среднего (Δ)	**2.16**			
17	интервал	Нижняя граница доверительного интервала	**12.84**	α=0,05		
18		Верхняя граница доверительного интервала	**17.16**			
19		Предельная ошибка среднего (Δ)	**3.27**			
20		Нижняя граница доверительного интервала	**11.73**	α=0,01		
21		Верхняя граница доверительного интервала	**18.27**			
22						

Рис. 3.3. Результаты расчета границ доверительного интервала среднего арифметического

Варианты заданий на определение стандартной ошибки и доверительных границ интервала среднего при малом числе наблюдений для самостоятельного выполнения

Вариант 2. Частота пульса (ударов в минуту) в состоянии покоя у 7 спортсменок занимающихся легкой атлетикой: 68, 70, 67, 75, 72, 69, 65.

Вариант 3. Численность населения в 5 районах области: 32150, 15000, 18542, 12385, 15795.

Вариант 4. Численность населения в четырех районах областного центра: 120000, 85000, 110000, 78000.

Вариант 5. Рост 10 школьников 9 класса средней школы (в см): 180, 175, 177, 174, 181, 167, 172, 182, 164, 173.

Вариант 6. Масса тела (в кг) 6 мужчин проходивших профосмотр при приеме на работу: 65, 80, 74, 93, 70, 105.

Вариант 7. Результат измерения вечерней температуры (в °С) у 5 стационарных пациентов терапевтического отделения: 37,1; 36,8; 37,5; 38,4; 36,6.

Вариант 8. Число пациентов состоящих на диспансерном учете у 7 врачей общей практики: 122, 134, 131, 127, 143, 148, 129.

Вариант 9. Число семей, обслуживаемых 5 врачами общей практики в амбулаториях врача общей практики: 354, 371, 402, 368, 387.

Вариант 10. Число детей до года, состоящих под наблюдением у 9 педиатров участковых детской поликлиники: 62, 54, 57, 52, 60, 53, 55, 61, 63.

Вариант 11. Число обращений с травмами в травматологический пункт районного центра в первые 7 дней месяца: 127, 145, 142, 134, 128, 132, 128.

Вариант 12. Число студентов, прошедших медицинский осмотр за каждый день первой недели учебного года: 12, 15, 10, 14, 22, 25, 18

Вариант 13. Уровень систолического артериального давления (мм рт. ст.) у 5 мужчин 45-летнего возраста: 130, 110, 135, 120, 140.

Вариант 14. Частота дыхания (дыхательных движений в минуту) у 5 мужчин 30-летнего возраста после 20 приседаний: 27, 23, 24, 28, 25.

ЗАДАНИЕ №2. ОПРЕДЕЛЕНИЕ СТАНДАРТНОЙ ОШИБКИ И ГРАНИЦ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА СРЕДНЕГО ПРИ БОЛЬШОМ ЧИСЛЕ НАБЛЮДЕНИЙ

Типовое задание

Используя данные о длительности временной нетрудоспособности пациентов районной поликлиники, определите стандартную ошибку среднего арифметического и его доверительные границы.

Длительность нахождения на бюллетене: 9, 1, 13, 8, 2, 7, 5, 11, 7, 14, 3, 4, 10, 9, 5, 6, 15, 7, 1, 8, 9, 7, 10, 8, 6, 11, 8, 8, 12, 8, 9, 6, 13, 12, 14, 4, 15, 16, 5, 10, 5, 17, 7, 18, 11.

Образец выполнения типового задания

1. Определяем число групп по таблице 2.1 (тема 2):

при N=45 число групп (g) соответствует 6.

2. Рассчитываем интервал:

$$\Delta x = \frac{x_{max} - x_{min}}{g} = \frac{18-1}{6} = 2,83 \approx 3$$

3. Определяем моду (Мо): наиболее часто встречается варианта со значением 8. Значит, Мо = 8.

Достаточно часто Мо равна условной средней (А). Значит, Мо = А = 8.

4. Составляем сгруппированный вариационный ряд, находим середины групп, частоту их встречаемости, условное отклонение в интервалах и его произведение на частоту. Ранее аналогичные расчеты приводились в теме 2.

Длительность нахождения на бюллетене	Середина группы (c)	Частота встречаемости вариант интервала (n)	Условное отклонение (a_i) $a_i = (c - A)/\Delta x$	$a_i n_i$	$a_i^2 n_i$
1-3	2	4	(2-8)/3 = -2	(-2)*4 = -8	$(-2)^2$*4=16
4-6	5	9	(5-8)/3 = -1	(-1)*9 = -9	$(-1)^2$*9=9
7-9	8	15	(8-8)/3 = 0	0*15 = 0	0*15=0
10-12	11	8	(11-8)/3 = 1	1*8 =8	1^2*8=8
13-15	14	6	(14-8)/3 = 2	2*6=12	2^2*6=24
16-18	17	3	(17-8)/3 = 3	3*3=9	3^2*3=27

N = 45;

$\sum_{i=1}^{k} a_i n_i = 12$; $\sum_{i=1}^{k} a_i^2 n_i = 84$

5. Определяем взвешенную среднюю арифметическую по способу моментов:

$$X_в = A + \frac{\sum_{i=1}^{k} a_i n_i}{N} * \Delta x = 8 + \frac{12}{45} * 3 = 8,8$$

6. Находим среднее квадратическое отклонение:

$$S = \Delta x * \sqrt{\frac{\sum_{i=1}^{N} a_i^2 n_i}{N} - \left(\frac{\sum_{i=1}^{N} a_i n_i}{N}\right)^2} =$$

$$= 3 * \sqrt{\frac{84}{45} - \left(\frac{12}{45}\right)^2} = 4{,}02$$

7. Рассчитываем стандартную ошибку среднего арифметического:

при N >30 $\quad \mu_{\bar{x}} = \frac{S}{\sqrt{N}} = \frac{4{,}02}{\sqrt{45}} = 0{,}6$

8. Находим доверительный интервал с учетом того, что при

$$\left.\begin{array}{ll} \alpha = 0{,}05\ (5\%) \Rightarrow \gamma = 95\% & T = \varepsilon_{95} = 1{,}96 \approx 2 \\ \alpha = 0{,}01\ (1\%) \Rightarrow \gamma = 99\% & T = \varepsilon_{99} = 2{,}7 \approx 3 \end{array}\right\}$$

$\Delta = \varepsilon_{95} * \mu = 2 \cdot 0{,}6 = 1{,}2$

Таким образом, доверительные интервалы **средней величины** ($\bar{x} - \Delta$ и $\bar{x} + \Delta$) с вероятностью 95%: $X = 8{,}8 \pm 1{,}2$ или средняя длительность нахождения на бюллетене от 7,6 до 10,0 или от 8 до 10 дней.

$\Delta = \varepsilon_{99} * \mu = 3 \cdot 0{,}6 = 1{,}8$

Значит, доверительные интервалы **средней величины** ($X - \Delta$ и $X + \Delta$) с вероятностью 99%: $X = 8{,}8 \pm 1{,}8$ или средняя длительность нахождения на бюллетене от 7,0 до 10,6 или от 7 до 11 дней.

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта 1 (рис. 3.4).

2. Введите исходные данные в блок ячеек с адресами **(A6:A51)**.

3. Скопируйте данные блока ячеек **(A6:A51)** в блок ячеек с адреса **B6**.

4. В блоке ячеек **(B6:B51)** выполните сортировку данных по возрастанию.

5. Рассчитайте среднее арифметическое ($\bar{x}$), применив встроенную функцию **СРЗНАЧ(B6:B51)**.

6. Рассчитайте среднее квадратическое отклонение – S (стандартное отклонение) с помощью функции **СТАНДОТКЛОН(B6:B51)**.

7. Вычислите стандартную ошибку среднего арифметического ($\mu_{\bar{x}}$) по формуле **S/КОРЕНЬ(45)**.

8. Определите коэффициент Стьюдента T, применив функцию **СТЬЮДРАСПРОБР(0,01;44)**, для разных значений уровня значимости (0,01 и 0,05).

9. Вычислите половину доверительного интервала Δ (предельную ошибку среднего), как T*μ $\bar{x}$ для разных значений уровня значимости.

10. Определите границы доверительного интервала для среднего($\bar{x}$ -Δ; $\bar{x}$ +Δ) для двух заданных уровней значимости.

	A	B	C	D	E
1	Макет для расчета стандартной ошибки среднего и границ доверительного интервала среднего				
3	Длительность амбулаторного лечения больных				
4			Результаты		
5	*Исходные данные*	*сортировка*			
6	9	1			
7	1	1			
8	13	2			
9	11	3			
10	8	4			
11	2	4			
12	7	5			
13	5	5			
14	11	5			
15	7	5			
16	14	6			
17	3	6			
18	4	6	Мода	=МОДА(B6:B51)	
19	10	7	Среднее арифметическое	=СРЗНАЧ(B6:B51)	
20	9	7	Стандартное отклонение	=СТАНДОТКЛОН(B6:B51)	
21	5	7	Стандартная ошибка среднего	=D20/КОРЕНЬ(45)	
22	6	7	Коэффициент Стьюдента (T)	=СТЬЮДРАСПОБР(0,05;44)	α=0,05
23	15	7	Коэффициент Стьюдента (T)	=СТЬЮДРАСПОБР(0,01;44)	α=0,01
24	7	8	Предельная ошибка среднего (Δ)	=D22*D21	
25	1	8	Нижняя граница доверительного инт	**=D19-D24**	α=0,05
26	8	8	Верхняя граница доверительного ин	**=D19+D24**	
27	9	8	Предельная ошибка среднего (Δ)	=D21*D23	
28	7	8	Нижняя граница доверительного инт	**=D19-D27**	α=0,01
29	10	8	Верхняя граница доверительного ин	**=D19+D27**	
30	8	9			
31	6	9			

Рис. 3.4 Фрагмент листа Excel с макетом для определения границ доверительного интервала среднего при большом числе наблюдений

11. Проанализируйте полученные результаты. Результаты вычислений для варианта 1 приведены на рис. 3.5.

Примечание: *Некоторые расхождения с расчетными значениями объясняются более точными расчетами на ПК.*

	A	B	C	D	E
1	Макет для расчета стандартной ошибки и границ доверительного интервала среднего				
3	Длительность амбулаторного лечения больных				
4	*Исходные*	*Результаты*			
5	*данные*	*сортировки*			
6	9	1			
7	1	1			
8	13	2			
9	11	3			
10	8	4			
11	2	4			
12	7	5			
13	5	5			
14	11	5			
15	7	5			
16	14	6			
17	3	6			
18	4	6	Мода	8	
19	10	7	Среднее арифметическое	8.8	
20	9	7	Стандартное отклонение	4.12	
21	5	7	Стандартная ошибка среднего	0.61	
22	6	7	Коэффициент Стьюдента (Т)	2.02	α=0,05
23	15	7	Коэффициент Стьюдента (Т)	2.69	α=0,01
24	7	8	Предельная ошибка среднего (Δ)	1.24	
25	1	8	Нижняя граница доверительного инт	**7.57**	α=0,05
26	8	8	Верхняя граница доверительного ин	**10.04**	
27	9	8	Предельная ошибка среднего (Δ)	1.65	
28	7	8	Нижняя граница доверительного инт	**7.15**	α=0,01
29	10	8	Верхняя граница доверительного ин	**10.46**	
30	8	9			

Рис. 3.5. Результаты расчета границ доверительного интервала среднего арифметического

Варианты заданий на определение стандартной ошибки среднего и его доверительных границ для самостоятельного выполнения

Вариант 2. Длительность временной утраты трудоспособности у 32 пациентов с диагнозом острое респираторное заболевание: 3, 6, 7, 6, 3, 4, 3, 11, 4, 5, 6, 7, 10, 8, 3, 4, 5, 3, 5, 9, 7, 3, 4, 6, 6, 3, 6, 7, 6, 3, 7, 8.

Вариант 3. Частота пульса (ударов в минуту) у 35 мужчин в возрасте 50 лет: 65, 78, 62, 64, 71, 75, 80, 64, 73, 72, 69, 78, 65, 67, 66, 65, 73, 68, 65, 72, 75, 82, 67, 64, 61, 67, 77, 79, 68, 63, 64, 67, 70, 73, 75.

Вариант 4. Число лиц, состоящих на диспансерном учете у 35 врачей общей практики: 80, 76, 85, 75, 79, 74, 78, 87, 88, 79, 76, 73, 82, 81, 80, 85, 76, 74, 81, 83, 78, 82, 83, 74, 79, 78, 77, 81, 83, 78, 72, 84, 81, 73, 75.

Вариант 5. Число лиц, прошедших обязательный медицинский осмотр перед устройством на работу у 32 врачей терапевтов за первую неделю месяца: 45, 37, 38, 39, 34, 37, 39, 43, 40, 36, 33, 43, 45, 39, 41, 44, 34, 32, 36, 37, 45, 43, 36, 35, 33, 42, 41, 38, 40, 32, 34, 41.

Вариант 6. Число лиц, взятых на диспансерный учет 33 терапевтами участковыми в течение года: 30, 32, 31, 24, 28, 25, 26, 27, 29, 33, 32, 25, 28, 21, 23, 35, 34, 31, 30, 28, 20, 25, 26, 22, 23, 25, 30, 32, 32, 33, 34, 27, 25.

Вариант 7. Число лиц, состоящих на диспансерном учете у 36 невропатологов центральной поликлиники: 32, 24, 28, 26, 27, 29, 32, 28, 23, 34, 30, 28, 20, 22, 25, 32, 32, 34, 25, 32, 37, 24, 29, 30, 31, 28, 22, 27, 33, 34, 29, 27, 26, 27, 32, 34.

Вариант 8. Длительность лечения (в днях) 32 стационарных пациентов с диагнозом острая пневмония: 14, 21, 24, 27, 25, 22, 21, 20, 16, 14, 12, 13, 15, 18, 21, 14, 24, 18, 17, 22, 14, 16, 13, 12, 15, 12, 12, 14, 16, 19, 22, 17.

Вариант 9. Масса тела 35 новорожденных мальчиков (в кг): 2,5; 2,3; 2,2; 2,5; 2,2; 3,1; 3,5; 3,2; 2,3; 3,7; 3,1; 2,7; 2,3; 2,6; 2,4; 2,5; 3,2; 3,1; 2,5; 2,2; 2,5; 3,6; 3,5; 3,8; 2,7; 2,4; 2,8; 2,9; 2,5; 3,4; 3,1; 2,6; 3,6; 3,7; 2,7.

Вариант 10. Длина тела 33 новорожденных девочек (в см): 53, 48, 49, 54, 53, 51, 47, 49, 50, 54, 50, 53, 49, 49, 50, 49, 54, 55, 49, 52, 49,54, 48, 56, 55, 48, 47, 50, 51, 47, 52, 54, 49.

Вариант 11. Рост 40 студентов первого курса медицинского университета (в см): 185, 176, 190, 184, 175, 178, 181, 179, 177, 184, 188, 177, 178, 182, 188, 174, 176, 175, 184, 182, 191, 189, 185, 178, 180, 168, 180, 179, 164, 172, 182, 173, 185, 176, 174, 179, 181, 180, 177, 168.

Вариант 12. Масса тела 34 абитуриенток медицинского колледжа (в кг): 58,2; 54,6; 60,2; 71,4; 70,5; 66,8; 65,5; 61,5; 56,5; 57,3; 60,8; 72,4; 76,3; 72,1; 69,4; 67,2; 69,7; 72,2; 74,1; 66,8; 66,3; 69,2; 59,7; 55,4; 69,5; 71,4; 58,3; 55,9; 60,3; 58,9; 57,4; 55,7; 58,2; 63,7.

Вариант 13. Температура тела у 32 стационарных пациентов терапевтического профиля: 37,4; 36,6; 36,9; 37,1; 38,2; 37,7; 37,4; 37,8;

38,1; 38,3; 37,9; 36,6; 36,5; 36,9; 37,7; 37,8; 36,4; 36,6; 36,8; 36,6; 38,5; 37,8; 37,4; 38,2; 38,4; 39,2; 36,6; 37,8; 36,4; 37,4; 37,5; 36,2.

Вариант 14. Длительность пребывания в стационаре 40 пациентов с хроническим панкреатитом (в койко-днях): 11, 13, 25, 12, 21, 21, 24, 23, 18, 15, 17, 24, 22, 23, 18, 17, 15, 12, 19, 22, 19, 21, 26, 23, 22, 11, 12, 14, 21, 24, 22, 23, 21, 10, 11, 13, 14, 20, 23, 18.

ЗАДАНИЕ № 3. ОПРЕДЕЛЕНИЕ ДОСТОВЕРНОСТИ РАЗЛИЧИЯ МЕЖДУ СРЕДНИМИ ВЕЛИЧИНАМИ ВЫБОРОК С ИСПОЛЬЗОВАНИЕМ ПАРАМЕТРИЧЕСКОГО КРИТЕРИЯ СТЬЮДЕНТА

Типовое задание (*Вариант 1*)

Используя данные о величине систолического давления у 10 студентов до и после занятий физической культурой, требуется определить достоверно ли повышение уровня систолического давления после занятия.

Систолическое давление до занятий (в мм рт. ст.)	115	120	120	125	120	130	125	120	115	120
Систолическое давление после занятий (в мм рт. ст.)	150	145	155	140	140	145	150	160	150	150

Образец выполнения типового задания

1. Определяем среднее арифметическое первого ряда:

$$\bar{x}_1 = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{115+120+120+125+120+130+125+120+115+120}{10} = 121$$

2. Определяем среднее арифметическое второго ряда:

$$\bar{x}_2 = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{150+145+155+140+140+145+150+160+150+150}{10} = 148,5$$

3. Рассчитываем среднее квадратическое отклонение в первом ряду (при $N \leq 30$):

$$S_1 = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N-1}} = \sqrt{\frac{\sum_{i=1}^{10} (\bar{x}_i - \bar{x}_1)^2}{9}} = \sqrt{\frac{190}{9}} = 4,6$$

4. Рассчитываем среднее квадратическое отклонение в втором ряду (при $N \leq 30$):

$$S_2 = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N-1}} = \sqrt{\frac{\sum_{i=1}^{10} (x_i - \bar{x}_2)^2}{9}} = \sqrt{\frac{355}{9}} = 6,3$$

5. Определяем стандартную ошибку среднего арифметического для первого ряда:

$$\mu_{\bar{x}_1} = \frac{S_1}{\sqrt{N-1}} = \frac{4,6}{\sqrt{9}} = 1,5$$

6. Определяем стандартную ошибку среднего арифметического для второго ряда:

$$\mu_{\bar{x}_2} = \frac{S_2}{\sqrt{N-1}} = \frac{6,3}{\sqrt{9}} = 2,1$$

7. Находим достоверность различия между средними величинами:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\mu^2_{\bar{x}_1} + \mu^2_{\bar{x}_2}}} = \frac{148,5 - 121}{\sqrt{(2,1)^2 + (1,5)^2}} = \frac{27,5}{2,58} = 10,7$$

Поскольку при N = 10 (N – 1 = 9) наблюдениях показатель T превысил указанный в таблице значений коэффициента Стьюдента (таблица 3.1) уровень достоверности различия более 99,9 %, т.е. средние двух групп статистически значимо различаются с доверительной вероятностью 99,9 %.

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории

Выявить достоверность различий средних в двух выборках с нормальными частотными законами распределения позволяет функция **ТТЕСТ(М1;М2;Хв;Тип)**,

где М1 – границы массива первой выборки,

М2 – границы массива второй выборки,

Хв – хвосты, для двустороннего распределения значение 2,

Тип – тип теста. Может принимать следующие значения:

1 – парный тест, если в двух выборках одни и те же пациенты, например, их показатели до приема и после приема лекарства,

2 – тест в группах с равными дисперсиями,

3 – тест в группах с разными дисперсиями (пациенты в двух группах разные, мы ничего о них не знаем).

Пример использования функции **ТТЕСТ(B4:K4; B5:K5;2;1)**.

Функция возвращает вероятность справедливости нулевой гипотезы (H_0). H_0 принимается, если ее вероятность больше уровня значимости (α), в противном случае считается справедливой альтернативная гипотеза, утверждающая что две группы статистически значимо различаются.

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта 1 (рис. 3.6).

2. Введите исходные данные в блок ячеек с адресами **(A4:K5)**.

3. Введите функцию **ТТЕСТ(B4:K4; B5:K5;2;1)**.

4. Проанализируйте результаты (рис. 3.7).

5. На основании полученного значения вероятности нулевой гипотезы сделайте вывод о статистической значимости различия средних значений в двух выборках.

Примечание: *Следует помнить о том, что критерий Стьюдента можно применять только для выборок с нормальным частотным распределением.*

	A	B	C	D	E	F	G	H	I	J	K
1	**Макет для определения достоверности различия среднего в двух выборках**										
2		***Исходные данные***									
3		**Систолическое давление у студентов (мм рт ст)**									
4	До занятий	115	120	120	125	120	130	125	120	115	120
5	После занятий	150	145	155	140	140	145	150	160	150	150
6	**ТТЕСТ**	=ТТЕСТ(B4:K4;B5:K5;2;1)									

Рис.3.6 Макет для определения вероятности справедливости нулевой гипотезы (парный тест)

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Макет для определения достоверности различия среднего в двух выборках**											
2		***Исходные данные***										
3		**Систолическое давление у студентов (мм рт ст)**										
4	До занятий	115	120	120	125	120	130	125	120	115	120	
5	После занятий	150	145	155	140	140	145	150	160	150	150	
6	**ТТЕСТ**	0,000004										
7	**Вывод**											
8	Так как вероятность нулевой гипотезы меньше 0,001, то нулевая гипотеза отвергается, принимается											
9	альтернативная гипотеза. Две группы статистически значимо различаются с достоверностью более, чем 99,9%.											

Рис.3.7. Результаты вычисления справедливости нулевой гипотезы (парный тест)

Варианты заданий на определение достоверности различий средних величин выборок для самостоятельного выполнения

Вариант 2. По уровню лейкоцитов в общем анализе крови (ОАК) у 8 пациентов с острой вирусной пневмонией до и после приема противовирусного препарата определить, достоверно ли снижение количества лейкоцитов в общем анализе крови после лечения (группы состоят из одних и тех же пациентов):

Лейкоциты (10^9) в ОАК **до приема препарата**	10,2	15,1	15,4	19,5	19,7	9,5	15,3	21,4
Лейкоциты (10^9) в ОАК **после приема препарата**	8,2	10,1	11,8	10,3	17,4	9,0	10,7	11,4

Вариант 3. По величине пульса у 9 студентов до и на 5 минуте после занятий физической культурой определить, достоверно ли повышение частоты сердечных сокращений (ЧСС) после занятия (о формировании выборок ничего неизвестно):

ЧСС (уд. в минуту) **до занятий**	72	78	62	64	72	75	72	79	64
ЧСС (уд. в минуту) **после занятий**	85	82	72	70	82	100	94	88	69

Вариант 4. По величине содержания натрия (Na^+) в плазме крови (Мм/л) у 10 пациентов с артериальной гипертензией I-III стадий до и после применения лизиноприла по 20 мг в сутки определить, достоверно ли повышение содержания натрия в плазме крови при артериальной гипертензии (группы состоят из одних и тех же пациентов):

Na^+ в плазме крови (Мм/л) **до приема препарата**	146	150	155	150	160	158	161	142	143	156
Na^+ в плазме крови (Мм/л) **после приема препарата**	137	136	132	140	132	131	133	129	134	131

Вариант 5. По величине температуры у 5 пациентов с острой респираторной инфекцией до и после приема жаропонижающего определить, достоверно ли снижение температуры тела после приема лекарства (группы состоят из различных пациентов, но температура измерялась одним прибором):

Температура тела (°С) **до приема лекарства**	37,2	37,3	37,4	37,5	37,6
Температура тела (°С) **после приема лекарства**	37,1	37,2	37,3	37,4	37,5

Вариант 6. По величине содержания натрия (Na^+) в плазме крови (Мм/л) у 10 пациентов с артериальной гипертензией I-III стадий и у 10 человек контрольной группы определить, достоверно ли повышение содержания натрия в плазме крови при артериальной гипертензии:

Na^+ в плазме крови (Мм/л) **исследуемая группа**	146	150	155	150	160	158	161	142	143	156
Na^+ в плазме крови (Мм/л) **контрольная группа**	137	136	132	140	132	131	133	129	134	131

Вариант 7. По уровню гемоглобина капиллярной крови у 8 беременных женщин с железодефицитной анемией до и после приема железосодержащего препарата определить, достоверно ли повышение гемоглобина капиллярной крови после лечения (группы состоят из одних и тех же женщин):

Гемоглобин (г/л) в ОАК **до приема препарата**	102	101	98	95	88	89	95	98
Гемоглобин (г/л) в ОАК **после приема препарата**	112	105	118	103	93	90	105	114

Вариант 8. По величине экскреции альдостерона с мочой у 8 пациентов с артериальной гипертензией I-III стадий и у 8 человек контрольной группы определить, достоверно ли повышение экскреции альдостерона с мочой при артериальной гипертензии (измерения проводились одним и тем же прибором):

Экскреция альдостерона (мкг/сут.) **у пациентов с АГ**	27,2	25,1	24,4	29,3	28,9	28,9	31,3	32,0
Экскреция альдостерона (мкг/сут.) **контрольная группа**	11,2	12,1	11,8	10,3	13,4	9,0	10,7	11,2

Вариант 9. По величине систолического давления ($АД_{сист.}$) у 10 студентов после занятий физической культурой и 10 студентов контрольной группы, определить, достоверно ли повышение уровня систолического давления после занятия:

$АД_{сист.}$ (в мм рт. ст.) **после занятий**	160	155	135	140	125	140	150	120	150	140
$АД_{сист.}$ (в мм рт. ст.) **контрольной группы**	110	120	140	135	100	120	125	120	120	120

Вариант 10. По величине температуры у 5 пациентов с острой респираторной инфекцией до и после приема жаропонижающего определить, достоверно ли снижение температуры тела после приема лекарства (группы состоят из одних и тех же пациентов):

Температура тела (°С) **до приема лекарства**	37,9	37,4	37,7	37,7	38,0
Температура тела (°С) **после приема лекарства**	37,1	37,2	37,3	37,3	37,5

Вариант 11. По величине пульса (ЧСС) у 9 пациентов с артериальной гипертензией до и после приема в течение 2 недель бисопролола определить, достоверно ли ЧСС после приема лекарства (группы из различных пациентов):

ЧСС (уд. в минуту) **до лечения**	84	78	72	74	78	95	88	79	74
ЧСС (уд. в минуту) **после 2 недель лечения**	60	65	58	58	64	68	70	65	64

Вариант 12. По уровню лейкоцитов в общем анализе крови (ОАК) у 8 пациентов с острой вирусной пневмонией до и после приема противовирусного препарата определить, достоверно ли снижение количества лейкоцитов в ОАК после лечения (группы из разных пациентов):

Лейкоциты (10^9) в ОАК **до приема препарата**	9,2	13,1	14,2	19,5	18,9	8,9	11,3	12,0
Лейкоциты (10^9) в ОАК **после приема препарата**	8,3	10,0	11,8	10,3	17,4	9,0	10,7	11,4

Вариант 13. По величине пульса (ЧСС) у 10 пациентов с артериальной гипертензией до и после приема в течение 2 недель бисопролола определить, достоверно ли снижение ЧСС после лечения (пациенты одни и те же):

ЧСС (уд. в минуту) **до лечения**	84	78	72	74	78	95	88	79	74	82
ЧСС (уд. в минуту) **после 2 недель лечения**	60	65	58	58	64	68	70	65	64	62

Вариант 14. По величине пульса (ЧСС) у 9 студентов до и на 5 минуте после занятий физической культурой определить, достоверно ли повышение ЧСС после занятия (группы из разных студентов):

ЧСС (уд. в минуту) **до занятий**	72	78	62	64	72	75	72	79	64
ЧСС (уд. в минуту) **после занятий**	85	82	72	70	82	97	94	88	69

ЗАДАНИЕ №4. ОПРЕДЕЛЕНИЕ ДОСТОВЕРНОСТИ СООТВЕТСТВИЯ МЕЖДУ ВИДАМИ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ ДВУХ ВЫБОРОК ПО КРИТЕРИЮ ХИ-КВАДРАТ

Типовое задание (*Вариант 1*)

Определить имеются ли различия в характере эпидемических процессов при дизентерии и прочих острых кишечных инфекциях, если помесячное число пациентов было следующим:

Таблица 3.2 Заболеваемость дизентерией и прочими острыми кишечными инфекциями

Месяцы	Исходные данные		Расчетные данные	
	Число пациентов		Всего	Процент
	Дизентерия	Прочие ОКИ		
1	2	3	4	5
Январь	8	110	118	11,2
Февраль	10	30	40	3,8
Март	7	90	97	9,2
Апрель	10	40	50	4,8
Май	8	20	28	2,7
Июнь	8	20	28	2,7
Июль	12	120	132	12,7
Август	10	100	110	10,5
Сентябрь	8	120	128	12,2
Октябрь	15	100	115	10,9
Ноябрь	7	140	147	14,0
Декабрь	6	50	56	5,3
Итого:	109	940	1049	100,0

Образец выполнения типового задания

1. Суммируем помесячное число пациентов обеими нозологическими формами:

8+110=118 – за январь (таблица 3.2, столбец 4).

2. Определяем удельный вес помесячного числа пациентов :

118/1049=0,112 (11,2%) (таблица 3.2, столбец 5).

3. «Ожидаемое» число пациентов (теоретический динамический ряд) (у) рассчитываем умножением общего количества заболевших дизентерией за год (109 человек) на долю заболевших за каждый месяц (за январь – 11,2%):

y_1 = 109 * 0,112 = 12,2 (таблица 3.3, столбец 4)

4. Находим отклонение фактически полученных данных от «ожидаемых», вычисляя разницу между х и у:

х – у = 8 – 12,2 = - 4,2 (таблица 3.3, столбец 5)

5. Вычисляем значение хи-квадрата путем суммирования частных от деления квадратов отклонений фактически полученных данных от «ожидаемых» на число «ожидаемых» (таблица 3.3, столбец 6).

Значения хи-квадрат могут возрастать от 0 до бесконечности. Если х = у, т.е. «ожидаемые» числа соответствуют фактическим, $\chi^2 = 0$. Такое положение подтверждает правильность нулевой гипотезы и свидетельствует об отсутствии различий между сравниваемыми явлениями.

Таблица 3.3 Параметры расчета критерия хи-квадрат

Нозология	Месяцы	Зарегистрировано больных (реальный динамический ряд) (х)	«Ожидаемое» число больных (теоретический динамический ряд) (у)	(х-у)	$(х-у)^2$	$\frac{(х-у)^2}{у}$
1	2	3	4	5	6	7
Дизентерия	1	8	12,2	-4,2	17,6	1,4
	2	10	4,1	5,9	34,8	8,5
	3	7	10,0	-3,0	9,0	0,9
	4	10	5,2	4,8	23,0	4,4
	5	8	2,9	5,1	26,0	8,9
	6	8	2,9	5,1	26,0	8,9
	7	12	13,8	-1,8	3,2	0,2
	8	10	11,4	-1,4	1,9	0,2
	9	8	13,3	-5,3	28,1	2,1
	10	15	11,9	3,1	9.6	0,8

	11	7	15,3	-8,3	68,9	4,5
	12	6	5,8	0,2	0,04	0
Прочие острые кишечные инфекции (ОКИ)	1	110	105,3	-4,7	22,1	0,2
	2	30	35,7	-5,7	32,5	0,9
	3	90	86,4	3,6	12,9	0,1
	4	40	45,1	-5,1	26,0	0,6
	5	20	25,4	-5,4	29,2	1,1
	6	20	25,4	-5,4	29,2	1,1
	7	120	119,4	0,6	0,4	0
	8	100	98,7	1,3	1.7	0
	9	120	114,7	5,3	28,1	0,3
	10	100	102,5	-2,5	6,3	0
	11	140	131,6	8,4	70,6	0,6
	12	50	49,8	0,2	0,04	0
				$\sum_{i=1}^{N} \frac{(x-y)^2}{y} = \chi^2 = 44,9$ где N – количество пар		

6. Оценку достоверности полученной величины $\chi^2 = 44,9$ проводим по таблице 3.4 с учетом числа степеней свободы.

Таблица 3.4 Критические значения величины χ^2 с n′ степенями свободы

n′ = N - 1	Уровни значимости		n′ = N - 1	Уровни значимости	
	0,05	0,01		0,05	0,01
1	3,84	6,64	15	25,0	30,6
2	5,99	9,21	16	26,3	32,0
3	7,82	11,34	17	27,6	33,4
4	9,49	13,28	18	28,9	34,8
5	11,07	15,09	19	30,1	36,2
6	12,59	16,81	20	31,4	37,6
7	14,07	18,48	21	32,7	38,9
8	15,51	20,1	22	33,9	40,3
9	16,92	21,7	23	35,2	41,6
10	18,31	23,2	24	36,4	43,0
11	19,68	24,7	25	37,7	44,3
12	21,0	26,2	26	38,9	45,6
13	22,4	27,7	27	40,1	47,0
14	23,7	29,1	28	41,3	48,3

В примере полученная величина $\chi^2 = 44,9$ выше критического уровня 19,68 при $p < 0,05$; и выше 24,7 при $p < 0,01$ и n' = 11 согласно таблице 3.4. Следовательно, есть основания отвергнуть нулевую гипотезу и говорить о существенности различий, а стало быть, и независимости характера эпидемического процесса при дизентерии и прочих кишечных инфекциях.

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории

Выявить соответствие между видами частотных распределений выборок позволяет ***критерий хи-квадрат*** К. Пирсона, который реализуется в среде электронных таблиц Excel функцией **ХИ2ТЕСТ(М1;М2)**, где

М1– границы массива с фактическим распределением выборок,

М2 – границы массива, в котором размещены теоретические значения, предварительно рассчитанные пользователем.

При вычислении критерия используются только численные показатели, причем в одной ячейке должно быть значение не меньше пяти, в противном случае следует уменьшить число групп, объединив несколько значений.

Пример использования функции: **ХИ2ТЕСТ(В4:К4;В5:К5).**

Функция возвращает вероятность справедливости нулевой гипотезы (H_0) *об отсутствии различия теоретического и ожидаемого распределений*, а следовательно, распределений в двух выборках. H_0 принимается, если ее вероятность больше уровня значимости (α), в противном случае считается справедливой альтернативная гипотеза, утверждающая что виды частотных распределений в двух группах статистически значимо различаются.

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере варианта 1 (рис. 3.8).

	A	B	C	D	E	F	G
1	Макет для определения достоверности соответствия видов распределений в двух выборках						
2	*Исходные данные*			*Вспомогательные расчеты*			
3		Число больных				Ожидаемое значение	
4	Месяцы	Дизентерия	Прочие ОКИ	Всего	Доля	Дизентерия	Прочие ОКИ
5	Январь	5	110	=СУММ(B5:C5)	=D5/$D$17	=E5*$B$17	=E5*$C$17
6	Февраль	10	30	=СУММ(B6:C6)	=D6/$D$17	=E6*$B$17	=E6*$C$17
7	Март	7	90	=СУММ(B7:C7)	=D7/$D$17	=E7*$B$17	=E7*$C$17
8	Апрель	10	40	=СУММ(B8:C8)	=D8/$D$17	=E8*$B$17	=E8*$C$17
9	Май	8	20	=СУММ(B9:C9)	=D9/$D$17	=E9*$B$17	=E9*$C$17
10	Июнь	8	20	=СУММ(B10:C10)	=D10/$D$17	=E10*$B$17	=E10*$C$17
11	Июль	12	120	=СУММ(B11:C11)	=D11/$D$17	=E11*$B$17	=E11*$C$17
12	Август	10	100	=СУММ(B12:C12)	=D12/$D$17	=E12*$B$17	=E12*$C$17
13	Сентябрь	8	120	=СУММ(B13:C13)	=D13/$D$17	=E13*$B$17	=E13*$C$17
14	Октябрь	15	100	=СУММ(B14:C14)	=D14/$D$17	=E14*$B$17	=E14*$C$17
15	Ноябрь	7	140	=СУММ(B15:C15)	=D15/$D$17	=E15*$B$17	=E15*$C$17
16	Декабрь	6	50	=СУММ(B16:C16)	=D16/$D$17	=E16*$B$17	=E16*$C$17
17	*Итого*	=СУММ(B5:B16)	=СУММ(C5:C16)	=СУММ(D5:D16)	=СУММ(E5:E16)	=СУММ(F5:F16)	=СУММ(G5:G16)
18							
19	ХИ2ТЕСТ=	=ХИ2ТЕСТ(B5:C16;F5:G16)					

Рис. 3.8 Макет для вычисления вероятности H_0 по критерию хи-квадрат

2. Введите исходные данные в блок ячеек с адресами **(A5:C16)**.

3. Выполните промежуточные расчеты.

Введите формулы для вычисления:

- сумм по строкам (в столбец **D5:D16**);
- сумм по столбцам (в строку **B17:G17**);
- доли числа пациентов по каждому месяцу по отношению к общему числу больных (в столбце **D5:$D$17**);
- ожидаемых значений по столбцам «Дизентерия» («Прочие ОКИ»), как доли заболевания (из соответствующей ячейки столбца Е), умноженной на общее количество дизентерией **($B$17)** («Прочие ОКИ» - **$C$17**), заполнив соответственно столбцы **(F5:F16), (G5:G16)**.

4. Вычислите вероятность нулевой гипотезы по критерию хи-квадрат – введите формулу **=ХИ2ТЕСТ(B5:C16;F5:G16)** в ячейку **B19**.

Результаты расчетов, представлены на рис. 3.9.

5. На основании полученного значения вероятности нулевой гипотезы сделайте вывод о статистической значимости различий частотных распределений в двух выборках, а, следовательно, и статистической значимости различий этих выборок.

Рис. 3.9 Результаты вычисления вероятности нулевой гипотезы по критерию хи-квадрат

	A	B	C	D	E	F	G
1	Макет для определения достоверности соответствия видов распределений в двух выборках						
2	*Исходные данные*			*Вспомогательные расчеты*			
3		Число больных				Ожидаемое значение	
4	Месяцы	Дизентерия	Прочие ОКИ	Всего	Доля	Дизентерия	Прочие ОКИ
5	Январь	8	110	118	0,1125	12,26	105,74
6	Февраль	10	30	40	0,0381	4,16	35,84
7	Март	7	90	97	0,0925	10,08	86,92
8	Апрель	10	40	50	0,0477	5,20	44,80
9	Май	8	20	28	0,0267	2,91	25,09
10	Июнь	8	20	28	0,0267	2,91	25,09
11	Июль	12	120	132	0,1258	13,72	118,28
12	Август	10	100	110	0,1049	11,43	98,57
13	Сентябрь	8	120	128	0,1220	13,30	114,70
14	Октябрь	15	100	115	0,1096	11,95	103,05
15	Ноябрь	7	140	147	0,1401	15,27	131,73
16	Декабрь	6	50	56	0,0534	5,82	50,18
17	*Итого*	109	940	1049	1	109	940
18							
19	ХИ2ТЕСТ=	0,000004					

Примечание: *Следует помнить о том, что элементарное значение частоты при использовании критерия хи-квадрат должно быть не меньше 5.*

Варианты заданий №4 на определение достоверности соответствия между видами частотных распределений двух выборок по критерию хи-квадрат для самостоятельного выполнения возьмите из задания №3.

Сравните результаты, полученные при использовании параметрического и непараметрического методов.

Внимание! *Если параметрический и непараметрический методы дают разные результаты, следует принимать те значения, которые соответствуют области действия примененного критерия. Для нормальных частотных распределений правомерно принятие результатов, полученных по критерию Стьюдента. Для частотных распределений, не соответствующих нормальному закону частотного распределения, принимают результаты, полученные по непараметрическому критерию хи-квадрат.*

ТЕМА 4

ИЗУЧЕНИЕ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

I. ЦЕЛЬ И ЗАДАЧИ ЗАНЯТИЯ

Цель: Изучить статистические методы оценки связи между исследуемыми явлениями с помощью корреляционного анализа

Задачи

1. Усвоить сущность корреляционной связи между признаками.
2. Овладеть методикой вычисления параметрического коэффициента корреляции Пирсона и непараметрического коэффициента ранговой корреляции Спирмена с оценкой характера, силы и достоверности связи.

II. СТУДЕНТ ДОЛЖЕН ЗНАТЬ

1. Теоретические основы статистических методов выявления и оценки связи между явлениями.
2. Понятия корреляционного и регрессионного анализа.
3. Практическое значение корреляционного и регрессионного анализа в медицине и здравоохранении.

III. СТУДЕНТ ДОЛЖЕН УМЕТЬ

1. Провести корреляционный анализ статистических совокупностей.

2. Правильно интерпретировать результаты корреляционного анализа с учетом его достоверности.

3. Определять область применения корреляционного анализа в медицине и здравоохранении.

IV. МОТИВАЦИЯ ТЕМЫ ЗАНЯТИЯ

Согласно законам диалектики все явления и процессы, совершающиеся в природе и обществе, взаимосвязаны и взаимообусловлены. Проведение статистического анализа связей, изучение закономерностей и определение влияющих факторов чрезвычайно актуальны для различных областей медицины, биологии, организации здравоохранения, социально-гигиенических и клинических исследований. Корреляционный анализ является методом статистического исследования, который позволяет оценить силу связи между явлениями, а также ее вид и направление.

V. ОСНОВНЫЕ ВОПРОСЫ ТЕМЫ

1. Понятие функциональной зависимости, аргумента, функции. Область применения. Понятие корреляционной связи. Виды корреляционной связи. Задачи корреляционного анализа.

2. Расчеты параметрического коэффициента корреляции методом квадратов (Пирсона), непараметрического коэффициента корреляции методом рангов (Спирмена). Определение достоверности коэффициента корреляции.

3. Понятие регрессионного анализа. Область применения. Расчет коэффициента регрессии, шкалы регрессии, сигма регрессии. Виды причинно - следственных связей и их характеристика. Алгоритм расчета коэффициента регрессии.

VI. МАТЕРИАЛЬНОЕ ОСНАЩЕНИЕ

1. Презентация с иллюстрацией методики проведения корреляционного и регрессионного анализов.

2. Плакат с логической структурой темы: «Изучение связи между явлениями. Корреляционный и регрессионный анализ».

3. Технические средства обучения: видео двойка MPEG 4 – DVD, персональный компьютер.

VII. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ

Тема изучается 3 академическихе часа и состоит из трех частей:

1. Вступительное слово преподавателя.

Во вступительном слове преподаватель обращает внимание студентов на значение изучения связи между явлениями в статистике, роль корреляционного и регрессионного анализа между явлениями и признаками.

2. Диагностика знаний студентов по теме: «Изучение связи между явлениями. Корреляционный и регрессионный анализ»

Преподаватель выясняет исходный теоретический уровень знаний в изучении корреляционного и регрессионного анализа в статистических исследованиях.

3. Инструктаж по выполнению работы и контроль над ходом ее выполнения.

На примере типового задания преподаватель рассматривает методику вычисления коэффициента корреляции и расчета качественной и количественной взаимосвязи статистических величин.

VIII. ИНФОРМАЦИОННЫЙ БЛОК ТЕМЫ

1. Понятие функциональной зависимости, аргумента, функции. Область применения. Понятие корреляционной связи. Виды корреляционной связи. Задачи корреляционного анализа

Еще Гиппократ в VI в. до н.э. обратил внимание на наличие связи между телосложением и темпераментом людей, между строением тела человека и предрасположенностью к тем или иным заболеваниям. Определенные виды подобной связи выявлены также в животном и растительном мире.

Для описания связей между переменными величинами применяют ***математическое понятие функции f***, которая ставит в соответствие каждому определенному значению независимой переменной x, определенное значение зависимой переменной y:

$$y = f(x),$$

где x – аргумент, y – соответствующее ему значение функции f(x).

Однозначную зависимость между переменными величинами x и y называют ***функциональной***, при которой любому значению одного из признаков соответствует строго определенное значение другого. Например, радиусу круга соответствует определенная площадь круга.

Функциональная связь характерна для физико-химических процессов.

Такого рода *однозначные или функциональные связи* между переменными величинами встречаются далеко не всегда. Известно, например, что между ростом и массой тела у человека существует положи-

тельная связь: более высокие индивиды имеют обычно и большую массу тела, чем низкорослые. Однако, из этого правила существуют исключения, когда люди невысокого роста оказываются тяжелее высокорослых. Причиной таких «исключений» является тот факт, что *каждый биологический признак представляет собой функцию многих переменных*: на него влияют и генетические, и средовые факторы, что и обусловливает варьирование признаков. Поэтому зависимость между биологическими признаками имеет не функциональный, а ***статистический характер***. При этом в статистической совокупности определенному значению одного признака, рассматриваемого в качестве аргумента, соответствует не одно и то же числовое значение, а целая гамма распределяющихся в вариационный ряд числовых значений другого признака, рассматриваемого в качестве зависимой переменной или функции. Эта зависимость между переменными величинами называется ***корреляционной*** или ***корреляцией***, при которой *значению каждой средней величины одного признака соответствует несколько значений другого взаимосвязанного с ним признака.*

В социально-гигиенических исследованиях, клинической медицине и биологии зависимость между явлениями имеет характер *корреляционной связи.*

Если функциональную связь легко обнаружить и измерить на единичных и групповых объектах, то этого нельзя выявить с корреляционными связями, которые можно изучать только на *групповых объектах методами математической статистики.*

Задачи корреляционного анализа сводятся к:

1) установлению направления и вида связи между варьирующими признаками;

2) измерению силы корреляционной связи;

3) проверке достоверности выборочных показателей корреляции.

Корреляционный анализ находит широкое применение в работе врача любой специальности: при оценке физического развития детей и подростков, для определения зависимости между условиями труда, быта и состояния здоровья; при определении зависимости показателей заболеваемости от пола, возраста, стажа работы и т.д.

Измерить и оценить статистическую достоверность корреляционной связи можно при помощи ***коэффициента корреляции*** (r_{xy}) и ***его средней ошибки*** (μ_r).

Коэффициент корреляции (r_{xy}) позволяет одним числом охарактеризовать силу связи между изучаемыми явлениями и дает представление о ее направлении.

Корреляционная связь между признаками бывает ***прямой (положительной)*** и ***обратной (отрицательной)***, ***прямолинейной*** и ***криволинейной***.

При ***прямой связи*** с увеличением значений одного признака возрастает среднее значение другого признака. Коэффициент корреляции, определяющий прямую связь, обозначается знаком плюс (+).

При ***обратной связи*** с увеличением одного признака убывает среднее значение другого. Коэффициент корреляции, характеризующий обратную связь, обозначается знаком (–).

По силе связи значение коэффициента корреляции колеблется *от единицы (полная связь) до нуля (отсутствие связи)*. Чем больше среднему значению одного признака соответствует значений другого признака, тем выше сила связи между ними.

Прямолинейная связь отличается относительно равномерным изменением средних значений одного признака при равных изменениях другого.

В случае ***криволинейной зависимости*** - при равномерном изменении одного признака могут наблюдаться возрастающие и убывающие средние значения другого признака.

Измерение и оценка связи между явлениями при прямолинейной корреляции осуществляется с помощью *коэффициента корреляции* (r_{xy}), а при криволинейной – *корреляционным отношением* η.

Коэффициент корреляции указывает лишь на направление и силу связи между двумя переменными величинами, но не дает возможности судить о том, как количественно меняются величины одного по мере изменения величин другого признака. Ответ на этот вопрос дает применение ***метода регрессии***.

2. Расчеты параметрического коэффициента корреляции методом квадратов (Пирсона), непараметрического коэффициента корреляции методом рангов (Спирмена). Определение достоверности коэффициента корреляции

Степень зависимости случайных величин x_i и y_i характеризуется значением коэффициента корреляции.

Параметрический коэффициент корреляции Пирсона применяется для оценки связи между количественными признаками, подчиняющимися нормальному закону частотного распределения. Принимается *нулевая гипотеза* (H_0), утверждающая, что признаки, представленные в разных выборках *независимы друг от друга*, а коэффициент корреляции близок к нулю. *Альтернативная гипотеза* (H_1) предполагает *наличие зависимости* между рассматриваемыми признаками, причем величина коэффициента корреляции указывает *на силу имеющейся зависимости*.

Для проверки достоверности гипотезы H_0 следует после нахождения значения коэффициента корреляции (r_{xy}) воспользоваться специальной *таблицей наибольших (критических) значений коэффициента корреляции* ($r_{кр}$), в которой при заданном уровне значимости α (0,01; 0,05) и

известном числе степеней свободы $n' = n - 2$ представлены *максимальные значения*, которые может иметь величина коэффициента корреляции при справедливости нулевой гипотезы. Если найденный $r_{xy} < r_{кр}$, то наблюдаемое отклонение от нуля незначимо, и нулевая гипотеза принимается: величины независимы, если $r_{xy} > r_{кр}$, то принимается гипотеза H_1: рассматриваемые величины зависимы.

Алгоритм расчета параметрического коэффициента корреляции *методом линейной корреляции Пирсона*:

1. Определить суммы рядов x и y ($\sum_{i=1}^{N} x_i$ и $\sum_{i=1}^{N} y_i$).
2. Вычислить средние арифметические рядов x и y ($\bar{x}$ и $\bar{y}$).
3. Рассчитать разность (d_x) каждого показателя ряда x и среднего арифметического этого ряда ($d_x = x_i - \bar{x}$), а также разность (d_y) каждого показателя ряда y и среднего арифметического $\bar{x}$ ($d_y = y_i - \bar{y}$).
4. Величины d_x и d_y возвести в квадрат (d_x^2 и d_y^2).
5. Показатели d_x и d_y перемножить между собой.
6. Рассчитать коэффициент линейной корреляции (r_{xy}):

$$r_{xy} = \pm \frac{\sum_{i=1}^{N}(d_x * d_y)}{\sqrt{\sum_{i=1}^{N} d_x^2 * \sum_{i=1}^{N} d_y^2}}$$

7. Оценить представительность (репрезентативность) коэффициента корреляции по таблице критических значений (таблица 4.1).

Таблица 4.1 Критические значения коэффициента корреляции (r_{xy})

$n' = n - 2$	Уровни значимости		$n' = n - 2$	Уровни значимости	
	0,05	0,01		0,05	0,01
5	0,75	0,87	27	0,37	0,47
6	0,71	0,83	28	0,36	0,46
7	0,67	0,80	29	0,36	0,46
8	0,63	0,77	30	0,35	0,45
9	0,60	0,74	45	0,33	0,42
10	0,58	0,71	40	0,30	0,39
11	0,55	0,68	45	0,29	0,37
12	0,53	0,66	50	0,27	0,35
13	0,51	0,64	60	0,25	0,33
14	0,50	0,62	70	0,23	0,30
15	0,48	0,61	80	0,22	0,28
16	0,47	0,59	90	0,21	0,27
17	0,46	0,58	100	0,20	0,25
18	0,44	0,56	125	0,17	0,23
19	0,43	0,55	150	0,16	0,21
20	0,42	0,54	200	0,14	0,18
21	0,41	0,53	300	0,11	0,15

22	0,40	0,52	400	0,10	0,13
23	0,40	0,51	500	0,09	0,12
24	0,39	0,50	700	0,07	0,10
25	0,38	0,49	900	0,06	0,09
26	0,37	0,48	1000	0,06	0,09

8. Оценить тесноту корреляционной связи по таблице 4.2.

Таблица 4.2 Степень тесноты связи коэффициента корреляции

Степень связи	Величина коэффициента
Малая (слабая)	0,01 – 0,30
Средняя (умеренная)	0,31 – 0,70
Большая (сильная)	0,71 – 1,00

Примечание: *О наличии тесной корреляции можно говорить при* r_{xy} *не менее 0,7.*

Непараметрический коэффициент ранговой корреляции Спирмена применяется для оценки связи между *количественными признаками, не подчиняющимися закону нормального частотного распределения* или *качественными (порядковыми)* характеристиками. В последнем случае, значение признака невозможно измерить, но разные значения можно сравнивать между собой и располагать их в порядке возрастания или убывания качества.

Например, 1 сорт, 2 сорт и т.д., 1-е место, 2-е место, 3-е место и т.д., т.е. любой последующий объект по качеству уступает предыдущему. Порядковый номер каждого объекта называется ***рангом***, упорядочивание объектов – ***ранжирование***.

Для исследования качественных показателей выдвигаются *непараметрические* гипотезы, которые проверяются с помощью соответствующих критериев, оперирующих рангами, не использующих какие-либо данные о законе распределения. Если применить ранговый критерий к количественным признакам, то он окажется менее эффективным, чем соответствующий параметрический критерий, так как *выборка, состоящая из количественных значений более содержательна* в информационном плане, по сравнению с порядковыми значениями. При работе с количественными признаками ранговые критерии реже, чем параметрические, отвергают нулевую гипотезу. И если нулевая гипотеза отвергнута ранговым критерием, то и параметрические критерии не приведут к иному результату.

Алгоритм расчета непараметрического коэффициента ранговой корреляции *по методу Спирмена:*

1. Определить ранги (x и y).
2. Рассчитать разность рангов (d).
3. Вычислить квадрат ранга (d^2).
4. Рассчитать коэффициент ранговой корреляции:

N

$$\rho_{xy} = 1 - 6 * \sum_{i=1} d^2/n\,(n^2 - 1),$$

где n – число пар коррелируемых рядов;

6 – постоянный коэффициент.

5. Установить репрезентативность коэффициента по таблице 4.3.

Если $\rho_{xy} < \rho_{кр}$ при данном уровне значимости, то принять нулевую гипотезу о том, что ряды признаков не связаны между собой. Если $\rho_{xy} > \rho_{кр}$ при данном уровне значимости - принять альтернативную гипотезу о существовании корреляционной связи рассматриваемых признаков.

Таблица 4.3 Критические значения коэффициента корреляции рангов ($\rho_{кр}$)

n	Уровни значимости		n	Уровни значимости	
	0,05	0,01		0,05	0,01
5	1,00		23	0,42	0,53
6	0,89	1,00	24	0,41	0,52
7	0,78	0,94	25	0,40	0,51
8	0,72	0,88	26	0,39	0,50
9	0,68	0,83	27	0,38	0,49
10	0,64	0,79	28	0,38	0,48
11	0,61	0,76	29	0,37	0,48
12	0,58	0,73	30	0,36	0,47
13	0,56	0,70	31	0,36	0,46
14	0,54	0,68	32	0,36	0,45
15	0,52	0,66	33	0,34	0,45
16	0,50	0,64	34	0,34	0,44
17	0,48	0,62	35	0,33	0,43
18	0,47	0,60	36	0,33	0,43
19	0,46	0,58	37	0,33	0,42
20	0,45	0,57	38	0,32	0,41
21	0,44	0,56	39	0,31	0,41
22	0,43	0,54	40	0,31	0,40

3. Понятие регрессионного анализа. Область применения. Расчет коэффициента регрессии, шкалы регрессии, сигма регрессии. Виды причинно - следственных связей и их характеристика

Регрессия – функция, позволяющая по величине одного корреляционного (связанного) признака определить средние величины другого признака. С помощью *регрессии* можно выяснить, как количественно меняется одна величина при изменении другой величины. Для определения размера этого изменения применяется ***коэффициент регрессии***.

Коэффициент регрессии $R_{x/y}$ – абсолютная величина, на которую в среднем изменяется признак при изменении другого признака на единицу.

Формула для расчета коэффициента регрессии:

$$R_{x/y} = r_{xy} \cdot \frac{\sigma_y}{\sigma_x},$$

где $R_{x/y}$ – коэффициент регрессии,

r_{xy} – коэффициент корреляции,

σ_x и σ_y – средние квадратические отклонения для ряда x и ряда y.

С помощью коэффициента регрессии без специальных измерений можно определить величину одного из признаков (например, массы тела), зная значение другого (роста).

Для этих целей служит ***уравнение линейной регрессии***. Рассмотрим его вид при определении зависимости массы тела человека от его роста:

$$y = M_y + R_{x/y} (x - M_x),$$

где y – искомая величина массы тела,

x – известная величина роста,

$R_{x/y}$ – коэффициент регрессии при определении зависимости массы тела человека от его роста,

M_y – среднее значение массы тела, характерное для данного возраста,

M_x – среднее значение роста, характерное для данного возраста.

В практике исследования физического развития детей и подростков существует *метод оценки физического развития по шкалам регрессии.*

Индивидуальные значения отдельных признаков очень разнообразны: так, у людей с одинаковым ростом показатели массы тела и окружности грудной клетки могут колебаться в самых широких пределах. *Меру разнообразия индивидуальных размеров признака* характеризует ***сигма регрессии*** ($\sigma_{Rx/y}$), которая рассчитывается по формуле:

$$\sigma_{Rx/y} = \sigma_y \sqrt{1 - r_{xy}^2}$$

где σ_y – среднее квадратическое отклонение измеряемого признака (например, массы тела),

r_{xy} – коэффициент корреляции.

Метод регрессионного анализа широко используется в научной деятельности и практической медицине.

Алгоритм расчета коэффициента регрессии:

1. Определить суммы рядов x и y ($\sum_{i=1}^{N} x_i$ и $\sum_{i=1}^{N} y_i$).
2. Вычислить средние арифметические рядов x и y ($\bar{x}$ и $\bar{y}$).
3. Рассчитать разность (d_x) каждого показателя ряда x и среднего арифметического этого ряда ($d_x = x_i - \bar{x}$), а также разность (d_y) каждого показателя ряда y и среднего арифметического $\bar{y}$ ($d_y = y_i - y$).
4. Величины d_x и d_y возвести в квадрат (d_x^2 и d_y^2).
5. Показатели d_x и d_y перемножить между собой.

6. Рассчитать коэффициенты регрессии (R_{xy} и R_{yx}):

$$R_{xy} = \frac{\sum_{i=1}^{N}(dx \cdot dy)}{\sum_{i=1}^{N} dy^2}; \quad R_{yx} = \frac{\sum_{i=1}^{N}(dx \cdot dy)}{\sum_{i=1}^{N} dx^2}$$

7. Оценить репрезентативность с вычислением ошибки выборки (μ_{Rxy} и μ_{Ryx}).

$$\mu_{Rxy} = \pm \sqrt{\frac{\sum \bar{d}_y^2 - \frac{\sum(\bar{d}_x \cdot \bar{d}_y)}{\sum \bar{d}_y^2}}{\sum \bar{d}_y^2 \cdot (n-2)}}; \quad \mu_{Ryx} = \pm \sqrt{\frac{\sum \bar{d}_y^2 - \frac{\sum(\bar{d}_x \cdot \bar{d}_y)}{\sum \bar{d}_x^2}}{\sum \bar{d}_x^2 \cdot (n-2)}}$$

Степень репрезентативности определяется по Т – критерию при n′= n – 2 и уровне значимости 0,05.

$$t_{xy} = \frac{R_{xy}}{\mu_{Rxy}}; \qquad t_{yx} = \frac{R_{yx}}{\mu_{Ryx}}$$

Между коэффициентами корреляции и регрессии имеется определенная *связь*, позволяющая вычислить один из них через другой. Ее *можно выразить формулой:*

$$r_{xy} = \pm\sqrt{R_{xy} \cdot R_{yx}}$$

XI. УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА СТУДЕНТОВ

ЗАДАНИЕ №1. ОПРЕДЕЛЕНИЕ ПАРАМЕТРИЧЕСКОГО КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ МЕТОДОМ ЛИНЕЙНОЙ КОРРЕЛЯЦИИ ПИРСОНА

Типовое задание (*Вариант 1*)

Вычислить коэффициент корреляции методом линейной корреляции и оценить его репрезентативность, используя исходные данные, приведенные в таблице 4.4.

Таблица 4.4 Рост и масса тела десяти школьников 2 класса:

Рост тела (см)	110	112	120	127	130	135	135	140	145	145
Масса тела (кг)	20	22	24	25	25	27	25	30	35	37

Образец выполнения типового задания

1. Составляем сводную таблицу исходных и рассчитанных данных:

Рост тела (x_i)	Масса тела (y_i)	$dx_i = x_i - \bar{x}$	$dy_i = y_i - \bar{y}$	$dx_i * dy_i$	dx_i^2	dy_i^2
110	20	-19,9	-7	139,3	396,01	49
112	22	-17,9	-5	89,5	320,41	25
120	24	-9,9	-3	29,7	98,01	9
127	25	-2,9	-2	5,8	8,41	4
130	25	0,1	-2	-0,2	0,01	4
135	27	5,1	0	0	26,01	0
135	25	5,1	-2	-10,2	26,01	4
140	30	10,1	3	30,3	102,01	9
145	35	15,1	8	120,8	228,01	64
145	37	15,1	10	151	228,01	100
$\sum_{i=1}^{N} x_i = 1299$	$\sum_{i=1}^{N} y_i = 270$			$\sum_{i=1}^{N} dx_i * dy_i = 556$	$\sum_{i=1}^{N} dx_i^2 = 1405,9$	$\sum_{i=1}^{N} dx_i^2 = 268$
$\bar{x} = 129,9$	$\bar{y} = 27$					

2. Рассчитываем коэффициент линейной корреляции:

$$r_{xy} = \frac{\sum_{i=1}^{N}(dx_i * dy_i)}{\sqrt{\sum_{i=1}^{N} dx_i^2 * \sum_{i=1}^{N} dy_i^2}} = \frac{556}{\sqrt{1405,9 * 268}} = 0,906$$

Данный результат говорит о наличии прямой сильной степени связи (таблица 4.2).

3. По таблице 4.1 критических значений коэффициента корреляции, учитывая, что $n' = 10 - 2$, определяем репрезентативность коэффициента корреляции:

$r_{xy} = 0,906 > r_{кр} = 0,77$, следовательно, отвергается нулевая гипотеза об отсутствии связи между ростом и массой тела и принимается альтернативная, о том что рост и масса тела коррелируют между собой с вероятностью 99%.

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории.

Коэффициент корреляции Пирсона применяется для выборок, содержащих количественный признак, с нормальным частотным распределением (в параметрической статистике). В электронных таблицах Excel его значение вычисляется с помощью встроенной статистической функции КОРРЕЛ().

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта 1 (рис. 4.1).

2. Введите исходные данные в блок ячеек с адресами **(A5:C14)**.

3. С помощью функции КОРРЕЛ() вычислите коэффициент линейной корреляции Пирсона.

	A	B	C
1	Макет для вычисления коэффициента линейной корреляции Пирсона		
2		Исходные данные	
3	Порядковый номер	Рост тела (см) (X)	Масса тела (кг) (Y)
4			
5	1	110	20
6	2	112	22
7	3	120	24
8	4	127	25
9	5	130	25
10	6	135	27
11	7	135	25
12	8	140	30
13	9	145	35
14	10	145	37
15		=КОРРЕЛ (B5:B14;C5:C14)	

Рис. 4.1 Макет для вычисления коэффициента линейной корреляции Пирсона

4. Проанализируйте результаты (рис. 4.2).

	A	B	C
1	Макет для вычисления коэффициента линейной корреляции Пирсона		
2		Исходные данные	
3	Порядковый номер	Рост тела (см) (X)	Масса тела (кг) (Y)
4			
5	1	110	20
6	2	112	22
7	3	120	24
8	4	127	25
9	5	130	25
10	6	135	27
11	7	135	25
12	8	140	30
13	9	145	35
14	10	145	37
15		КОРРЕЛ= 0,897	

Рис. 4.2 Результаты расчета коэффициента линейной корреляции Пирсона

ЗАДАНИЕ №2. ОПРЕДЕЛЕНИЕ НЕПАРАМЕТРИЧЕСКОГО КОЭФФИЦИЕНТА РАНГОВОЙ КОРРЕЛЯЦИИ СПИРМЕНА

Типовое задание (*Вариант 1*)

Вычислить коэффициент корреляции методом рангов и оценить его достоверность, используя исходные данные:

Таблица 4.4 Рост и масса тела десяти школьников 2 класса:

Рост тела (см)	110	112	120	127	130	135	135	140	145	145
Масса тела (кг)	20	22	24	25	25	27	25	30	35	37

Образец выполнения типового задания

1. Составляем сводную таблицу исходных и рассчитанных данных:

№	Рост тела (x)	Масса тела (y)	Номера рангов		Разность рангов (d)	Квадрат разности рангов (d^2)
			Роста тела(x)	Массы тела (y)		
1	110	20	1	1	0	0
2	112	22	2	2	0	0
3	120	24	3	3	0	0
4	127	25	4	5,3	–1,3	1,7
5	130	25	5	5,3	–0,3	0,09
6	135	27	6,5	6	0,5	0,25
7	135	25	6,5	5,3	1,2	1,44
8	140	30	8	7	1	1
9	145	35	9,5	8	1,5	2,25
10	145	37	9,5	9	0,5	0,25

2. Рассчитываем коэффициент ранговой корреляции, вычислив что:

3.

$$\sum_{i=1}^{N} d^2 = 6{,}98$$

$\rho_{xy} = 1 - 6 \cdot \sum d^2/n \cdot (n^2 - 1) = 1 - 6 \cdot 6{,}98/[10 \cdot (100 - 1)] = 1 - 41{,}88/990 = +0{,}96$

Учитывая, что n=10, сравниваем полученное значение ρ_{xy} с табличным значением $\rho_{кр} = 0{,}79$ (таблица 4.3). Даже при самом высоком уровне значимости $\alpha=0{,}01$ значение $\rho_{xy} > \rho_{кр}$. Следовательно, отвергается нулевая гипотеза об отсутствии связи и принимается альтернативная, утверждающая, что рост коррелирует с массой тела школьников с вероятностью 99%

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Сведения из теории.

Коэффициент ранговой корреляции Спирмена применяется в непараметрической статистике и в электронных таблицах Excel его вычисление в виде встроенной функции не реализовано.

Вычислить коэффициент ранговой корреляции Спирмена можно, используя приведенный ниже алгоритм.

Ранги элементов выборки позволяет вычислить процедура ***Ранг и персентиль*** из пакета **Анализ данных**, который является «надстройкой» электронных таблиц Excel.

Вызов процедуры ***Ранг и Персентиль*** осуществляется командой:

Сервис ⇒ Анализ данных ⇒ Ранг и персентиль

В появившемся диалоговом окне ***Ранг и персентиль***, которое представлено на рис. 4.3, следует указать:

- входной интервал данных (блок данных **(B3:C14)**, в котором размещены исходные данные вместе с соответствующими заголовками);
- флажок в поле ***Метки в первой строке***;
- переключатель ***Выходной интервал***;
- адрес первой ячейки выходного интервала ***H4***;
- нажать кнопку ОК.

Результаты работы процедуры представлены на рис. 4.4.

Примечание: *- процедура* ***Ранг и персентиль*** *элементам массивов с равными значениями присваивает одинаковые ранги, что не влияет на конечные результаты вычислений;- большим значениям присваиваются самые низкие ранги.*

Рис. 4.3 Фрагменг листа Excel с диалоговым окном ***Ранг и персентиль***

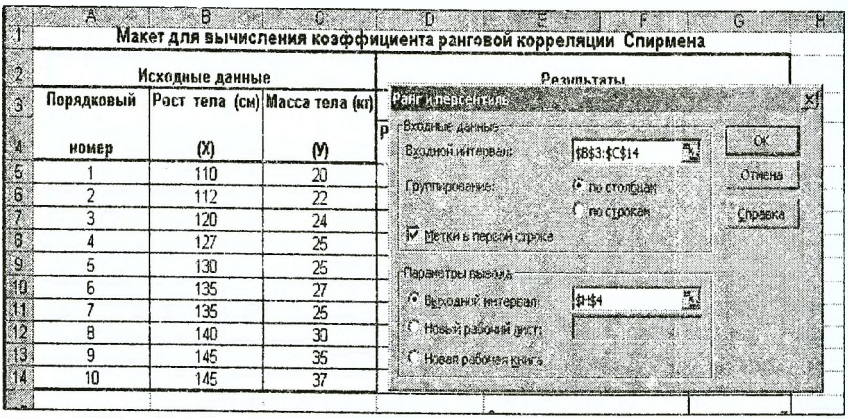
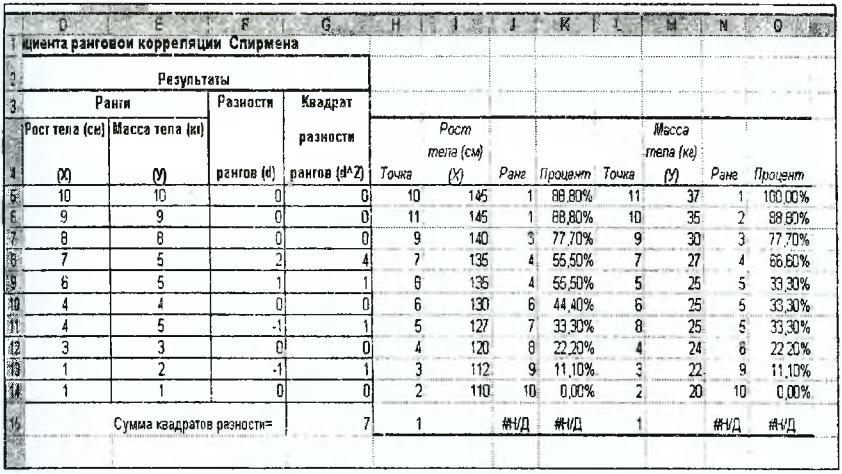


Рис. 4.4 Результаты работы функции ***Ранг и персентиль***



	D	E	F	G	H	I	J	K	L	M	N	O
1	циента ранговои корреляции Спирмена											
2		Результаты										
3	Ранги		Разности	Квадрат разности		Рост тела (см)				Масса тела (кг)		
4	Рост тела (см) (X)	Масса тела (кг) (Y)	рангов (d)	рангов (d^2)	Точка	(X)	Ранг	Процент	Точка	(Y)	Ранг	Процент
5	10	10	0	0	10	145	1	88,80%	11	37	1	100,00%
6	9	9	0	0	11	145	1	88,80%	10	35	2	88,80%
7	8	8	0	0	9	140	3	77,70%	9	30	3	77,70%
8	7	5	2	4	7	135	4	55,50%	7	27	4	66,60%
9	6	5	1	1	6	135	4	55,50%	5	25	5	33,30%
10	4	4	0	0	6	130	6	44,40%	6	25	5	33,30%
11	4	5	-1	1	5	127	7	33,30%	8	25	5	33,30%
12	3	3	0	0	4	120	8	22,20%	4	24	6	22,20%
13	1	2	-1	1	3	112	9	11,10%	3	22	9	11,10%
14	1	1	0	0	2	110	10	0,00%	2	20	10	0,00%
15		Сумма квадратов разности=		7	1		#Н/Д	#Н/Д	1		#Н/Д	#Н/Д

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта №1 (рис. 4.5).

2. Введите исходные данные в блок ячеек с адресами **(A5:C14)**.

3. Рассчитайте ранги элементов массива, применив процедуру ***Ранг и персентиль***.

4. Выполните сортировку данных по возрастанию поля ***Рост тела*** в блоке ячеек **(H5:K14)**.

5. Выполните сортировку данных по возрастанию поля ***Масса тела*** в блоке ячеек **(L5:O14)**.

6. Вручную скопируйте соответствующие ранги в блоки ячеек **(D5:D14)** и **(E5:E14)**.

7. В соответствии с макетом вычислений, приведенным на рис. 4.5 введите формулы для вычислений разности рангов, квадрата разности рангов, суммы квадратов разности рангов, коэффициента ранговой корреляции.

8. Проанализируйте результаты, сравните их со значениями, полученными по методу линейной корреляции Пирсона.

Результаты вычислений для варианта 1 приведены на рис. 4.6.

***Примечание**: При вводе формул следует учитывать, что диапазон размещения массивов исходных данных и количество объектов в выборках в различных вариантах отличается. В соответствии с этим, формулы для вычислений следует вводить исходя из их смысловой нагрузки.*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Макет для вычисления коэффициента ранговой корреляции Спирмена														
2		Исходные данные		Результаты											
3		Порядковый номер	Рост тела (см) (X)	Масса тела (кг) (Y)	Ранги		Разности рангов (d)	Квадрат разности рангов (d^2)			Рост		Масса		
4					Рост тела (см) (X)	Масса тела (кг)					Точка тела	Ранг	Процент	Точка тела	Ранг	Процент
5	1	110	20	10	10	=D5-E5	=F5^F5	2	110	10	0,888	2	20	10	1
6	2	112	22	9	9	=D6-E6	=F6^F6	3	112	9	0,888	3	22	9	0,888
7	3	120	24	8	8	=D7-E7	=F7^F7	4	120	6	0,777	4	24	8	0,777
8	4	127	25	7	7	=D8-E8	=F8^F8	5	127	7	0,555	5	25	5	0,666
9	5	130	25	6	5	=D9-E9	=F9^F9	6	130	6	0,555	6	25	5	0,333
10	6	135	27	4	4	=D10-E10	=F10^F10	7	135	4	0,444	8	25	5	0,333
11	7	135	25	4	5	=D11-E11	=F11^F11	8	135	4	0,333	7	27	4	0,333
12	8	140	30	3	3	=D12-E12	=F12^F12	9	140	3	0,222	9	30	3	0,222
13	9	145	35	1	2	=D13-E13	=F13^F13	10	145	1	0,111	10	35	2	0,111
14	10	145	37	1	1	=D14-E14	=F14^F14	11	145	1	0	11	37	1	0
15				Сумма квадратов разности=			=СУММ(G5:G14)	1		#Н/Д	#Н/Д	1		#Н/Д	#Н/Д
16	Коэффициент ранговой корреляции		Rxy =1-6*7//(10*(100-1))												

Рис. 4.5 Макет для вычисления коэффициента ранговой корреляции Спирмена

	A	B	C	D	E	F	G
1	Макет для вычисления коэффициента ранговой корреляции Спирмена						
2	Исходные данные			Результаты			
3	Порядковый	Рост тела (см)	Масса тела (кг)	Ранги		Разности	Квадрат разности
				Рост тела (см)	Масса тела (кг)		
4	номер	(X)	(Y)	(X)	(Y)	рангов (d)	рангов (d^2)
5	1	110	20	10	10	0	0
6	2	112	22	9	9	0	0
7	3	120	24	8	8	0	0
8	4	127	25	7	5	2	4
9	5	130	25	6	5	1	1
10	6	135	27	4	4	0	0
11	7	135	25	4	5	-1	1
12	8	140	30	3	3	0	0
13	9	145	35	1	2	-1	1
14	10	145	37	1	1	0	0
15					Сумма квадратов разности =		7
16	Коэффициент ранговой корреляции	Pxy=	0,96				

Рис. 4.6 Результаты расчета коэффициента ранговой корреляции Спирмена

Варианты заданий на вычисление коэффициентов линейной корреляции Пирсона и ранговой корреляции Спирмена для самостоятельного выполнения

Вариант 2. Уровни систолического и диастолического давления у 7 лиц призывного возраста проходящих медицинскую комиссию:

Систолическое давление (в мм рт. ст)	120	115	125	115	120	110	125
Диастолическое давление (в мм рт. ст)	80	70	75	70	65	65	80

Вариант 3. Частота сердечных сокращений у 7 школьников начальных классов в дневное и ночное время суток (ударов в мин):

День	80	82	85	82	75	80	85
Ночь	60	60	65	62	50	55	60

Вариант 4. Показатели роста и массы тела у 6 мужчин 50 лет:

Рост тела (в см)	164	170	185	180	182	195
Масса тела (в кг)	80	75	92	80	90	102

Вариант 5. Число послеоперационных осложнений при остром аппендиците в зависимости от сроков обращаемости:

Время (в часах)	1	2	3	4	5	6
Число осложнений	2	2	5	7	8	10

Вариант 6. Показатели систолического и диастолического давления у 7 студентов прошедших медицинский осмотр:

Систолическое давление (в мм рт. ст)	120	130	115	125	125	110	120
Диастолическое давление (в мм рт. ст)	75	80	70	65	75	65	85

Вариант 7. Уровень систолического артериального давления у 10 стационарных больных в утреннее и вечернее время суток (в мм рт. ст):

Утром	125	115	110	120	125	130	130	140	135	130
Вечером	120	100	95	110	100	110	100	120	115	100

Вариант 8. Показатели роста и массы тела 10 абитуриентов медицинского университета:

Рост тела (см)	190	175	185	180	195	198	176	184	182	187
Масса тела (кг)	90	80	80	85	83	90	70	72	75	75

Вариант 9. Уровень диастолического артериального давления у 6 студентов в дневное и ночное время суток (в мм рт. ст):

Днем	85	70	80	85	75	80
Ночью	65	60	65	70	55	65

Вариант 10. Частота сердечных сокращений у 6 больных в зависимости от температуры тела:

Температура тела (°С)	37,2	37,5	38,8	38,3	39,2	39,7
Частота сердечных сокращений (в мин)	80	85	95	95	100	105

Вариант 11. Длина и масса тела 6 новорожденных мальчиков:

Длина тела (в см)	50	53	54	50	55	56
Масса тела (в кг)	4,1	4,3	4,2	4,0	4,3	4,6

Вариант 12. Число дыхательных движений у 8 больных в зависимости от температуры тела:

Температура тела (°С)	36,7	37,2	37,9	38,3	38,4	38,7	38,4	39,1
Частота дыхания (в мин)	17	19	19	22	25	23	24	25

Вариант 13. Температура тела (в °С) у 7 больных в утреннее и вечернее время суток:

Утром	36,6	37,4	37,4	38,0	38,5	38,5	39,0
Вечером	36,8	38,2	37,9	38,5	38,7	39,2	39,5

Вариант 14. Число послеоперационные осложнений от начала приступа острого холецистита до начала операции (в часах):

Время до начала операции	3	7	11	16	20	24
Число осложнений	2	4	5	5	20	22

ТЕМА 5

СТАНДАРТИЗАЦИЯ ПОКАЗАТЕЛЕЙ

I. ЦЕЛЬ И ЗАДАЧИ ЗАНЯТИЯ

Цель: Изучить прямой метод стандартизации при статистическом анализе неоднородных совокупностей.

Задачи

1. Усвоить область применения метода стандартизации.
2. Овладеть методом прямой стандартизации.
3. Ознакомиться с методами косвенной и обратной стандартизации, а также условиями их применения.

II. СТУДЕНТ ДОЛЖЕН ЗНАТЬ

1. Сущность метода стандартизации и условия его применения.
2. Отличительные особенности прямого, косвенного и обратного методов стандартизации.
3. Условия выбора адекватного метода стандартизации.
4. Методику прямой стандартизации статистических совокупностей.

III. СТУДЕНТ ДОЛЖЕН УМЕТЬ

1. Определять случаи, при которых необходимо проведение прямого метода стандартизации.
2. Рассчитывать общие и специальные интенсивные показатели (или средние величины) для двух сравниваемых совокупностей.
3. Выбирать и рассчитывать стандарт.
4. Рассчитывать «ожидаемые величины» для каждой группы стандарта.
5. Определять стандартизованные показатели.
6. Сравнивать исследуемые группы по общим интенсивным (или средним) и стандартизованным показателям.
7. Анализировать полученные результаты и формулировать выводы.
8. Применять метод прямой стандартизации в практической деятельности врача.

IV. МОТИВАЦИЯ ТЕМЫ ЗАНЯТИЯ

Как известно, одним из основных требований статистики является качественная однородность сравниваемого материала, которая обеспечивает достоверность результатов и правильность выводов.

Неоднородный состав совокупностей, который иногда наблюдается при изучении заболеваемости, демографической ситуации и других аспектов общественного здоровья и здравоохранения ограничивает применение общепринятых методов статистического анализа без специальных приемов «элиминирования» этой неоднородности.

Статистический метод стандартизации позволяет исключить влияние различных факторов неоднородного состава (пол, возраст, стаж и т.д.) на общие показатели заболеваемости, смертности и др. Поэтому умение правильно выбирать и применять метод стандартизации является одним из требований к уровню профессиональной подготовки будущего специалиста.

V. ОСНОВНЫЕ ВОПРОСЫ ТЕМЫ

1. Сущность и методы стандартизации показателей, применяемые в статистическом анализе.
2. Прямой, косвенный и обратный методы стандартизации, этапы проведения прямого метода. Алгоритм вычисления стандартизованных показателей методом прямой стандартизации
3. Интерпретация результатов, полученных при проведении прямого метода стандартизации.

VI. МАТЕРИАЛЬНОЕ ОСНАЩЕНИЕ

1. Презентация с изображением логической структуры темы - «Стандартизация показателей».
2. Наглядные иллюстрации методики стандартизации показателей.
3. Технические средства обучения: видео двойка MPEG 4 – DVD, персональный компьютер, мультимедийный проектор.

VII. МЕТОДИКА ПРОВЕДЕНИЯ ЗАНЯТИЯ

Тема изучается в течение 3 академических часов для студентов лечебного факультета и состоит из трех частей:

3. Вступительное слово преподавателя.

Во вступительном слове преподаватель обращает внимание студентов на значение метода стандартизации.

2. Диагностика знаний студентов по теме: «Стандартизация показателей»

Преподаватель выясняет исходный теоретический уровень знаний в изучении методов стандартизации показателей.

3. Инструктаж по выполнению работы и контроль над ходом ее выполнения.

На примере типового задания разбирается методика стандартизации показателей. Далее студенты самостоятельно под контролем преподавателя выполняют варианты индивидуальных заданий по данной теме.

VIII. ИНФОРМАЦИОННЫЙ БЛОК ТЕМЫ

1. Сущность и методы стандартизации показателей, применяемые в статистическом анализе

Как практическому врачу, так и научному работнику нередко приходится сравнивать статистические показатели, вычисленные в *качественно неоднородных по составу группах*. Иногда при проведении социально-гигиенических, а также клинических исследований получить статистически однородные группы сравнения не представляется возможным. Прежде всего, это касается показателей заболеваемости, рождаемости, общей смертности по странам, городам, областям, районам, имеющим разный состав населения, как по возрасту, так и по полу. Аналогичная ситуация возникает и при анализе заболеваемости с временной утратой трудоспособности. Так, различия в показателях болевших лиц, случаях и днях нетрудоспособности, вычисленных для основной и контрольной групп, могут быть обусловлены не только условиями труда рабочих, но и неодинаковым составом исследуемых групп по полу, возрасту, уровню образования, семейному положению, среднему доходу и другим основным факторам, которые могут участвовать в формировании этих показателей.

Поэтому для сравнения общих показателей, вычисленных из неоднородных по своему составу совокупностей, применяются *методы стандартизации.*

Стандартизация – метод расчета условных (стандартизованных) показателей, заменяющих общие интенсивные (или средние) величины в тех случаях, когда их сравнение затруднено из-за несопоставимости состава групп.

Рассчитанные при помощи метода стандартизации показатели условны, т.к. они, устраняя влияния того или иного фактора на истинное значение показателей, указывают, какими были бы эти показатели, если бы влияние данного фактора отсутствовало. Поэтому стандартизованные показатели используются только для сравнения выборок.

Существует три метода стандартизации: *прямой, косвенный и обратный.*

2. Прямой, косвенный и обратный методы стандартизации, этапы проведения прямого метода. Алгоритм вычисления стандартизованных показателей методом прямой стандартизации

Прямой метод стандартизации включает 5 этапов.

I этап – расчет общих и специальных (по каждой группе - половой, возрастной и др.) интенсивных показателей (или средних величин) для двух сравниваемых совокупностей;

II этап – выбор и расчет стандарта;

III этап – расчет «ожидаемых величин» для каждой группы стандарта;

IV этап – определение стандартизованных показателей;

V этап – сравнение групп по общим интенсивным (или средним) и стандартизованным показателям. Выводы.

Сущность ***косвенного метода*** стандартизации заключается в том, что за стандарт принимают возрастные коэффициенты заболеваемости или смертности какой-либо определенной группы населения и высчитывают количество заболевших или умерших для сравниваемых групп населения.

Применение прямого и косвенного методов стандартизации требует знания возрастного состава сравниваемых групп населения.

В тех случаях, когда сведения о возрастном составе анализируемой группы отсутствуют, применяется ***«обратный» метод*** стандартизации, особенностью которого является то, что при исчислении стандартизованных коэффициентов используются сведения о возрастном составе больных или умерших.

При использовании обратного метода стандартизации необходимо помнить, что он дает *более точные результаты*, когда раздельно вычисляются стандартизованные коэффициенты для населения, находящегося в молодом возрасте (до 25-30 лет), и для населения более старшего возраста, а также при употреблении более дробных возрастных интервалов (пятилетних вместо десятилетних) для вычисления повозрастных коэффициентов стандарта.

Алгоритм вычисления стандартизованных показателей методом прямой стандартизации

1. Расчет стандартизованных показателей

I этап – расчет стандартизированных интенсивных показателей в двух сравниваемых совокупностях.

На первом этапе стандартизации требуется определить, как часто изучаемый признак встречается в разных группах и подгруппах наблюдения.

II этап – выбор стандарта.

На втором этапе за стандарт принимается тот состав статистической совокупности, для которой были бы характерны все особенности сравниваемых групп.

III этап – расчет ожидаемых чисел в каждой группе наблюдения.

На третьем этапе необходимо вычислить, какова могла бы быть величина сравниваемого явления, учитывая, что частные интенсивные показатели оставались прежними, а распределение среды соответствовало выбранному стандарту.

IV этап – расчет стандартизованных показателей и их сравнение.

На четвертом этапе стандартизованные показатели вычисляются как обычные интенсивные показатели с учетом стандарта.

2. Анализ полученных данных

3. Выводы и предложения

IX. УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА СТУДЕНТОВ

ЗАДАНИЕ № 1. ВЫЧИСЛЕНИЕ СТАНДАРТИЗОВАННЫХ ПОКАЗАТЕЛЕЙ МЕТОДОМ ПРЯМОЙ СТАНДАРТИЗАЦИИ

Типовое задание (*Вариант 1*)

Используем данные по числу случаев нетрудоспособности на двух автобазах.

Возрастная группа	Автобаза № 1		Автобаза № 2	
	Численность работающих	Число случаев нетрудоспособности	Численность работающих	Число случаев нетрудоспособности
до 20 лет	150	85	250	135
21 – 39	420	450	450	500
40 – 59	350	540	100	150
60 и старше	100	175	45	65
Всего:	**1020**	**1250**	**845**	**850**

Образец выполнения типового задания

1. Расчет стандартизованных показателей

I этап – рассчитываем число случаев нетрудоспособности по каждой возрастной группе в двух совокупностях на 100 работающих.

Автобаза № 1:

1. Рассчитываем общий интенсивный показатель на 100 работающих для автобазы № 1:

1020 работающих – 1250 случаев нетрудоспособности

100 работающих – x

$$x = \frac{1250 \cdot 100}{1020} = 122{,}6 \text{ (случая)}$$

2. Рассчитываем интенсивный показатель для группы работающих в возрасте до 20 лет:

150 работающих – 85 случаев нетрудоспособности
100 работающих – x

$$x = \frac{85 \cdot 100}{150} = 56{,}7 \text{ (случаев)}$$

3. Рассчитываем интенсивный показатель для группы работающих в возрасте 21-39 лет:

420 работающих – 450 случаев нетрудоспособности
100 работающих – x

$$x = \frac{450 \cdot 100}{420} = 107{,}1 \text{ (случай)}$$

4. Рассчитываем интенсивный показатель для группы работающих в возрасте 40-59 лет:

350 работающих – 540 случаев нетрудоспособности
100 работающих – x

$$x = \frac{540 \cdot 100}{350} = 154{,}3 \text{ (случая)}$$

5. Рассчитываем интенсивный показатель для группы работающих в возрасте 60 лет и старше:

100 работающих – 175 случаев нетрудоспособности
100 работающих – x

$$x = \frac{175 \cdot 100}{100} = 175 \text{ (случаев)}$$

Автобаза № 2:

1. Рассчитываем общий интенсивный показатель на 100 работающих для автобазы №2:

845 работающих – 850 случаев нетрудоспособности
100 работающих – x

$$x = \frac{850 \cdot 100}{845} = 100{,}6 \text{ (случаев)}$$

2. Рассчитываем интенсивный показатель для группы работающих в возрасте до 20 лет:

250 работающих – 135 случаев нетрудоспособности
100 работающих – x

$$x = \frac{135 \cdot 100}{250} = 54 \text{ (случая)}$$

3. Рассчитываем интенсивный показатель для группы работающих в возрасте 21-39 лет:

450 работающих – 500 случаев нетрудоспособности
100 работающих – x

$$x = \frac{500 \cdot 100}{450} = 111{,}1 \text{ (случая)}$$

4. Рассчитываем интенсивный показатель для группы работающих в возрасте 40-59 лет:

100 работающих -- 150 случаев нетрудоспособности
100 работающих – x

$$x = \frac{150 \cdot 100}{100} = 150 \text{ (случаев)}$$

5. Рассчитываем интенсивный показатель для группы работающих в возрасте 60 лет и старше:

$$x = \frac{65 \cdot 100}{45} = 144{,}4 \text{ (случая)}$$

45 работающих – 65 случаев нетрудоспособности
100 работающих -- x

6. Полученные результаты вносим в таблицу:

Возрастная группа	Частота случаев нетрудоспособности (на 100 работающих)	
	Автобаза № 1	Автобаза № 2
до 20 лет	56,7	54
21 – 39	107,1	111,1
40 – 59	154,3	150
60 и старше	175	144,4
Всего:	**122,6**	**100,6**

II этап – выбор стандарта.

За стандарт принимаем сумму работающих в двух автобазах по каждой возрастной группе.

Таким образом, в первой возрастной группе (до 20 лет) число работающих будет равно 150 + 250 = 400 человек,

во второй (21 - 39) лет - 420 + 450 = 870 человек,

в третьей (40 - 59 лет) - 350 + 100 = 450 человек,

в четвертой (60 лет и старше) - 100 + 45 = 145 человек.

Общее число работающих составляет 1865 человек.

III этап – расчет ожидаемых чисел в каждой группе наблюдения.

Определяем ожидаемое число случаев нетрудоспособности в соответствии со стандартом по возрастным группам.

Автобаза № 1:

1. Определяем ожидаемое число случаев нетрудоспособности в первой возрастной группе (до 20 лет):

100 работающих – 56,7 случаев нетрудоспособности
400 работающих – x

$$x = \frac{56,7400}{100} = 226,8 \text{ (случаев)}$$

2. Определяем ожидаемое число случаев нетрудоспособности во второй возрастной группе (21 - 39 лет):

100 работающих – 107,1 случая нетрудоспособности
870 работающих – x

$$x = \frac{107,1870}{100} = 931,7 \text{ (случая)}$$

3. Определяем ожидаемое число случаев нетрудоспособности в третей возрастной группе (40-59 лет):

100 работающих – 154,3 случая нетрудоспособности
450 работающих – x

$$x = \frac{154,3 \cdot 450}{100} = 694,4 \text{ (случая)}$$

4. Определяем ожидаемое число случаев нетрудоспособности в четвертой возрастной группе (60 лет и старше):

100 работающих – 175 случаев нетрудоспособности

$$x = \frac{175 \cdot 145}{100} = 253,8 \text{ (случая)}$$

145 работающих – x

Автобаза № 2:

1. Определяем ожидаемое число случаев нетрудоспособности в первой возрастной группе (до 20 лет):

$$x = \frac{54 \cdot 400}{100} = 216 \text{ (случаев)}$$

100 работающих – 54 случая нетрудоспособности
400 работающих – x

2. Определяем ожидаемое число случаев нетрудоспособности во второй возрастной группе (21 - 39 лет):

100 работающих – 111,1случая нетрудоспособности
870 работающих – x

$$x = \frac{111,1 \cdot 870}{100} = 966,5 \text{ (случая)}$$

3. Определяем ожидаемое число случаев нетрудоспособности в третей возрастной группе (40-59 лет):

$$x = \frac{150 \cdot 450}{100} = 675 \text{ (случаев)}$$

4. Определяем ожидаемое число случаев нетрудоспособности в четвертой возрастной группе (60 лет и старше):

100 работающих – 144,4случая нетрудоспособности
145 работающих – x

$$x = \frac{144,4 \cdot 145}{100} = 209,4 \text{ (случая)}$$

5. Полученные результаты вносим в таблицу:

Возрастная группа	Стандарт по числу работающих	Предполагаемое число случаев нетрудоспособности	
		Автобаза № 1	Автобаза № 2
до 20 лет	400	226,8	216
21 - 39	870	931,7	966,5
40 - 59	450	694,4	675
60 и старше	145	253,8	209,4
Всего:	**1865**	**2106,7**	**2066,9**

IV этап – расчет стандартизованных показателей и их сравнение.

100 работающих – 150 случаев нетрудоспособности
450 работающих – x

Рассчитываем стандартизованные показатели по двум автобазам на 100 работающих.

Автобаза № 1:

1865 работающих – 2106,7 случаев нетрудоспособности
100 работающих – x

$$x = \frac{2106,7 \cdot 100}{1865} = 112,9 \text{ (случаев)}$$

Автобаза № 2:

1865 работающих – 2066,9 случаев нетрудоспособности
100 работающих – x

$$x = \frac{2066,9 \cdot 100}{1865} = 110,8 \text{ (случаев)}$$

2. Анализ полученных данных

Анализ полученных данных показал, что в автобазе №1 предполагаемое число случаев нетрудоспособности составило 112,9, тогда как в автобазе №2 - 110,8 случаев.

3. Выводы

При условии одинакового возрастного состава коллективов двух автобаз частота случаев заболеваемости в автобазе №1 была бы выше, что обусловлено большим числом лиц в четвертой возрастной группе (60 и более лет) и меньшим в первой (до 20 лет).

Обработка данных на персональном компьютере в среде электронных таблиц Microsoft Excel

Алгоритм обработки данных

1. Введите текстовые значения на первом листе книги документа Excel, в соответствии с макетом для обработки данных на примере задания варианта 1 (рис. 5.1).

2. Введите исходные данные в блок ячеек с адресами **(B5:F8)**.

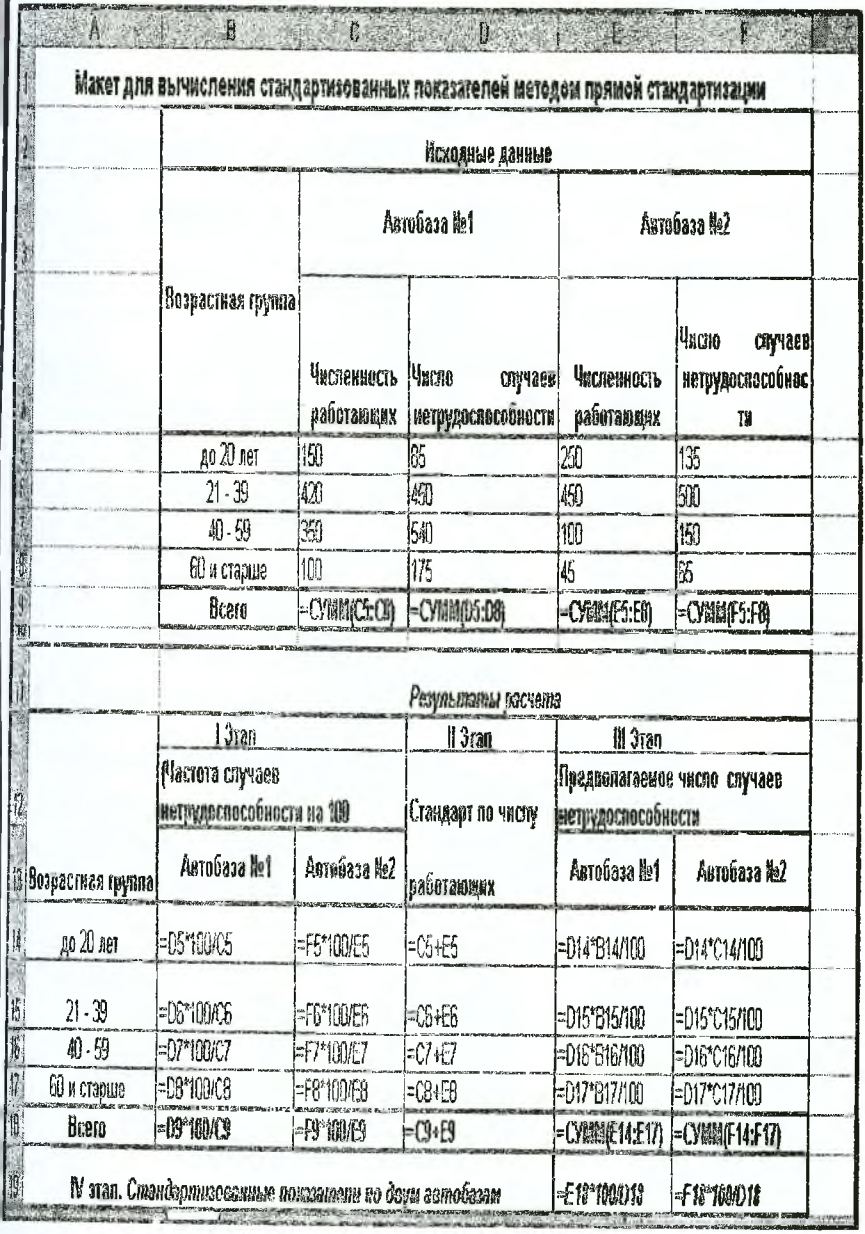
3. Вычислите в соответствующих ячейках суммы численности работающих и случаев нетрудоспособности по каждой автобазе (рис. 5.1).

4. Последовательно для каждого этапа введите формулы для расчета требуемых показателей в соответствии с представленным макетом автоматизации вычислений (рис. 5.1).

5. Проанализируйте результаты.

Результаты вычислений стандартизованных показателей методом прямой стандартизации для варианта 1 приведены на рис. 5.2.

Примечание: *При вводе формул следует учитывать, что диапазон размещения массивов исходных данных и количество объектов в выборках в различных вариантах отличается. В соответствии с этим, формулы для вычислений следует вводить исходя из их смысловой нагрузки.*



Макет для вычисления стандартизованных показателей методом прямой стандартизации

	Исходные данные				
	Возрастная группа	Автобаза №1		Автобаза №2	
		Численность работающих	Число случаев нетрудоспособности	Численность работающих	Число случаев нетрудоспособности
	до 20 лет	150	85	250	135
	21 - 39	420	450	450	500
	40 - 59	350	540	100	150
	60 и старше	100	175	45	65
	Всего	=СУММ(C5:C8)	=СУММ(D5:D8)	=СУММ(E5:E8)	=СУММ(F5:F8)

	Результаты расчета				
	I Этап		II Этап	III Этап	
	Частота случаев нетрудоспособности на 100		Стандарт по числу работающих	Предполагаемое число случаев нетрудоспособности	
Возрастная группа	Автобаза №1	Автобаза №2		Автобаза №1	Автобаза №2
до 20 лет	=D5*100/C5	=F5*100/E5	=C5+E5	=D14*B14/100	=D14*C14/100
21 - 39	=D6*100/C6	=F6*100/E6	=C6+E6	=D15*B15/100	=D15*C15/100
40 - 59	=D7*100/C7	=F7*100/E7	=C7+E7	=D16*B16/100	=D16*C16/100
60 и старше	=D8*100/C8	=F8*100/E8	=C8+E8	=D17*B17/100	=D17*C17/100
Всего	=D9*100/C9	=F9*100/E9	=C9+E9	=СУММ(E14:E17)	=СУММ(F14:F17)
IV этап. Стандартизованные показатели по двум автобазам				=E18*100/D18	=F18*100/D18

Рис. 5.1 Макет для автоматизации вычислений стандартизованных показателей методом прямой стандартизации

	A	B	C	D	E	F
1	Макет для вычисления стандартизованных показателей методом прямой стандартизации					
2		Исходные данные				
3		Возрастная	Автобаза №1		Автобаза №2	
4		группа	Численность работающих	Число случаев нетрудоспособности	Численность работающих	Число случаев нетрудоспособности
5		до 20 лет	150	85	250	135
6		21 - 39	420	450	450	500
7		40 - 59	350	540	100	150
8		60 и старше	100	175	45	65
9		Всего	1020	1250	845	850
10						
11		*Результаты расчета*				
12		Возрастная	I Этап (Частота случаев нетрудоспособности на 100		II Этап Стандарт по числу работающих	III Этап Предполагаемое число случаев нетрудоспособности
13		группа	Автобаза №1	Автобаза №2		Автобаза №1	Автобаза №2
14		до 20 лет	56,7	54,0	400	226,7	216,0
15		21 - 39	107,1	111,1	870	932,1	966,7
16		40 - 59	154,3	150,0	450	694,3	675,0
17		60 и старше	175,0	144,4	145	253,8	209,4
18		Всего	122,5	109,6	1865	2106,8	2067,1
19		IV этап. *Стандартизованные показатели по двум автобазам*				113,0	110,8

Рис. 5.2 Результаты расчета стандартизованных показателей методом прямой стандартизации

Варианты заданий на вычисление стандартизованных показателей методом прямой стандартизации для самостоятельного выполнения

Вариант 2. Рассчитать стандартизованные показатели смертности от злокачественных новообразований пищевода. За стандарт принять полусумму численности населения территориальных областей.

Распределение городского и сельского населения по областям и числу умерших от новообразований пищевода.

Область	Городские жители		Сельские жители	
	Численность населения (тыс. чел.)	Число умерших (абсолютные числа)	Численность населения (тыс. чел.)	Число умерших (абсолютные числа)
Минская	801,8	31	662,5	29
Брестская	924,2	16	513,1	22

Вариант 3. Рассчитать стандартизованные показатели травматизма на предприятии. За стандарт принять полусумму состава рабочих по полу.

Распределение рабочих слесарей и монтажников по полу и числу случаев травматизма.

Пол	Слесари		Монтажники	
	Число работающих (чел.)	Число случаев	Число работающих (чел.)	Число случаев
М	120	26	190	17
Ж	80	7	65	3

Вариант 4. Рассчитать стандартизованные показатели смертности от отравлений алкоголем. За стандарт принять полусумму численности населения анализируемых территориальных областей.

Распределение городского и сельского населения по областям и числу умерших от отравлений алкоголем.

Область	Городские жители		Сельские жители	
	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)
Минская	801,8	115	662,5	215
Витебская	914,4	263	363,9	215

Вариант 5. Рассчитать стандартизованные показатели летальности в 2-х больницах при срочных хирургических вмешательствах. За стандарт принять полусумму госпитализированных больных обеих больниц.

Распределение госпитализированных и умерших больных в 2-х больницах в зависимости от срока госпитализации.

Сроки госпитализации (сутки)	Больница №1		Больница №2	
	Число больных (чел.)	Число умерших (чел.)	Число больных (чел.)	Число умерших (чел.)
до 1	800	3	400	2
более 1	280	5	150	3

Вариант 6. Рассчитать стандартизованные показатели смертности от транспортных несчастных случаев. За стандарт принять полусумму численности населения анализируемых территориальных областей.

Распределение городского и сельского населения по областям и числу умерших от транспортных несчастных случаев.

Область	Городские жители		Сельские жители	
	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)
Минская	801,8	182	662,5	243
Гродненская	746,2	110	364,1	82

Вариант 7. Рассчитать стандартизованные показатели травматизма на промышленном предприятии. За стандарт принять полусумму состава рабочих по стажу.

Распределение рабочих шлифовщиков и фрезеровщиков по стажу и числу случаев травматизма.

Стаж	Шлифовщики		Фрезеровщики	
	Число рабочих (чел.)	Число случаев	Число рабочих (чел.)	Число случаев
5-9	120	12	210	16
10-15	200	9	300	13

Вариант 8. Рассчитать стандартизованные показатели смертности от всех причин. За стандарт принять полусумму численности населения анализируемых областей.

Распределение городского и сельского населения по областям и числу умерших от всех причин.

Область	Городские жители		Сельские жители	
	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)
Минская	801,8	8522	662,5	14508
Витебская	914,4	10975	363,9	8786
Гомельская	1472,2	11801	420,2	9340

Вариант 9. Рассчитать стандартизованный коэффициент рождаемости. За стандарт принять полусумму численности женщин каждого возраста.

Распределение коэффициента рождаемости по территориальным областям и возрастным группам.

Возраст (лет)	Минская		Брестская		Витебская	
	Численность женского населения (чел.)	Число родившихся на 1000 женщин	Численность женского населения (чел.)	Число родившихся на 1000 женщин	Численность женского населения (чел.)	Число родившихся на 1000 женщин
20-24	20905	58,23	18080	87,19	15180	55,66
25-29	30010	114,25	25900	100,74	20008	54,4
30-34	8005	186,11	7805	80,79	14005	46,32

Вариант 10. Рассчитать стандартизованные показатели смертности от менингококковой инфекции. За стандарт принять полусумму численности населения анализируемых территориальных областей.

Распределение городского и сельского населения по областям и числу умерших от менингококковой инфекции.

Область	Городские жители		Сельские жители	
	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)
Минская	801,8	1	662,5	2
Могилевская	846,6	2	286,9	1

Вариант 11. Рассчитать стандартизованные показатели смертности от всех форм туберкулеза. За стандарт принять полусумму численности населения анализируемых территориальных областей.

Распределение городского и сельского населения по областям и числу умерших от всех форм туберкулеза.

Область	Городские жители		Сельские жители	
	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)
Брестская	924,2	48	513,1	56
Гомельская	1052	136	420,2	94
Могилевская	846,6	81	286,9	41

Вариант 12. Рассчитать стандартизованные показатели смертности от острого инфаркта миокарда с гипертонической болезнью. За стандарт принять полусумму численности населения анализируемых территориальных областей.

Распределение городского и сельского населения по областям и числу умерших от острого инфаркта миокарда с гипертонической болезнью.

Область	Городские жители		Сельские жители	
	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)	Численность населения (тыс.чел.)	Число умерших (абсолютные числа)
Минская	801,8	42	662,5	36
Гродненская	746,2	53	364,1	31
Могилевская	846,6	47	286,9	18

ТЕМА 6

ОРГАНИЗАЦИЯ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

I. ЦЕЛЬ И ЗАДАЧИ ЗАНЯТИЯ

Цель: Изучить организацию и способы проведения статистического исследования.

Задачи

1. Научиться определять цель статистического исследования.
2. Научиться формулировать рабочую гипотезу статистического исследования.
3. Овладеть «стратегией» статистического исследования – методикой составления программы, включающей программу сбора, разработки и анализа статистического материала.
4. Освоить «тактику» статистического исследования – методику составления плана, включающего организационные моменты формирования объекта исследования с учетом охвата, способа отбора, времени, вида и необходимого объема наблюдений.
5. Научиться проводить обработку, сводку и анализ материала.

II. СТУДЕНТ ДОЛЖЕН ЗНАТЬ

1. Основные понятия, используемые при организации статистического исследования.
2. Этапы статистического исследования и их содержание.
3. Методику составления программы и плана исследования.
4. Методику оценки репрезентативности выборочной совокупности.
5. Правила отбора, регистрации, разработки и сводки статистического материала.
6. Методику статистического анализа.
7. Типичные ошибки при проведении статистического исследования.

III. СТУДЕНТ ДОЛЖЕН УМЕТЬ

1. Определять цель статистического исследования и формулировать рабочую гипотезу.
2. Составлять программу и план статистического исследования.
3. Выбирать единицу наблюдения и определять учетные признаки.
4. Обосновывать необходимое число наблюдений.

5. Производить регистрацию статистического материала путем заполнения учетных документов.

6. Группировать полученные данные.

7. Составлять макеты статистических таблиц.

8. Осуществлять статистическую сводку и рассчитывать статистические показатели.

9. Анализировать статистический материал и делать выводы.

IV. МОТИВАЦИЯ ТЕМЫ ЗАНЯТИЯ

Социально-гигиенические, клинико-статистические, физиологические и другие исследования с начального этапа планирования и до завершающего этапа внедрения результатов исследования в практику представляют собой единый взаимосвязанный научно-обоснованный процесс. В практической деятельности врача часто возникает необходимость проведения статистического исследования в целях получения данных, необходимых для разработки предложений по улучшению медицинской помощи. Решение многочисленных задач, поставленных перед практическим здравоохранением, требует от будущих специалистов знаний теоретических основ и навыков проведения статистического исследования, грамотного подхода к анализу результатов и внедрению их в практику.

V. ОСНОВНЫЕ ВОПРОСЫ ТЕМЫ

1. Цель и задачи статистического исследования, рабочая гипотеза. Этапы статистического исследования. Методика проведения статистического исследования

2. Составление программы и плана исследования. Регистрация и сбор материала. Формирование репрезентативной выборочной статистической совокупности

3. Обработка и сводка статистических данных. Анализ статистических данных

VI. МАТЕРИАЛЬНОЕ ОСНАЩЕНИЕ

1. Мультимедийная презентация по теме «Организация статистического исследования».

2. Наглядные иллюстрации этапов статистического исследования.

3. Технические средства обучения: видео двойка MPEG 4 – DVD, персональный компьютер, мультимедийный проектор.

VII. МЕТОДИКА ПРОВЕДЕНИЯ ЗАНЯТИЯ

Тема изучается в течение 6 академических часов для студентов лечебного факультета и состоит из трех частей:

4. Вступительное слово преподавателя.

Во вступительном слове преподаватель обращает внимание студентов на важность правильной организации и проведения статистического исследования.

2. Диагностика знаний студентов по теме: «Организация статистического исследования»

Преподаватель выясняет исходный теоретический уровень знаний в изучении организации статистического исследования.

3. Инструктаж по выполнению работы и контроль над ходом ее выполнения.

На примере типового задания рассматривается методика проведения статистического исследования. Далее студенты самостоятельно под контролем преподавателя выполняют варианты индивидуальных заданий по данной теме.

VIII. ИНФОРМАЦИОННЫЙ БЛОК ТЕМЫ

1. Цель и задачи статистического исследования, рабочая гипотеза. Этапы статистического исследования. Методика проведения статистического исследования

Прежде, чем приступить к проведению статистического исследования, необходимо иметь точное представление о том, что, когда, в каком объеме и при помощи каких методов предполагается изучить. Детальное знакомство лица, проводящего исследование, с изучаемой проблемой позволит избежать грубых ошибок, а, следовательно, фиктивных выводов и обобщений.

Первоначально определяется ***цель*** исследования, которая должна давать ответ на вопрос: *зачем проводится данное исследование.* Четкое определение цели необходимо для *составления программы* и *плана* статистического исследования.

Разрабатываются ***задачи***, отвечающие на вопрос: *какие конкретные действия последовательно ведут к достижению намеченной цели.*

Формулируется ***рабочая гипотеза*** – основная идея исследования, которую предстоит доказать или опровергнуть в ходе дальнейшей работы.

Определяется ***объект*** исследования, которым может быть *генеральная* или *выборочная* статистическая совокупность. При использовании в качестве объекта последней следует обеспечить ее *репрезентативность* (тема 3).

Этапы статистического исследования включают:

- Составление программы и плана исследования;
- Регистрацию и сбор материала;
- Обработку и сводку данных;
- Анализ полученных результатов.

Методика проведения статистического исследования

Для каждого этапа статистического исследования характерна своя методика проведения.

Первый этап – составление программы и плана статистического исследования:

1) Разработка программы статистического исследования.

1.1. Выбор единицы наблюдения как основного признака изучаемого явления.

1.2. Определение учетных признаков (расшифровка единицы наблюдения).

1.3. Разработка образца первичного статистического наблюдения.

1.4. Выбор правильных группировок (определение групп по учетным признакам).

1.5. Составление макета статистических таблиц (сводка единичных случаев по качественной однородности).

2) Составление плана статистического исследования.

2.1. Определение (постановка) цели исследования.

2.2. Выбор объекта исследования.

2.3. Определение места (территории) исследования.

2.4. Установление сроков исследования.

2.5. Определение сил (кадров) и средств (финансирование) для проведения исследования.

2.6. Определение организационно - методического руководства исследованием.

2.7. Установление охвата наблюдения.

2.8. Определение путей формирования объекта исследования (статистической совокупности).

Второй этап – регистрация и сбор статистических данных:

1) Получение (сбор) необходимой информации.

2) Заполнение учетных документов.

Третий этап – обработка и сводка данных:

1) Контроль (проверка) качества учетных документов (первичной документации).

2) Шифровка (кодирование) статистического материала.

3) Раскладка (группировка) материала по однородным группам.

4) Сводка материала (заполнение статистических таблиц).

5) Вычисление статистических критериев (статистические расчеты).

6) Графическое представление полученных результатов.

7) Текстовое оформление полученных данных.

Четвертый этап – проведение статистического анализа:

1) Интерпретация полученных результатов и их графическое изображение.

2) Окончательное текстовое оформление работы.

3) Формулировка выводов исследования.

4) Предложения по внедрению полученных результатов в практику.

2. Составление программы и плана исследования. Регистрация и сбор материала. Формирование репрезентативной выборочной статистической совокупности

Программа определяет, что и в каком направлении надо исследовать, и представляет собой комплекс задач по сбору, обработке и анализу статистического материала. Программа также предусматривает *вид статистического наблюдения* (сплошное или выборочное; единовременное или текущее), *способ получения информации, силы и средства*, необходимые для выполнения работы.

Программа статистического исследования включает *программу сбора материала*, *программу обработки материала* и *программу анализа данных*.

I. В состав ***программы сбора материала*** входит выбор *единицы наблюдения* и определение *учетных признаков*.

Единица наблюдения является основным признаком изучаемого явления. Из единиц наблюдения в итоге складывается *статистическая совокупность*, представляющая собой *объект исследования*. Дополняют и расформировывают каждую единицу наблюдения ***учетные признаки***.

Перечень учетных признаков, подлежащих регистрации, заносится в ***первичный статистический документ*** официального образца или образца, разработанного исследователем. Первичный статистический документ может иметь вид бланка, карты, таблицы или журнала. Государственные статистические документы (статистический талон для регистрации уточненных диагнозов, листок нетрудоспособности, медицинская карта амбулаторного больного и др.) могут быть использованы в качестве официальной программы сбора материала по заболеваемости. В клинико-статистических исследованиях часто используются анкеты, специальные карты для выборки данных из медицинских документов, опросники и т.д.

По времени наблюдения формирование совокупности осуществляется путем *текущего* или *единовременного наблюдения*.

При текущем наблюдении статистические данные собирают путем регистрации каждого отдельного случая по мере его возникновения за определенный промежуток времени (изучение рождаемости, смертности, заболеваемости и т.д.).

При единовременном наблюдении регистрацию данных приурочивают к моменту времени, т.е. сбор всех наблюдений осуществляется

единовременно. Единовременное наблюдение отражает состояние изучаемого явления на определенный момент (перепись населения, медицинские осмотры населения и др.).

Сбор сведений различается ***по способу наблюдений***: *непосредственный, выкопировка, анкетный, опрос.*

Способ выкопировки широко распространен при проведении социально-гигиенических исследований. При выкопировке из учетных документов берутся лишь необходимые данные, соответствующие цели исследования и предусмотренные его программой.

Анамнестический способ получения информации основан на регистрации сведений, полученных *анкетным методом* или *методом опроса.*

Анкетный метод целесообразно применять в тех случаях, когда необходимо получить объективные сведения в глубоко личных, интимных сторонах жизни. Анкета в этих случаях должна быть анонимной.

Для использования ***метода опроса*** необходимо наличие важного условия – доверия опрашиваемого. При этом сведения получают во время беседы врача с обследуемым лицом по определенному кругу вопросов.

II. ***Программа обработки данных*** представляет собой составление макетов статистических таблиц. В статистических таблицах выделяют следующие элементы:

- ***Заглавие***, отражающее содержание таблицы;
- ***Статистическое подлежащее*** – основной анализируемый признак;
- ***Статистическое сказуемое*** – признаки, которые дополняют и раскрывают подлежащее;
- ***Итоги*** (по горизонтали и вертикали), которые используются для проверки заполнения таблицы и вычисления показателей.

Статистические таблицы бывают 3-х видов:

- *С одним статистическим подлежащим* (сводка материала производится по одному признаку). Например, распределение больных в отделении А по полу;
- ***Групповая*** статистическая таблица *с одним статистическим подлежащим и несколькими несвязанными между собой статистическими сказуемыми.* Например, распределение больных в том же отделении А, но уже не только по полу, но и по возрасту;
- ***Комбинационная*** таблица – имеет *одно или несколько взаимосвязанных подлежащих и несколько взаимосвязанных сказуемых.* Например, распределение больных в вышеупомянутом отделении А по эффективности лечения, полу и возрасту.

III. ***Программа анализа данных*** – представляет собой перечень адекватных статистических методов, с помощью которых будут выявлены закономерности изучаемого явления.

План является организационной частью и указывает где, в каком количестве, как, когда и кем будет выполняться работа по реализации поставленных задач.

В плане статистического исследования должны быть отражены *сроки* выполнения его этапов. Если предполагается изучение выборочной совокупности, то в плане статистического исследования необходимо указать ***способ отбора единиц*** для выборочной совокупности: *случайный, механический, типологический, серийный*.

Случайный – это такой отбор, когда все единицы генеральной совокупности имеют равную возможность попасть в выборку. Способом такого отбора является жеребьевка (лотерея) и применение таблиц случайных чисел.

Жеребьевка заключается в том, что на каждую единицу наблюдения заготавливают стандартного вида карточку, в которой отмечают порядковый номер по списку. Затем их перемешивают в закрытом ящике и в случайном порядке отбирают для данного наблюдения определенное количество карточек.

Правило пользования таблицей случайных чисел.

Всем единицам наблюдения, из которых предстоит отобрать их необходимое количество, присваивают номера от 1 до n. Затем в таблице случайных чисел находят необходимое количество таких чисел, которые не превышают n.

Механический отбор производится по какому-либо признаку (первая буква фамилии, номер истории болезни и т.д.), от которого зависят результаты исследования. При механическом отборе единицы совокупности сначала располагаются в каком-либо порядке (по алфавиту или по номеру истории болезни). После этого производится механический отбор единиц через какой-либо интервал (каждую 4-ю, или каждую 10-ю историю болезни). Интервал при механическом отборе предварительно рассчитывается. Для этого общее число единиц совокупности делится на число единиц, которые надо отобрать.

Типологический отбор позволяет производить выбор единиц наблюдения из типичных групп всей генеральной совокупности. Для этого внутри генеральной совокупности все единицы группируются по какому-нибудь признаку в типичные группы (например, по возрасту). Затем из каждой такой группы производят отбор (случайным или механическим способом) необходимого числа единиц таким образом, чтобы соотношение размеров групп (например, возрастных) в выборочной совокупности сохранялось такое же, как и в генеральной.

Серийный отбор предусматривает выбор из генеральной совокупности не отдельных единиц, а выбор серий (гнезд). Для этого вся генеральная совокупность разбивается на относительно однородные серии. Отбор серий осуществляется путем случайной или механической выборки. При этом отбор должен производиться так, чтобы каждой серии генеральной совокупности была бы обеспечена одинаковая воз-

можность быть отобранной в выборочную совокупность. В каждой серии обследуются все составляющие ее единицы наблюдения.

Особенности формирования выборочных совокупностей контрольных и опытных групп.

Важнейшим принципом этого процесса - опытная и контрольная группы максимально похожи во всем, кроме изучаемого фактора воздействия (например, метода лечения). Одним из способов, удовлетворяющих этим требованиям, является ***парно-сопряженный отбор*** или ***метод уравновешивания групп***. При этом каждой единице наблюдения в исследуемой группе подбирают пару в контрольной группе.

Формирование выборочной совокупности можно осуществлять по комплексу признаков, т.е. применять *метод направленного отбора.*

Направленный отбор определяет границы совокупности (например, определенный возрастно-половой состав) и обеспечивает возможность углубленного изучения отдельных групп населения. Однородность таких групп по ряду признаков позволяет уменьшить численность наблюдений при организации выборочных исследований и получить обоснованные выводы.

Одним из видов направленного отбора является *когортный метод.*

Когорта – это статистическая совокупность, которая состоит из относительно однородных элементов, объединенных наступлением определенного признака и прослеженного за один и тот же интервал времени (например, для определения численности детей, рожденных в семье за определенный интервал супружеской жизни, исследуется когорта лиц, имеющих единый срок вступления в брак и единую продолжительность супружеской жизни).

Репрезентативность *выборочной статистической совокупности обеспечивается предварительным расчетом необходимого объема выборки (числа наблюдений).*

Если в результате исследования конечный результат должен быть выражен средними величинами, то необходимый объем выборки рассчитывается по формуле:

$$N = \frac{t^2 \cdot \sigma^2}{\Delta^2}$$

где **N** – число наблюдений;

t – доверительный коэффициент, который при вероятности безошибочного прогноза (γ), равной 95%, составляет 2, при γ = 99% равен 3;

σ – среднее квадратическое отклонение;

Δ – предельная ошибка (задается в соответствии со спецификой измеряемого признака, целью и задачами исследования).

Если в результате исследования конечный результат должен быть выражен в процентах (показателях), объем выборки рассчитывается по формуле:

$$N = \frac{t^2 \cdot pq}{\Delta^2}$$

где N -- число наблюдений;

t – доверительный коэффициент, который при вероятности безошибочного прогноза (р), равной 95%, составляет 2, при р= 99% равен 3;

p – частота вариант обладающих данным признаком (выбирается с расчетом, чтобы произведение pq было максимальным, так если p = q = 50%);

q – частота вариант, не обладающих данным признаком (p + q = 1 или q = 100%– p%);

Δ – предельная ошибка (обычно не превышает 5%).

3. Обработка и сводка статистических данных. Анализ статистических данных

Обработка и сводка данных – третий этап статистического исследования. Этот этап включает в себя ряд операций:

- ***Контроль качества*** учетных документов;
- ***Шифровка*** (кодирование);
- ***Группировка*** (распределение единиц наблюдения по однородным группам для подсчета);
- ***Составление сводки*** (заполнение статистических таблиц).
- ***Вычисление статистических показателей***;
- ***Графическое изображение*** показателей.

Четвертый этап – это ***статистический анализ полученных данных***. Условиями его проведения являются:

- Понимание сущности изучаемого явления;
- Владение методикой статистического исследования;
- Корректировка предыдущих этапов статистического исследования.

При ***анализе*** материала проводится более глубокое изучение всех взаимодействующих факторов. На этом этапе осуществляется осмысление, сравнение, проводится обсуждение всех полученных результатов с последующей формулировкой выводов.

В практическом здравоохранении любое исследование должно быть направлено на получение объективных выводов с последующим их внедрением, что позволяет выделить еще один пятый этап – **внедрение полученных данных**.

IX. УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА СТУДЕНТОВ

(использовано руководство к практическим занятиям по социальной гигиене и организации здравоохранения./ Под редакцией Ю.П. Лисицина, Н.Я.Копыта. М.Медицина,1984г.).

Типовое задание (*Вариант 1)*

Цель исследования – изучить длительность пребывания в терапевтическом отделении стационара и факторы, на нее влияющие, чтобы наметить пути повышения эффективности использования коечного фонда.

Требуется

1. В соответствии с целью:
– сформулировать основные задачи исследования;
– определить единицу наблюдения;
– составить программу сбора данных (см. образец – карту больного, госпитализированного в терапевтическое отделение стационара) (учетный документ);
– рассчитать необходимый объем наблюдений для вычисления средней длительности пребывания в стационаре больных при вероятности безошибочного прогноза (γ), равной 95%, среднем квадратическом отклонении (σ), рассчитанном при пробном исследовании, 2,5 дня и предельной ошибке (Δ) 0,5 дня;
– составить макет комбинационной таблицы для разработки и сводки данных о влиянии на длительность пребывания в стационаре возраста и пола больных.

2. Провести шифровку (в соответствии с международной классификацией болезней, травм и причин смерти или на другой основе) диагнозов в картах больных, выбывших из стационара.

3. Провести разработку карт больных, выбывших из стационара, и заполнить макет статистической таблицы, используя данные таблицы 6.1.

Таблица 6.1 Распределение больных, прошедших через терапевтическое отделение стационара, по длительности лечения, возрасту, полу (в абс. числах)

Длительность лечения в днях	Возраст в годах												Итого
	до 20		21–30		31–40		41–50		51–60		старше 60		
	М.	Ж.	М.	Ж.	М.	Ж.	М.	Ж.	М.	Ж.	М.	Ж.	
до 10	0	8	2	0	0	0	0	0	0	0	0	0	**10**
10–14	0	0	0	0	8	4	2	2	0	0	0	0	**16**
15–19	0	0	0	0	0	6	2	4	4	4	0	0	**20**
20–24	0	0	0	0	0	0	2	2	2	10	0	0	**16**
25–29	0	0	0	0	0	0	0	0	0	6	8	10	**24**
30 и более	0	0	0	0	0	0	0	0	0	0	6	8	**14**
Всего	**0**	**8**	**2**	**0**	**8**	**10**	**6**	**8**	**6**	**20**	**14**	**18**	**100**

4. Вычислить показатели, характеризующие распределения больных по длительности пребывания в стационаре, и представить их в виде диаграммы.

Образец выполнения типового задания

Основные задачи исследования:

1. Провести анализ длительности стационара лечения больных терапевтического профиля в зависимости от вида патологии, возраста, пола, метода диагностики и лечения.

2. Изучить влияние на сроки стационарного лечения лечебно-профилактических мероприятий на догоспитальном этапе.

3. Разработать мероприятия, направленные на повышение эффективности использования коечного фонда, сокращение длительности стационарного лечения и повышение качества диагностики и лечения.

Единица наблюдения – больной, лечившийся в стационаре терапевтического отделения в истекшем году.

Определение необходимого объема наблюдений:

$$n = \frac{t^2\sigma^2}{\Delta^2} = \frac{2^2 \cdot 2{,}5^2}{0{,}5^2} = \frac{4 \cdot 6{,}25}{0{,}25} = 100$$

КАРТА

больного, госпитализированного в терапевтическое отделение стационара в 20__ г.

№ карты стационарного больного_____
Фамилия, имя, отчество_____
Дата поступления _____Дата выписки_____
Длительность стационарного лечения (дней)_____
Диагноз при поступлении_____
Вид госпитализации (плановая, экстренная, без направления)
Тяжесть состояния при поступлении (удовлетворительное, средней тяжести, тяжелое)
Диагноз при выписке_____
Совпадение диагноза поликлиники и стационара (полное, частичное, несовпадение)
_____
Диагностические исследования на догоспитальном этапе_____
_____
Диагностические исследования в стационаре_____
_____
Дублирование диагностических исследований поликлиники и стационара (да, нет)
_____
Длительность диагностического периода в стационаре (дни)_____
Лечебные мероприятия догоспитального _____
Лечебные мероприятия госпитального этапа_____
Исход заболевания (выздоровление, улучшение, без изменений, летальный исход)
_____
Экспертная оценка:
диагностических исследований на догоспитальном этапе (в полном или неполном объеме)_____
_____
лечения на догоспитальном этапе (без дефектов или с дефектами)_____
диагностических исследований на госпитальном этапе (проведены своевременно или несвоевременно)_____
лечения на госпитальном этапе (без дефектов или с дефектами)_____
_____

Шифровка. При шифровке заключительного диагноза карты больного, выбывшего из стационара, по международной классификации болезней, травм причин смерти необходимо найти заболевание в алфавитном указателе и по номеру рубрики, соответствующей заболеванию, определить класс, которому соответствует заболевание, указанное в диагнозе (см. статистическую карту выбывшего из стационара).

Варианты заданий на организацию статистического исследования для самостоятельного выполнения

Вариант 2. Изучить первичную заболеваемость населения, обслуживаемого поликлиникой А. выявить контингенты лиц, отличающихся повышенными уровнями первичной заболеваемости. Разработать

комплекс мероприятий, направленных на улучшение здоровья выявленного контингента.

Единицей наблюдения считать первое посещение в изучаемом году по поводу данного заболевания. В качестве учетных документов использовать статистические талоны (ф. 025–2/у), талоны амбулаторного пациента (ф. 025-6/у, ф. 025–7/у) и амбулаторные карты (ф. 025/у).

Вариант 3. Изучить общую заболеваемость населения, обслуживаемого поликлиникой **Б**. Выявить контингенты лиц, отличающихся самыми низкими уровнями общей заболеваемости с целью активизации первичной и вторичной профилактики в обслуживаемом районе. Разработать комплекс мероприятий по сохранению и укреплению здоровья лиц названных контингентов.

Единицей наблюдения считать обращение населения за медицинской помощью. В качестве учетных документов использовать статистические талоны (ф. 025–2/у), талоны амбулаторного пациента (ф. 025–6/у, ф. 025–7/у) и амбулаторные карты (ф. 025/у).

Вариант 4. Изучить госпитализированную заболеваемость населения районного центра **А** в динамике за 3 года. Разработать комплекс мероприятий в работе поликлиники и стационара ЦРБ.

Единицей учета считать случай госпитализации больного в стационар по поводу заболевания. В качестве учетных документов использовать «Статистическую карту выбывшего из стационара» (ф. 066/у), Статистический талон (ф. 025–2/у), а также карту стационарного больного (ф. 003).

Вариант 5. Изучить острую инфекционную заболеваемость в населенном пункте Б за 5 лет. Разработать комплекс мероприятий по снижению острой инфекционной заболеваемости.

Единицей учета считать каждый случай острого инфекционного заболевания. В качестве учетных документов использовать «Журнал учета инфекционных заболеваний» ф. 60/у, «отчет о заболеваниях гриппом и другими острыми респираторными заболеваниями» (ф. 3), «Отчет об отдельных инфекционных и паразитарных заболеваниях за год» (ф. 2)

Вариант 6. Изучить заболеваемость с временной утратой трудоспособности (ВУТ) на промышленном предприятии **А** за 5-летний период. Выявить возрастные и профессиональные группы с наиболее высокими уровнями ВУТ. Разработать комплекс мероприятий по ее снижению.

Единицей учета считать случай потери трудоспособности в связи с заболеванием у работающих. В качестве учетных документов использовать листки нетрудоспособности, а также справки (ф. 095/у).

Вариант 7. Провести углубленное изучение заболеваемости с ВУТ за 3 года на промышленном предприятии **Б** в цеху №1. Выявить наиболее значимые производственные факторы, влияющие на состояние здоровья работающих. Разработать комплекс мероприятий. Направленных на оптимизацию условий труда на рабочих местах цеха №1.

Единицей учета считать каждый случай потери трудоспособности в связи с ВУТ у «круглогодовых» рабочих. В качестве учетных документов использовать листки нетрудоспособности.

Вариант 8. Изучить заболеваемость на промышленном предприятии **В** по данным периодического осмотра (одномоментная форма проведения).

Единицей учета считать каждый случай впервые выявленного заболевания. В качестве учетных документов использовать «Вкладыш в амбулаторную карту» и заключительный акт проведения периодического медицинского осмотра.

Вариант 9. Изучить заболеваемость населения в городе **А** по данным о причинах смерти. Разработать комплекс мероприятий по снижению смертности взрослого трудоспособного населения.

Единицей учета считать каждый случай смерти среди лиц трудоспособного возраста. В качестве учетных документов использовать «Врачебное свидетельство о смерти» (ф. –93).

Вариант 10. Изучить первичную инвалидность в населенном пункте **С**, связанную с трудовым увечьем. Разработать комплекс мероприятий, направленных на снижение первичной инвалидности с производственным травматизмом.

Единицей учета считать каждый случай первичной инвалидности связанную с трудовым увечьем. При составлении программы статистического исследования использовать «Статистический талон учета экспертной и консультативной деятельности МРЭК». В качестве учетных документов использовать «Отчет о первичной инвалидности по причинам» (ф. №1–инв.), «Отчет о числе инвалидов в разрезе городов и районов» (ф. №4–инв.).

Ответы к заданиям по темам

Тема 1. СТАТИСТИЧЕСКИЕ СОВОКУПНОСТИ. СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ.

Ответы к заданию №1
«Вычисление относительных величин»

Показатели	№ варианта			
	2	3	4	5
Посещаемость на 1000 населения	10000	7216	8036	5455
Уровень обращаемости (общий)	1143	866	800	945
Доля обращений лиц 70 и более лет	2,5	4	3	3
Число врачей на 1000 населения	7,1	5,2	7,1	4,5

Показатели	№ варианта			Показатели	№ варианта		
	6	7	8		9	10	11
Посещаемость на 1000 детей	4167	3667	4000	Число педиатрических коек на 1000 населения	1	1	4
Уровень обращаемости (общий)	833	800	800	Уровень проф.обращаемости (общий)	1400	1250	1500
Доля обращений детей 4-14 лет	50	44	44	Доля проф.обращений детей 4-14 лет	67	64	47
Число врачей на 1000 детского населения	3	3,1	3	Показатель наглядности (дети 4-14 лет)	24	22	27

Показатели	№ варианта	Показатели	№ варианта	Показатели	№ варианта
	12		13		14
Число прошедших проф. осмотр на 1000 населения	200	Число прошедших проф.осмотр на 1000 населения	133	Посещаемость на 1000 женского населения	1667
Уровень проф. выявляемости нарушений желудочно-кишечного тракта (общий)	14	Уровень выявляемости снижения остроты зрения среди детей (общий)	55	Уровень обращаемости (общий)	333
Доля проф. выявлений больных учащихся ВУЗов среди осмотренных учащихся	14	Доля детей дошкольного возраста со снижением остроты зрения среди осмотренных	36	Доля обращений женщин (35-49 лет)	58
Показатель наглядности (учащиеся ВУЗов)	66	Показатель наглядности (учащиеся 2-8 классов)	108	Показатель наглядности (35-49 лет)	199

Ответы к заданию №2
«Вычисление показателей динамического ряда»

Итого за все годы	№ варианта						
	2	3	4	5	6	7	8
Абсолютный прирост (убыль)	-159,3	62,6	-221,9	1,7	-0,8	2,5	-2,5
Темп прироста (убыли) %	-1,60	0,90	-7,9	18,9	-5,5	-45,5	-32,5
Значение 1% прироста (убыли)	99,56	69,56	28,1	0,09	0,15	-0.05	0,08
Темп роста (убыли)	98,4%	100,9%	92,1%	118,9%	94,5%	54,5%	67,5%

Итого за все годы	№ варианта					
	9	10	11	12	13	14
Абсолютный прирост (убыль)	-1,3	-3,4	-9	1,0	-0,1	-5,9
Темп прироста (убыли) %	-1,1	-9,5	-30,0	8,5	-0,4	-12,1
Значение 1% прироста (убыли)	1,18	0,36	0,30	0,12	0,25	0,49
Темп роста (убыли)	98,9%	90,5%	70,0%	108,5%	99,6%	87,9%

Тема 2. ВАРИАЦИОННАЯ СТАТИСТИКА. СРЕДНИЕ ВЕЛИЧИНЫ

Ответы к заданию №1
«Составление простого вариационного ряда и расчет простого среднего арифметического при малом числе наблюдений (N < 30)»

Показатели	№ варианта					
	2	3	4	5	6	7
Среднее арифметическое	25,4	69,4	18774	98250	174,5	81,2

Показатели	№ варианта						
	8	9	10	11	12	13	14
Среднее арифметическое	37,3	133	376	57	134	17	127

Ответы к заданию №2
«Составление простого вариационного ряда и расчет медианы, моды, взвешенного среднего арифметического»

Показатели	№ варианта					
	2	3	4	5	6	7
Мода	21	3	64	78	34	25
Медиана	19,5	6	68	79	38,5	28
Взвешенное среднее арифметическое	18,5	5,56	69,83	79,14	38,50	28,06

Показатели	№ варианта						
	8	9	10	11	12	13	14
Мода	32	14	2,5	49	176	58,2	36,6
Медиана	28,5	16,5	2,7	50	179,5	64,6	37,4
Взвешенное среднее арифметическое	28,69	17,5	2,85	50,85	179,58	64,22	37,40

Ответы к заданию №3
«Составление сгруппированного вариационного ряда, расчет среднего арифметического по способу моментов и определение среднего квадратического отклонения»

Показатели	№ варианта					
	2	3	4	5	6	7
Среднее квадратическое отклонение	4,7	2,2	5,7	4,2	4,0	4,1

Показатели	№ варианта						
	8	9	10	11	12	13	14
Среднее квадратическое отклонение	4,0	4,3	0,5	2,7	6,1	6,5	0,8

Тема 3. ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Ответы к заданию №1
«Определение стандартной ошибки среднего и границ доверительного интервала при малом числе наблюдений (N < 30)»

Показатели	№ варианта					
	2	3	4	5	6	7
Предельная ошибка среднего (Δ) при α=0,05	3,31	10818,08	34156,77	4,43	17,41	0,99
Предельная ошибка среднего (Δ) при α=0,01	5,01	17939,29	62689,71	6,36	27,31	1,64

Показатели	№ варианта						
	8	9	10	11	12	13	14
Предельная ошибка среднего (Δ) при α=0,05	9,13	25,68	3,39	7,17	5,41	16,72	2,88
Предельная ошибка среднего (Δ) при α=0,01	13,84	42,59	4,93	10,87	8,19	27,72	4,77

Ответы к заданию №2
«Определение стандартной ошибки среднего и границ доверительного интервала при малом числе наблюдений (N > 30)»

Показатели	№ варианта					
	2	3	4	5	6	7
Предельная ошибка среднего (Δ) при α=0,05	0,79	1,96	1,43	1,45	1,45	1,37
Предельная ошибка среднего (Δ) при α=0,01	1,06	2,64	1,92	1,94	1,95	1,83

Показатели	№ варианта						
	8	9	10	11	12	13	14
Предельная ошибка среднего (Δ) при α=0,05	1,56	0,17	0,94	1,97	2,26	0,27	1,51
Предельная ошибка среднего (Δ) при α=0,01	2,10	0,23	1,27	2,63	3,04	0,36	2,02

Ответы к заданию №3
«Определение достоверности различий между средними выборок по критерию Стьюдента»

Показатели	№ варианта					
	2	3	4	5	6	7
T	2,4	2,6	7,3	0,9	7,3	2,4
TTECT	0,006	0,015	0,00006	0,35	0,00004	0,005

Показатели	№ варианта						
	8	9	10	11	12	13	14
T	15,2	3,6	3,4	5,5	1,2	6,3	2,7
TTECT	0,000000002	0,0013	0,009	0,00007	0,21	0,000003	0,014

Ответы к заданию №4
«Определение достоверности соответствия между видами частотных распределений двух выборок по критерию хи-квадрат»

Показатели	№ варианта					
	2	3	4	5	6	7
χ^2	2,2	2,2	2,02	$1*10^{-8}$	2,02	1,2
ХИ2ТЕСТ	0,95	0,98	0,99	1,00	0,99	0,95

Показатели	№ варианта						
	8	9	10	11	12	13	14
χ^2	1,7	9,4	0,0025	1,25	1,95	1,25	2,2
ХИ2ТЕСТ	0,98	0,40	0,99	0,99	0,96	0,99	0,98

Тема 4. ИЗУЧЕНИЕ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

Ответы к заданию №1
«Определение параметрического коэффициента корреляции методом линейной корреляции Пирсона»

Показатели	№ варианта					
	2	3	4	5	6	7
Коэффициент корреляции Пирсона	0,689	0,878	0,882	0,982	0,438	0,607

Показатели	№ варианта						
	8	9	10	11	12	13	14
Коэффициент корреляции Пирсона	0,6	0,773	0,982	0,877	0,907	0,968	0,889

Ответы к заданию №2
«Определение непараметрического коэффициента ранговой корреляции Спирмена»

Показатели	№ варианта					
	2	3	4	5	6	7
Коэффициент корреляции Спирмена	0,93	0,88	0,97	0,97	0,89	0,9

Показатели	№ варианта						
	8	9	10	11	12	13	14
Коэффициент корреляции Спирмена	0,98	0,91	0,97	0,94	0,96	0,96	0,97

Тема 5. СТАНДАРТИЗАЦИЯ ПОКАЗАТЕЛЕЙ
Ответы к заданию №1
«Вычисление стандартизованных показателей методом прямой стандартизации»

Показатели	№ варианта					
	2	3	4	5	6	7
Стандартизованные показатели	1,4	8,8	10,5	0,4	9,6	3,3
	2,2	3,8	22,4	0,4	15,3	2,8

Показатели	8	9	10	11	12
Стандартизованные показатели	728,4	0,7	0,1	6,2	3,7
		0,5			
	1659,8	0,3	0,2	10,8	4,1

ИСПОЛЬЗОВАНИЕ ПАКЕТА АНАЛИЗА В EXCEL

В электронных таблицах Excel помимо мастера функций имеется набор более мощных инструментов для анализа данных, называемый *пакетом анализа*, который может быть использован для решения задач обработки данных.

Установка раздела *Анализ данных* в Excel:

- в меню *Сервис* выбрать команду *Надстройки*;
- в появившемся списке установить флажок *Пакет анализа* (при его отсутствии запустить программу инсталляции Office и добавить в Excel пакет анализа);
- при следующем вызове меню *Сервис* в списке добавится строка *Анализ данных.*

Ввод данных. Для использования пакета анализа исследуемые данные следует представить в виде таблицы, где столбцами являются соответствующие показатели, а строками – наблюдения. Информация вводится в отдельные ячейки. Совокупность ячеек, содержащих анализируемые данные, называется входным диапазоном.

Рассмотрим некоторые процедуры пакета анализа.

Описательная статистика используется для создания одномерного статистического отчета. В результате анализа в указанном выходном диапазоне для каждого столбца данных выводятся следующие статистические характеристики: среднее значение, стандартная ошибка (среднего), медиана, мода, стандартное отклонение, дисперсия выборки, эксцесс, асимметричность, интервал, минимум, максимум, сумма, счет, наибольшее и наименьшее значение, уровень надежности.

Двухвыборочный t-тест. Анализ двух выборок (t-тест Стьюдента). Об особенностях выбора t-теста для конкретных данных рассказывалось выше.

Результаты анализа. В выходной диапазон будут выведены: средняя, дисперсия и число наблюдений для каждой переменной, гипотетическая разность средних, число степеней свободы (df), значение t-статистики, P (T<=t) одностороннее, t критическое одностороннее, P (T<=t) двухстороннее, t критическое двухстороннее.

Интерпретация результатов. Если величина вероятности случайного появления анализируемых выборок (P(T<=t) двухстороннее) меньше уровня значимости (0,005), принято считать, что различия между выборками не случайные, т.е. различия достоверные.

Корреляционный анализ. Процедура позволяет получить корреляционную матрицу, содержащую коэффициенты корреляции между различными параметрами.

Результаты анализа. В выходной диапазон будет выведена корреляционная матрица, в которой на пересечении соответствующих строки и столбца находится коэффициент корреляции между соответст-

вующими параметрами. Ячейки выходного диапазона, имеющие совпадающие координаты строк и столбцов, содержат значение 1, так как каждый столбец во входном диапазоне полностью коррелирует с самим собой.

Регрессионный анализ. Линейный регрессионный анализ заключается в подборе графика и его уравнения для набора наблюдений. Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных. В пакете анализа экспериментальные данные апроксимируются линейным уравнением, содержащим до 16 переменных:

$$Y=a_1x_1+a_2x_2+ \ldots +a_{16}x_{16},$$

где Y – зависимая переменная, x_1-x_{16} – независимые переменные, a_0-a_{16} – искомые коэффициенты регрессии.

Результаты анализа. Выходной диапазон будет включать в себя результаты дисперсионного анализа, коэффициенты регрессии, стандартную погрешность вычисления Y, среднеквадратичные отклонения, число наблюдений, стандартные погрешности для коэффициентов.

Интерпретация результатов. Значения коэффициентов регрессии находятся в столбце *Коэффициенты* и соответствуют:

Y-пересечение – a_0;

Переменная x_1 – a_1;

Переменная x_2 – a_2 и т.д.

В столбце *P-значение* приводится достоверность отличия соответствующих коэффициентов от нуля. В случаях, когда $p>0,05$, коэффициент может считаться нулевым; это означает, что соответствующая независимая переменная практически не влияет на зависимую переменную.

Приводимое значение *R-квадрат* характеризует, с какой степенью точности полученное регрессионное уравнение апроксимирует исходные данные. Если R-квадрат больше 0,95, говорят о высокой степени апроксимации (модель хорошо описывает явление). Если R-квадрат лежит в диапазоне от 0,8 до 0,95, говорят об удовлетворительной апроксимации (модель в целом адекватна описываемому явлению). Если R-квадрат меньше 0,6, принято считать, что точность апроксимации недостаточна и модель требует улучшения (введения независимых переменных, учета нелинейности и т.д.).

Дисперсонный анализ предназначен для исследования задачи о действии на применяемую величину (отклик) одного или нескольких факторов. Причем в однофакторном, двухфакторном и т.д. анализе влияющие на результат факторы считаются известными, и речь идет только о выяснении существенности или оценке этого влияния.

Результаты анализа. Выходной диапазон будет включать в себя результаты дисперсионного анализа: средние, дисперсии, критерий Фишера и другие показатели.

Интерпретация результатов. Влияние исследуемого фактора определяется по величине значимости критерия Фишера, который нахо-

дится в таблице *Дисперсионный анализ* на пересечении строки *Между группами* и столбца *Р-значение*. В случаях, когда Р-Значение <0,05, критерий Фишера значим и влияние исследуемого фактора можно считать доказанным.

Другие возможности пакета анализа.

Мы рассмотрели далеко не все возможности применения пакета анализа в статистических исследованиях. Например, Excel является эффективным средством проведения численного моделирования. При этом часть параметров предполагается известными (исходные параметры), а часть рассчитывается по формулам. Меняя все возможные сочетания значения исходных параметров, можно наблюдать за изменением расчетных параметров и анализировать получаемые результаты. Excel производит такие расчеты быстро и без ошибок, предоставляя пользователю в считанные секунды множество вариантов решения поставленной задачи, на основании которой можно выбрать наиболее приемлемое.

ОСНОВНЫЕ МОДУЛИ СИСТЕМЫ STATISTICA

Статистические процедуры системы STATISTICA сгруппированы в нескольких специализированных статистических модулях. В каждом модуле вы можете выполнить определенный способ обработки, не обращаясь к процедурам из других модулей. Ниже дается краткое описание статистических модулей и их назначение.

Модуль «Основные статистики и таблицы»

Если вы хотите провести предварительную обработку данных, осуществить разведочный анализ, определить зависимости между ними, разбить их различными способами на группы, просмотреть эти группы визуально и определить взаимосвязи между данными, запустите модуль ***Basic Statistics/Tables — Основные статистики/Таблицы.*** Обычно с этого модуля начинается работа в системе. Этот статистический модуль включает в себя описательные статистики, группировки, разведочный анализ.

STATISTICA предлагает широкий выбор методов разведочного статистического анализа. Система может вычислить практически все описательные статистики, включая медиану, моду, квартили, определенные пользователем процентили, средние и стандартные отклонения, доверительные интервалы для среднего, коэффициенты асимметрии, эксцесса (с их стандартными ошибками), гармоническое и геометрическое среднее, а также многие другие описательные статистики. Возможен выбор критериев для тестирования нормальности распределения (критерий Колмогорова-Смирнова, Лилиефорса, Шапиро-Уилкса). Проведению разведочного анализа помогает широкий выбор графиков.

Практически все описательные статистики могут быть вычислены для данных, разделенных на группы с помощью одной или нескольких группирующих переменных. Например, всего несколько щелчков мыши позволяют разбить данные на группы, например, *пол* и *возраст*, и посмотреть гистограммы (по категориям) на графике, а также графики значений, матричные и многие другие типы графиков. Если выбрано более двух групповых переменных, то каскады соответствующих графиков могут быть получены автоматически.

Корреляции. Этот раздел включает большое количество средств, позволяющих исследовать зависимости между переменными. Можно вычислть практически все общие меры зависимости, включая коэффициент корреляции Пирсона, коэффициент ранговой корреляции Спирмена, Tau (b,c) Кендалла, Гамма, коэффициент сопряженности признаков С и многие другие. Корреляционные матрицы могут быть вычислены и для данных с пропусками, если использовать специальные методы обработки пропущенных значений.

t - критерии (и другие критерии для групповых различий). t-критерии для зависимых и независимых выборок, а также статистика Хоттелинга также вычисляются (см. также ANOVA/MANOVA).

Таблицы частот и таблицы кросстабуляций. В модуле содержится обширный набор процедур, обеспечивающих табулирование непрерывных, категориальных, дихотомических переменных, переменных, полученных в результате многовариативных опросов. Вычисляются как кумулятивные, так и относительные частоты, а также так называемые кросстабулированные частоты. Вычисляются статистики Пирсона, максимального правдоподобия, Йетс-коррекция, хи-квадрат, статистики Фишера, Макнемера и многие другие.

Модуль «Множественная регрессия»

Если вы хотите построить зависимости между многомерными переменными, подобрать простую линейную модель и оценить ее адекватность, воспользуйтесь модулем ***Multiple Regression – Множественная регрессия.***

Модуль Множественной регрессии включает в себя исчерпывающий набор средств множественной линейной и фиксированной нелинейной (в частности, полиномиальной, экспоненциальной, логарифмической и др.) регрессии, включая пошаговые, иерархические и другие методы, а также ридж-регрессию.

Можно анализировать чрезвычайно большой объем данных для регрессионных задач (до 300 переменных в процедуре разведочной регрессии). STATISTICA также содержит Модуль нелинейного оценивания, с помощью которого могут быть оценены практически любые определенные пользователем нелинейные модели, включая логит, пробит регрессию и др.

Модуль «Дисперсионный анализ. Анализ вариантов» – General ANOVA/MANOVA

Если вам известны факторы, которые воздействуют на измеряемые вами данные, и вы хотите оценить степень этого воздействия, понять, существенно оно или нет, воспользуйтесь модулем ANOVA/MANOVA.

ANOVA/MANOVA – модуль представляет собой набор процедур общего одномерного и многомерного дисперсионного и ковариационного анализа.

В модуле возможно решение задач практически неограниченной сложности. Можно определить все решения в наиболее прямых "функциональных" терминах и даже пользователи, имеющие малый опыт работы с ANOVA, могут анализировать очень сложные задачи.

Допускается анализ планов с межгрупповыми и повторно изме-

ренными факторами, включая планы с фиксированными/случайными факторами, а также несбалансированные планы, планы с изолированными контрольными группами и т.д. В любом виде планов могут быть использованы фиксированные или изменяющиеся ковариации. Возможен анализ неполных планов (латинские, греко-латинские, планы с единственным наблюдением в ячейке, рандомизированно-блоковые и т. д.). Дополнительно пользователь может ввести собственные слагаемые в ошибку для анализа дисперсии.

В модуле может быть выполнен также контрастный анализ практически неограниченной сложности. Для анализа неполных планов со случайным размещением пропущенных ячеек средства контрастного анализа будут автоматически отмечать все пустые ячейки и помогут вам при построении гипотез.

В модуле представлен самый широкий выбор статистических процедур для проверки основных предположений дисперсионного анализа, в частности, критерии Бартлетта, Кохрана, Хартли, Бокса и других. Как и во всех других модулях, здесь обеспечена всесторонняя графическая поддержка статистического анализа.

Модуль «Дискриминантный анализ»

Если перед вами стоит задача, как по результатам измерений отнести объект к одному из нескольких классов, например, в медицине при обследовании больных, или в геологии при оценке перспективности месторождений, или в банковской деятельности и управлении, следует применить модуль ***Discriminant Analysis – Дискриминантный анализ*** из раздела многомерных исследовательских методов.

Методы дискриминантного анализа позволяют построить на основе ряда предположений классификационное правило отнесения объекта к одному из нескольких классов, минимизируя некоторый разумный критерий, например, вероятность ложной классификации или заданную пользователем функцию потерь. Выбор критерия определяется пользователем из соображений ущерба, который он понесет из-за ошибок классификации. Исторически первой в дискриминантном анализе была модель Фишера, в которой предполагается, что наблюдаемые векторы имеют многомерное нормальное распределение с невырожденной ковариационной матрицей и вектором средних, разным для разных классов.

В простейших задачах распределение наблюдаемых величин задано точно для каждого класса. В сложных задачах распределение известно частично и для построения классификатора приходится привлекать дополнительную информацию.

Модуль дискриминантного анализа системы STATISTICA содержит полный набор процедур для множественного пошагового функционального дискриминантного анализа. STATISTICA позволяет выполнять пошаговый анализ, как вперед, так и назад, а также внутри опреде-

ленного пользователем блока переменных в модели. В дополнении к многочисленным графикам и диагностикам, описывающим дискриминантные функции, предлагается широкий набор опций и статистик для классификации "новых" или "старых" случаев (при оценивании справедливости модели).

Модуль «Непараметрической статистики и подгонка распределений»

Для проверки различных гипотез о характере распределения ваших данных следует применить модуль ***Nonparametrics/Distribution – Непараметрическая статистика/ Распределения.***

Модуль содержит обширный набор непараметрических критериев согласия, в частности, критерий Колмогорова-Смирнова, ранговые критерии Манна-Уитни, Вальда-Вольфовица, Вилкоксона и многие другие.

Все реализованные ранговые критерии применимы и в случае совпадающих рангов и используют поправки для малых выборок.

Статистические процедуры модуля позволяют пользователю легко сравнить распределение наблюдаемых величин с большим количеством различных теоретических распределений. Вы можете подогнать к данным нормальное, равномерное, линейное, экспоненциальное, Гамма, логнормальное, хи-квадрат, Вейбулла, Гомпертца, биномиальное, Пуассоновское, геометрическое распределения, распределение Бернулли. Точность подгонки оценивается с помощью критерия хи-квадрат или одновыборочного критерия Колмогорова-Смирнова (параметры подгонки можно контролировать). Выполняются также тесты Лиллифорса и Шапиро-Уилкса.

Дополнительно подгонка гипотетического распределения к эмпирическому может быть оценена в пользовательских гистограммах (стандартных или кумулятивных) с наложением на них выбранных функций; линейные и столбчатые графики ожидаемых и наблюдаемых частот могут быть получены непосредственно из таблицы с выходными данными. Определенные пользователем функции от одной и двух переменных так же легко могут быть построены и наложены на графики.

Модуль «Факторный анализ»

Если вы хотите сжать данные или выделить основные общие факторы, влияющие на наблюдаемые вами характеристики сложного объекта и объясняющие связи между ними, например, основные социально-экономические факторы или факторы, влияющие на результаты голосования, работайте в модуле ***Factor Analysis – Факторный анализ*** из раздела мнегомерных исследовательских методов.

Модуль Факторного анализа содержит широкий набор методов и опций, снабжающих пользователя исчерпывающими средствами фак-

торного анализа.

Он, в частности, включает в себя метод главных компонент, метод минимальных остатков, метод максимального правдоподобия и др. с расширенной диагностикой и чрезвычайно широким набором аналитических и разведочных графиков.

Модуль «Многомерное шкалирование»

Если вы хотите представить данные о близости объектов какой-либо простой пространственной моделью, в которой объекты интерпретируются, например, как города на обычной карте, а различия между ними есть просто расстояния, в частности, данные о странах, политических партиях, автомобилях и т. д., и всесторонне диагностировать модель, используйте модуль ***Multidimensional Scaling – Многомерное шкалирование.***

Модуль содержит всесторонний инструментарий для выполнения многомерного шкалирования. Матрицы подобия, различия и корреляции могут быть вычислены для большого числа переменных (до 90 переменных) с размерностью до 9 компонент. Начальная конфигурация может быть вычислена автоматически с помощью анализа главных компонент либо задана пользователем. Имеются всевозможные метрики.

Модуль «Анализ надежности»

Модуль включает широкий набор процедур для разработки и вычислений надежности сложных объектов на основе результатов обследований и диагностики отдельных узлов. Так же, как во всех модулях, STATISTICA могут быть проанализированы чрезвычайно большие объемы данных, например, таблицы, содержащие до 300 наименований.

Пользователь может вычислять статистики надежности для всех позиций таблицы, интерактивно выбирать подмножества или производить сравнения между подмножествами позиций.

Если в процессе исследований интерактивно удаляются какие-либо позиции таблицы, то немедленно вычисляется новая надежность. Вывод включает корреляционные матрицы и описательную статистику, полную таблицу дисперсионного анализа и многое другое. Интеллектуальную поддержку процессу анализа обеспечивают большое число специализированных графиков и широкий выбор интерактивных процедур типа "что-если".

Модуль «Кластерный анализ»

Если вы хотите произвести сложную иерархическую классификацию данных или выделить в них кластеры, обратитесь к модулю ***Claster***

Analysis – Кластерный анализ.

Модуль содержит всесторонний инструментарий для кластеризации (метод k-средних, иерархическая классификация и др.). Возможные метрики: Евклидова, Манхэттенская, Чебышевская и др. Может быть выполнен кластерный анализ чрезвычайно большого объема.

Модуль «Лог-линейный анализ»

Если вам нужно проанализировать сложные многоуровневые таблицы, как это бывает, например, в социологии при анализе массовых опросов, воспользуйтесь модулем ***Log-Linear Analysis – Лог-линейный анализ.***

Модуль содержит полный инструментарий лог-линейных процедур для многоуровневых таблиц частот. Могут быть проанализированы оба типа таблиц – полные и неполные (со структурными нулями). Таблицы частот могут быть построены из необработанных данных или непосредственно введены в систему. В процессе обработки пользователь может просматривать как исходную таблицу, так и маргинальные таблицы. Также имеется процедура автоматического выбора наилучшей модели.

Вывод результатов включает в себя G-квадрат, стандартный пирсоновский хи-квадрат, с подходящими степенями свободы и уровнями значимости, наблюдаемые и ожидаемые таблицы, маргинальные таблицы и другие характеристики. Графические возможности, доступные в модуле, включают разнообразные 2- и 3-мерные графики, спроектированные для визуализации таблиц частот (включая интерактивные, управляемые пользователем каскады гистограмм, 3-мерные гистограммы, позволяющие просматривать послойно таблицы с несколькими входами), графики наблюдаемых и подогнанных частот, разнообразные графики остатков и многое другое.

Модуль «Нелинейное оценивание»

Если вы хотите определить нелинейные зависимости в данных, подогнать к ним функциональные кривые, обратитесь к модулю ***Nonlinear Estimation – Нелинейное оценивание.***

Модуль предоставляет возможность осуществить подгонку к наблюдаемым данным кривой, по существу, любого типа.

Одна из уникальных возможностей модуля (в отличие от традиционных программ нелинейного оценивания) в том, что в нем не накладываются ограничения на размер файла обрабатываемых данных.

Оценки могут быть построены с помощью оценок метода наименьших квадратов, метода максимального правдоподобия или заданной пользователем функции потерь.

Пользователь может выбрать одну из четырех вычислительных

процедур: квазиньютоновский метод, симплекс-метод, метод Хука-Дживса, метод Розенброка. Пользователь может определить любой тип нелинейной модели, набрав соответствующее уравнение в редакторе. Кроме того, допускается использование логических операторов. Благодаря этому могут быть оценены разрывные регрессионные модели и модели, включающие индикаторные переменные.

Программа осуществляет полный контроль за всеми аспектами вычислительных процедур (начальное значение, размер шага, критерий сходимости и т. д.). Большинство обычных нелинейных регрессионных моделей задано в модуле и может быть просто выбрано из меню.

Модуль «Каноническая корреляция»

Модуль включает в себя широкий набор процедур для выполнения канонического корреляционного анализа; исследования связи между двумя множествами переменных. Модуль может обрабатывать векторные данные или корреляционные матрицы и вычислять все стандартные канонические корреляционные статистики (включая собственные вектора, собственные значения, коэффициенты избыточности, канонические веса, нагрузки, критерии значимости для каждого корня и т.д.).

Модуль «Анализ длительностей жизни»

Если ваши данные представляют собой таблицы жизней, как, например, в медицине или страховании, и вы хотите проанализировать их, воспользуйтесь модулем ***Survival Analysis – Анализ длительностей жизни.***

Модуль предлагает обширный набор методов анализа цензурированных данных из социологии, биологии, медицины, а также процедуры, используемые в инженерии и маркетинге (например, контроль качества, оценка надежности и др.). Дополнительно вычисляются таблицы жизни с различными описательными статистиками и оценками Каплана-Мейера.

Программа обеспечивает полную поддержку 4 основных разведочных моделей анализа: модель Кокса пропорциональных рисков, экспоненциальную регрессионную модель, логнормальную и нормальную регрессионные модели с расширенными диагностиками.

Если вы хотите оценить вероятность отказов сложной системы на основе статистики отказов, воспользуйтесь методами этого модуля.

Модуль «Анализ временных рядов и прогнозирование»

Предположим, вы исследуете какой-либо физический процесс и хотите оценить спектр мощности или исследуете экономический ряд данных, например курс акций, и хотите спрогнозировать его на не-

сколько дней вперед. Здесь полезен модуль ***Time Series/Forecasting – Анализ временных рядов и прогнозирование.***

Анализ временных рядов активно применяется в деловых, научных и инженерных исследованиях. Модуль предлагает широкий набор методов анализа. Имеются всесторонние возможности преобразования исходного временного ряда, большой набор графиков и статистических процедур. Для восстановления пропущенных наблюдений могут быть использованы интерполяционные методы.

Модуль состоит из нескольких общих процедур, объединенных вокруг динамического графического представления временных рядов и их сглаживающих/моделирующих преобразований. Пользователь может смоделировать одновременно несколько рядов и выполнять интерактивный "что-если" анализ, наблюдая графическое предаствление ряда. Подобно всем графикам в системе STATISTICA, графики рядов могут быть представлены на экране в увеличенном режиме, сжатыми, развернутыми или сдвинутыми один относительно другого.

Методы преобразования рядов включают: исключение среднего, тренда, взвешенное скользящее среднее, медианное сглаживание, фильтрацию, взятие разностей с любым сдвигом и многое другое.

Модуль также включает процедуры автокорреляционного анализа. Процедуры ARIMA (авторегрессии и проинтегрированного скользящего среднего) содержат полный набор инструментов для сезонной и несезонной процедуры ARIMA с встроенными преобразованиями, прогнозирование и специальную графику. Также воможен анализ моделей с интервенцией. В модуле имеется всесторонний набор процедур Фурье (спектрального) анализа с графиками периодограмм, спектральной плотности, амплитудами и фазами и т. д. При построении оценок спектральной плотности доступен широкий выбор спектральных окон, в том числе задаваемых пользователем окон.

Диалог *Анализ распределенных лагов* позволяет построить регрессию одного ряда на другой. В нем можно оценить модели как с фиксированными лагами, так и с неизвестными заранее.

Модуль «Моделирование структурными уравнениями» (SEPATH)

Если вы хотите построить и тестировать различные модели, объясняющие структуру связей между наблюдаемыми переменными, воспользуйтесь модулем ***SEPATH.***

Моделирование структурными уравнениями – мощная техника многомерного статистического анализа, развитая в последние годы и имеющая целью соединить статистические методы с методами теории систем.

SEPATH является сложным модулем. Требуется приложить немало усилий, чтобы овладеть им.

В заключение напомним, что компьютерные методы анализа – это

лишь реализация, способ применения математических методов. Применяя математические методы необходимо учитывать, что математический анализ экспериментальных наблюдений -- это только часть исследования, которую следует рассматривать в контексте всех проблем, определяющих качество научной работы. Формулировка основной задачи исследования, выбор соответствующего задаче метода и способа организации исследования, особенности отбора и характер полученных данных – все это является первичным и определяет выбор адекватного метода математического анализа и, в конечном счете, влияет на достоверность полученных результатов. При этом активное использование компьютерных методов позволит поднять на более высокий уровень применение математических методов анализа в проводимом научном исследовании.

ЛИТЕРАТУРА

1. Перечень использованной литературы

1. **Боровиков, В.** STATISTICA: искусство анализа данных на компьютере. Для профессионалов. – Спб.: Питер, 2001. – 656 с.
2. Организационные технологии в системе здравоохранения XXI века: монография // А.Н. Косинец [и др.]; под ред. А.Н. Косинца, В.С. Глушанко. – Витебск: ВГМУ, 2006. – 328 с.
3. **Бочкин, А.И., Колбасич, Л.П.** Применение статистических методов анализа данных в научно-исследовательских работах: Учебно-метод. пособ. – Витебск: Изд-во ВГУ им. П.М. Машерова, 2006. – 39 с.
4. **Вялков, А.И.** Управление и экономика здравоохранения: учеб. пособие / Под ред. А.И. Вялкова; Вялков А.И., Кучеренко В.З., Райзберг Б.А. и др. – 3-е изд., доп. – М.: ГЭОТАР-Медиа, 2009. – 644 с.
5. **Гараничева, С.Л.** Основы информационных технологий: Учеб. пособ. - Витебск: Изд-во ВГМУ, 2001 – 201 с.
6. **Гараничева, С.Л.** Теория и практика подготовки студентов медицинских вузов к применению информационных технологий: монография; под ред. Ю.Я. Родионова. – Витебск: Изд-во ВГМУ, 2004 – 152 с.
7. **Гельман, В.Я.** Решение математических задач средствами Excel. – СПб: Питер, 2003. – 237 с.
8. **Гельман, В.Я.** Электронная таблица Excel для врачей – СПб.: СПб МАПО. 2000 – 384 с.
9. **Глушанко, В.С.** Общественное здоровье и здравоохранение: пособие, 2-е изд., перераб. и доп. / В.С. Глушанко. Под ред. д-ра мед. наук, профессора В.П. Дейкало. – Витебск: ВГМУ, 2011. – 491с
10. **Зайцев В.М., Лифляндский В.Г.**: Прикладная медицинская статистика. – СПб: СПб ГМА им. И.И. Мечникова. 2000 – 302 с.
11. **Кучеренко, В.З., Флек, О.В., Вялкова, Г.М.** и др. Оценка эффективности деятельности медицинских организаций / под ред. А.И. Вялкова. – М.: ГЭОТАР-МЕД, 2004.
12. **Лисицын, Ю.П.** Общественное здоровье и здравоохранение. – М.: ГЭОТАР-МЕД, 2002.
13. **Макарова, Н.В., Трофимец В.Я.** Статистика в Excel: учеб. пособие. М.: Финансы и статистика, 2006. – 368 с.
14. **Мак-Федрис, П.** Формулы и функции в Microsoft Excel 2003. Пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 576 с.
15. **Мартыненко, В.Ф., Вялкова, Г.М., Полесский, В.А.** и др. Информационные технологии в системе управления здравоохранением Российской Федерации / под ред. А.И. Вялкова. – М.: ГЭОТАР-МЕД, 2006.
16. **Медик, М.А., Токмачев, М.С.** Математическая статистика в медицине: учеб. пособие / В.А. Медик, М.С. Токмачев. – М.: Финансы и статистика, 2007. -800 с.:ил

17. **Мидлтон, М.Р.** Анализ статистических данных с использованием Microsoft Excel для Officce XP. Пер. с англ.; Под ред.Г.М. Кобельникова. – М: БИНОМ, 2005. – 296 с.
18. **Рудикова Л.В.** Microsoft Excel для студента. – СПб.: БХВ-Петербург, 2006. – 369 с.
19. **Урбах, В.Ю.** Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975. – 296 с.
20. **Шупляк, В.И.** Математическая статистика: курс лекций / В.И. Шупляк.– Минск: РИВШ, 2011. – 228 с.

2. Перечень рекомендуемых изданий кафедры ВГМУ

1. **Глушанко, В.С.** Система контроля качества медицинской помощи в организациях здравоохранения / **В.С. Глушанко, А.П. Тимофеева, Д.В. Мороз.** - Витебск: ВГМУ, 2006. – 74 с.
2. **Глушанко, В.С.** Организация стационарной медицинской помощи населению. Анализ деятельности стационарных организаций здравоохранения / **В.С. Глушанко, А.П. Тимофеева, Д.В. Мороз, Т.В. Колосова.** – Витебск: ВГМУ, 2006. – 95 с.
3. **Глушанко, В.С.** Организация стационарной медицинской помощи населению. Анализ деятельности стационарных организаций здравоохранения (часть II) / **В.С. Глушанко, А.П. Тимофеева, Д.В. Мороз, Т.В. Колосова**. – Витебск: ВГМУ, 2008. – 16 с.
4. **Глушанко, В.С.** Общественное здоровье и здравоохранение как наука и предмет преподавания. / **В.С. Глушанко, А.П. Тимофеева**. – Витебск: ВГМУ, 2004. – 18 с.
5. **Глушанко, В.С.** Заболеваемость населения, ее виды. Методы изучения: методическая разработка для студентов высших медицинских учреждений образования по дисциплине «Общественное здоровье и здравоохранение» / **В.С. Глушанко, А.П. Тимофеева, А.В. Плиш**; под ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ, 2004. – 45 с.
6. **Глушанко, В.С.** Организация амбулаторно-поликлинической, скорой и неотложной помощи городскому населению. Диспансеризация: учебно-методическое пособие для студентов высших медицинских учреждений образования по дисциплине «Общественное здоровье и здравоохранение» / **В.С. Глушанко, А.В. Плиш, А.П. Тимофеева**; под ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ, 2004. – 79 с.
7. **Глушанко, В.С.** Медико-социальная экспертиза: учебно-методическое пособие для студентов высших медицинских учреждений образования по дисциплине «Общественное здоровье и здравоохранение» / **В.С. Глушанко, А.В. Плиш**; под ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ, 2004. – 103 с.
8. **Глушанко, В.С.** Анализ деятельности организаций здравоохранения службы охраны здоровья матери и ребенка: учебно-методическое пособие / **В.С. Глушанко, А.П. Тимофеева, Т.В. Колосова, Д.В. Мороз**; под ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ, 2010. – 239 с.
9. **Глушанко, В.С. Колосова, Т.В. Мороз, Д.В. Тимофеева, А.П. Плиш, А.В.** / Врачебная производственная практика по дисциплине "Общественное здоровье и здравоохранение" // Под ред. д.м.н., профессора

В.С.Глушанко: учебно-методическое пособие. - Витебск: Издательство ВГМУ, 2004. - 69 с.

10. **Глушанко, В.С.** Врачебная производственная практика по дисциплине «Общественное здоровье и здравоохранение» (II часть): Учеб.-метод. пособие / **В.С. Глушанко, Т.В. Колосова, А.П. Тимофеева**; под ред. доктора мед. наук, проф. В.С. Глушанко. – Витебск, ВГМУ, 2009. – 50 с.
11. **Глушанко, В.С.** Врачебная производственная практика по дисциплине «Общественное здоровье и здравоохранение»: Учеб.-метод. указания для иностранных студентов высших медицинских учреждений образования / **В.С. Глушанко, Т.В.Колосова**; под ред. доктора мед. наук, проф. В.С. Глушанко. – Витебск, ВГМУ, 2008. – Ч. II. – 6 с.
12. **Глушанко, В.С.** Современные проблемы совершенствования медико-социальной помощи: монография / **В.С. Глушанко, В.В, Колбанов, И.В. Левицкая, И.Н. Мороз, С.А. Морозова, Т.Г. Светлович, С. Эглофф**: под ред. Ген. Секретаря БОКК В.В. Колбанова. – Витебск: ВГМУ, 2008. – 185с.
13. **Петрище, Т.Л.** История медицины и фармации: Курс лекций / Т.Л. Петрище; под ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ, 2010. – 567 с.
14. **Плиш, А.В.** Медико-экономическая модель управления ресурсосберегающими технологиями в здравоохранении / **А.В. Плиш, В.С. Глушанко**; под ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ. – 2005. –324 с.
15. Сборник вопросов и ответов по клинической медицине / **Под общей редакцией чл.-корр. БелАМН, профессора А.Н. Косинца**. – Витебск: ВГМИ, 1998. – 948 с.
16. Экономика здравоохранения. Экономический анализ деятельности организаций здравоохранения: методическая разработка для студентов высших медицинских учреждений образования по дисциплине «Общественное здоровье и здравоохранение» / **В.С. Глушанко, Т.В. Колосова, А.В. Хоменко**; под. ред. д.м.н., проф. В.С. Глушанко. – Витебск: ВГМУ, 2004. - 27 с.
17. **Glushanko, V.S.** Public health and public health services: Lectures-presentations, computer tests and practical skills for the students of the faculties of preparation of foreign citizens with the Russian and English languages training: educational supply / V.S.Glushanko, T.V.Kolosova, A.P.Timofeeva: edited by V.S.Glushanko. - Vitebsk: VSMU, 2008. – 442 p. (**Глушанко, В.С.** Общественное здоровье и здравоохранение: Лекции-презентации, компьютерные тесты и практические навыки для студентов факультетов подготовки иностранных граждан с русским и английским языком обучения: учеб. пособие / **В.С. Глушанко, Т.В. Колосова, А.П. Тимофеева**; под ред. В.С. Глушанко. – Витебск: ВГМУ, 2008. – 442 с.).
18. **Glushanko, V.S.** Public health and public health services: Lectures-presentations, computer tests and practical skills for the students of the faculties of preparation of foreign citizens with the Russian and English languages training: educational supply / V.S.Glushanko, T.V.Kolosova, A.P.Timofeeva: edited by V.S.Glushanko. – [Электронный ресурс]. – Vitebsk: VSMU, 2008. – 442 p. – Электрон.дан. и прогр. (17,4 Мб). – 1 электрон. опт. Диск (CD-room).

ОГЛАВЛЕНИЕ

БЛОК 1	СТАТИСТИЧЕСКИЕ СОВОКУПНОСТИ. СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ.....	**3 – 33**
БЛОК 2	ВАРИАЦИОННАЯ СТАТИСТИКА. СРЕДНИЕ ВЕЛИЧИНЫ.....	**34 – 53**
БЛОК 3	ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ.....	**54 – 87**
БЛОК 4	ИЗУЧЕНИЕ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ.....	**87 – 104**
БЛОК 5	СТАНДАРТИЗАЦИЯ ПОКАЗАТЕЛЕЙ.....	**105 – 119**
БЛОК 6	ОРГАНИЗАЦИЯ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ.....	**120– 133**
БЛОК 7	ОТВЕТЫ К ЗАДАНИЯМ ПО ТЕМАМ.....	**134– 138**
БЛОК 8	ИСПОЛЬЗОВАНИЕ ПАКЕТА АНАЛИЗА В EXCEL.....	**139 141**
БЛОК 9	ОСНОВНЫЕ МОДУЛИ СИСТЕМЫ STATISTICA.....	**1432– 150**
БЛОК 10	ЛИТЕРАТУРА.....	**151– 153**

Библиотека ВГМУ
0 0 0 3 4 9 9 2

Учебное издание

ГЛУШАНКО Василий Семенович,
ГРУЗНЕВИЧ Алла Павловна,
ГАРАНИЧЕВА Светлана Леонидовна,
АЛЯХНОВИЧ Наталья Сергеевна и др.

ОСНОВЫ МЕДИЦИНСКОЙ СТАТИСТИКИ

Учебно-методическое пособие

Редактор *Глушанко В.С.*
Технический редактор *Борисов И.А.*
Компьютерная верстка *Симоненко А.А.*
Корректор *Симоненко А.А.*

Подписано в печать 24.02.2012 г. Формат бумаги 60x84 1/16.
Бумага типографская № 2.
Ризография. Усл. печ. 9,01. Уч.-изд. л. 9,69. Тираж 200. Заказ № 177.

Издатель и полиграфическое исполнение:
УО «Витебский государственный медицинский университет»
ЛИ № 02330/0549444 от 8.04.09 г.
Пр-т Фрунзе, 27, 210602, г. Витебск.