

МП-19

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-ОСЕТИНСКАЯ ГОСУДАРСТВЕННАЯ МЕДИЦИНСКАЯ АКАДЕМИЯ»
МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**СБОРНИК
МЕТОДИЧЕСКИХ РЕКОМЕНДАЦИЙ
ПО ИНФОРМАТИКЕ, МЕДИЦИНСКОЙ ИНФОРМАТИКЕ И
СТАТИСТИКЕ ДЛЯ СТУДЕНТОВ 1 КУРСА МЕДИКО-
ПРОФИЛАКТИЧЕСКОГО ФАКУЛЬТЕТА**

ЧАСТЬ I

Владикавказ 2023 г.

Сборник методических рекомендаций составлен в соответствии:

1. ФГОС ВО по специальности «Медико-профилактическое дело», утвержденным Министерством образования и науки РФ 16.07.2017 г.
2. Учебным планом по специальности «Медико-профилактическое дело», утвержденным Ученым Советом ФГБОУ ВО СОГМА Минздрава России.
3. Рабочей программы учебной дисциплины информатика, медицинская информатика и статистика.

СОСТАВИТЕЛИ:

- доцент А.В. Бабенко

Рецензенты:

доцент кафедры химии и физики ФГБОУ ВО СОГМА Минздрава России,
к.п.н. Боцьева Н.И.

Утверждена на заседании ЦКУМС
“24”мая 2023 г. Протокол № 5

ОГЛАВЛЕНИЕ

Тема 1 «Основные понятия и методы математической статистики».....	4
Тема 2: «Основные понятия и методы математической статистики. «Применение пакета Statistica для анализа временных рядов».....	24
Тема 3: «Методы математической статистики. Подготовка данных к статистическому анализу. Элементы организации медико-статистического исследования. Статистическая совокупность. Статистические величины. Вычисление статистических величин» величин».....	34
Тема 4-5: «Меры оценки разнообразия признака в совокупности и типичности средних. Оценка достоверности результатов исследования. Определение доверительных границ	48
Тема 6: «Методы математической статистики. Обзор графических возможностей системы Statistica»	63
Тема 7-8: «Оптимизация диагностики. МИС диагностических служб»	75

Тема 1: «Основные понятия и методы математической статистики»

1. Научно-методическое обоснование темы:

Теория вероятностей – математическая наука, изучающая закономерности в явлениях и опытах, результаты которых не могут быть заранее предсказаны.

Возникновение теории вероятностей как науки относят к средним векам. Первоначальным толчком к развитию теории вероятностей послужили задачи, относящиеся к азартным играм, таким, как орлянка, кости, карты, рулетка, когда в них начали применять количественные подсчеты и прогнозирование шансов на успех.

Другим толчком для развития теории вероятностей послужило страховое дело, а именно с конца XVII века на научной основе стало производиться страхование от несчастных случаев и стихийных бедствий. Стала зарождаться новая наука, вырисовываться ее специфика и методология: определения, теоремы, методы.

В ТВ вводятся специальные понятия и строятся специфические математические модели. Исходными понятиями в ТВ являются понятие случайного события и вероятности.

Слово «статистика» происходит от латинского слова «status» - состояние, положение. Впервые это слово в середине XVIII века применил немецкий ученый Ахенваль при описании состояния государства (нем. Statistk, от итал. stato - государство).

Статистика:

1) вид практической деятельности, направленной на сбор, обработку, анализ и публикацию статистической информации, характеризующей количественные закономерности жизни общества (экономики, культуры, политики и др.).

2) отрасль знаний (и соответствующие ей учебные дисциплины), в которой излагаются общие вопросы сбора, измерения и анализа массовых количественных данных.

Статистика как наука включает разделы: общая теория статистики, экономическая статистика, медицинская статистика, отраслевые статистики и др.

Общая теория статистики излагает общие принципы и методы статистической науки.

Как каждая наука, статистика имеет свой **предмет исследования** – массовые явления и процессы общественной жизни, свои **методы исследования** – статистические, математические, разрабатывает системы и подсистемы показателей, в которых отражаются размеры и качественные соотношения общественных явлений.

Статистика изучает количественные уровни и соотношения общественной жизни в неразрывной связи с их качественной стороной.

Статистика возникла на базе математики, и широко пользуется **математическими методами**. Это выборочный метод исследования, основанный на математической теории вероятности и законе больших чисел, различные методы обработки вариационных и динамических рядов, измерение корреляционных связей между явлениями и др.

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

Главная задача статистики, как и всякой другой науки, заключается в установлении закономерностей изучаемых явлений.

2. Краткая теория:

1. Элементы теории вероятности.

1.1. Основные понятия теории вероятности.

Испытание (опыт, эксперимент) – это выполнение определенного комплекса условий, в которых наблюдается то или иное явление, фиксируется тот или иной результат.

В теории вероятностей рассматриваются **испытания**, результаты которых нельзя предсказать заранее, а сами испытания можно повторять, хотя бы теоретически, произвольное число раз при неизменном комплексе условий. Испытаниями, например, являются: подбрасывание монеты, выстрел из винтовки, проведение денежно-вещевой лотереи.

Испытание, в котором событие A наступило, называется **успешным**, в противном случае – **неудачным**.

Случайным событием (*возможным событием* или просто *событием*) называется любой факт, который в результате испытания может произойти или не произойти.

Случайное событие – это не какое-нибудь происшествие, а лишь возможный **исход**, результат испытания (опыта, эксперимента). События обозначаются прописными (заглавными) буквами латинского алфавита: A, B, C .

1. Событие называется **достоверным**, если оно всегда происходит.
2. Событие, которое никогда не произойдет, называется **невозможным**.
3. **Суммой** двух событий называется такое событие, которое наступает тогда и только тогда, когда наступает хотя бы одно из слагаемых ($A \cup B$).
4. **Произведением** двух событий называется такое событие, которое наступает тогда и только тогда, когда наступают оба события ($A \cap B$).
5. **Разностью** двух событий называется событие, которое наступает тогда, когда наступает одно событие и не наступает другое ($A \setminus B$).
6. Говорят, что событие A **влечет за собой** событие B , если из наступления A всегда следует наступление B ($B \subset A$).

7. Если одновременно $B \subset A$ и $A \subset B$, то в этом случае события A и B называются **равносильными**.
8. События A и B называются **несовместными**, если наступление одного из них исключает появление другого в одном и том же испытании.
9. События A и B называются **совместными** если они могут произойти вместе в одном и том же испытании.

Пример 1. Испытание состоит в однократном подбрасывании игральной кости с шестью гранями. Событие A – появление трех очков, событие B – появление четного числа очков, C – появление нечетного числа очков. События A и C совместны, поскольку число 3 – нечетное, а значит, если выпало 3 очка, то произошло и событие A и событие C . Кроме того, событие A влечет за собой событие C . События A и B несовместны, т.к. если произошло и событие A , то не произойдет событие B , а если произошло событие B , то не произойдет событие A . События B и C также являются несовместными.

10. События называются **попарно несовместными** (или **взаимоисключающими**), если любые два из них несовместны.

Пример 2. Испытание – сдача студентом экзамена по определенной дисциплине. События – соответственно студент получит на экзамене один балл, два, три и т.д. Эти события являются попарно несовместными.
1021, ..., AAA

События образуют **полную группу** для данного испытания, если они попарно несовместны и в результате испытания обязательно появится одно из них. $nAAA, \dots, 21$

В примере 2 события образуют полную группу, а события – нет.
1021, ..., AAA621, ..., AAA

11. События называются **равновозможными**, если нет оснований считать, что одно из них является более возможным, чем другое.

Примеры равновозможных событий: выпадение любого числа очков при броске игральной кости, появление любой карты при случайном извлечении из колоды, выпадение герба или цифры при броске монеты и т.п.

1.2. Классическое определение вероятности.

Для практической деятельности важно уметь сравнивать события по степени возможности их наступления. Например, интуитивно ясно, что при последовательном извлечении из колоды пяти карт более возможна ситуация, когда появились карты разных мастей, чем появление пяти карт одной масти;

при десяти бросках монеты более возможно чередование гербов и цифр, нежели выпадение подряд десяти гербов, и т.д. поэтому для сравнения событий нужна определенная мера.

Численная мера степени объективной возможности наступления события называется **вероятностью события** и является, наряду с понятием случайного события, вторым основным понятием теории вероятности.

Множество всех взаимоисключающих исходов эксперимента называется **пространством элементарных событий**. Пространство элементарных событий будем обозначать буквой Ω , а его исходы – буквой ω , т.е. $\omega \in \Omega$.

Пусть производится испытание с конечным числом равновозможных исходов $\omega_1, \omega_2, \dots, \omega_n$, образующих полную группу событий.

Пусть число возможных исходов равно n (общее число элементарных исходов), а при m из них происходит некоторое событие A (число благоприятных исходов), тогда при сделанных ранее предположениях на испытание, **вероятностью $P(A)$ случайного события A** , наступившего в данном испытании вычисляется по формуле:

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

Это, так называемое, **классическое определение вероятности**.

Пример 3. из урны, содержащей 6 белых и 4 черных шара, наудачу вынут шар. Найти вероятность того, что он белый.

Решение. Будем считать элементарными событиями, или исходами опыта, извлечение из урны каждого из имеющихся в ней шаров. Число возможных исходов равно 10, а число благоприятных исходов (появлению белого шара) – 6 (количество белых шаров). Значит,

$$P(A) = \frac{m}{n} = \frac{6}{10} = 0.6$$

Данная формула справедлива только в случае всех равновозможных исходов. Она может применяться только для очень узкого класса задач. В случае, когда исходы не равновозможные, требуется определять вероятность события другим способом. Для этого введем вначале понятие **относительной частоты (частоты) $W(A)$** события A как отношение числа опытов, в которых наблюдалось событие A , к общему количеству проведенных испытаний:

$$W(A) = \frac{M}{N}$$

где N – общее число опытов, M – число опытов, в которых появилось событие A . Т.е. **статистической вероятностью события** считается его относительная частота или число, близкое к ней.

1.3. Основные теоремы.

Теорема сложения вероятностей.

Теорема 1. Вероятность $P(A+B)$ суммы событий A и B равна

$$P(A+B) = P(A) + P(B) - P(AB).$$

$$P(A+B) = \frac{m_A + m_B + m_{AB}}{n} = \frac{m_A}{n} + \frac{m_B}{n} - \frac{m_{AB}}{n} = P(A) + P(B) - P(AB)$$

что и требовалось доказать.

Следствие. Сумма вероятностей противоположных событий равна 1:

$$P(A) + P(\bar{A}) = 1$$

Теорема умножения вероятностей

Остановимся более подробно на следующем примере иллюстративного характера. Допустим, что студент из 30 билетов успел выучить билеты с 1-го по 3-й и с 28-го по 30-й. На экзамен он пришел одиннадцатым, и оказалось, что к его приходу остались только билеты с 1-го по 20-й (событие A). Вероятность события $B = \{\text{студент получил выученный билет}\}$ без дополнительной информации о том, что событие A произошло, может быть вычислена по классическому определению с $\Omega = \{1, 2, \dots, 30\}$.

Согласно формуле $P(B) = \frac{6}{30} = \frac{1}{5}$

При дополнительной информации (событие A произошло) множество возможных исходов A состоит из 20 элементарных исходов, а событие B вместе с A наступает в 3 случаях. Следовательно, в рассматриваемом примере естественно определить **условную вероятность $P(B|A) = P_A(B) =$**

$P(B|A) = P_A(B) = \frac{P(AB)}{P(A)}$ **события B при условии, что событие A произошло,**

как $P_A(B) = \frac{3}{20}$.

Теорема 2 (теорема умножения). Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое событие произошло:

$$P(AB) = P(A) * P_A(B)$$

Доказательство. Воспользуемся обозначениями теоремы 1. Тогда для вычисления $P_A(B)$ множеством возможных исходов нужно считать m_A (так как A произошло), а множеством благоприятных исходов – те, при которых произошли и A , и B (m_{AB}). Следовательно,

$$P_A(B) = \frac{m_{AB}}{m_A} = \frac{m_{AB}}{n} \cdot \frac{n}{m_A} = \frac{P(AB)}{P(A)}$$

откуда следует утверждение теоремы.

1.4. Формула полной вероятности

Теорема 3 (формула полной вероятности). Если событие A может произойти только при условии появления одного из событий (гипотез) $H_i=1, \dots, n$ образующих полную группу, то вероятность события A равна

$$P(A) = \sum_1^n P(H_i) \cdot P_{H_i}(A)$$

Группа событий H_1, \dots, H_n называется **полной группой событий**, если выполняются следующие условия:

1. $P(A_i) > 0, i = \overline{1, n}$
2. $A_i \cdot A_j = \emptyset, i, j = \overline{1, n}, i \neq j$
3. $\sum_1^n A_i = \Omega$

Теорема 3 (формула Байеса). Пусть H_1, \dots, H_n полная группа, ($P(A) > 0$). Тогда имеем место формула

$$P_A(H_i) = \frac{P(H_i) \cdot P_{H_i}(A)}{P(A)} = \frac{P(H_i) \cdot P_{H_i}(A)}{\sum_{i=1}^n P(H_i) \cdot P_{H_i}(A)}$$

Доказательство. По определению условной вероятности $P_A(B) = \frac{P(AB)}{P(A)}$. В числителе применим теорему умножения, а в знаменателе – формулу полной вероятности.

1.5. Случайные величины.

Наряду с понятием случайного события в теории вероятности используется понятие случайной величины.

Случайной величиной называется переменная величина, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений, причем заранее неизвестно, какое именно.

Примеры:

- число очков, выпавших при броске игральной кости;

- число появлений герба при 10 бросках монеты;
- число выстрелов до первого попадания в цель;
- расстояние от центра мишени до пробойны при попадании.

Случайные величины подразделяются на две группы: *дискретные* и *непрерывные*.

Дискретные случайные величины

Случайная величина называется **дискретной** (ДСВ), если множество $\{x_1, x_2, \dots, x_n, \dots\}$ ее возможных значений конечно или счетно (т.е. если все ее значения можно занумеровать).

Такие из перечисленных выше случайных величин, как количество очков, выпадающих при бросании игрального кубика, число посетителей аптеки в течение дня, количество яблок на дереве являются дискретными случайными величинами.

Наиболее полную информацию о дискретной случайной величине дает **закон распределения** этой величины.

Закон распределения – это соответствие между всеми возможными значениями этой случайной величины и соответствующими им вероятностями.

Закон распределения дискретной случайной величины часто задают в виде двухстрочной таблицы, в первой строке которой перечислены все возможные значения этой величины (в порядке возрастания), а во второй – соответствующие этим значениям вероятности:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Пример 4. Два стрелка делают по одному выстрелу по мишени. Вероятности их попадания при одном выстреле равны соответственно 0,6 и 0,7. Составить ряд распределения случайной величины X – числа попаданий после двух выстрелов.

Решение. Очевидно, что X может принимать три значения: 0, 1 и 2. Найдем их вероятности: Пусть события A_1 и A_2 – попадание по мишени соответственно первого и второго стрелка. Тогда

$$P(X = 0) = P(\overline{A_1} \overline{A_2}) = 0.4 * 0.3 = 0.12$$

$$P(X = 1) = P(\overline{A_1} A_2 + A_1 \overline{A_2}) = 0.4 * 0.7 + 0.6 * 0.3 = 0.46$$

$$P(X = 2) = P(A_1 A_2) = 0.6 * 0.7 = 0.42$$

Следовательно, ряд распределения имеет вид:

x_i	0	1	2
p_i	0.12	0.46	0.42

Для описания определенных особенностей дискретной случайной величины используют ее *основные числовые характеристики*: математическое ожидание, дисперсию и среднее квадратическое отклонение (стандарт).

Математическим ожиданием $M(X)$ (используется также обозначение « μ ») дискретной случайной величины x называется сумма произведений каждого из всех ее возможных значений на соответствующие вероятности:

$$M(x) = \mu = \sum_{i=1}^n x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

Основной смысл математического ожидания дискретной случайной величины состоит в том, что оно представляет собой *среднее значение* данной величины. Другими словами, если произведено некоторое количество испытаний, по результатам которых найдено среднее арифметическое всех наблюдавшихся значений дискретной случайной величины X , то это среднее арифметическое приближенно равно (тем точнее, чем больше количество испытаний) математическому ожиданию данной случайной величины. Приведем некоторые свойства математического ожидания.

1. *Математическое ожидание постоянной величины равно этой постоянной величине:*

$$M(C) = C$$

2. *Математическое ожидание произведения постоянного множителя на дискретную случайную величину равно произведению этого постоянного множителя на математическое ожидание данной случайной величины:*

$$M(kX) = kM(X)$$

3. *Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий этих величин:*

$$M(X+Y) = M(X) + M(Y)$$

4. *Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий:*

$$M(X \cdot Y) = M(X) \cdot M(Y)$$

Отдельные значения дискретной случайной величины группируются около математического ожидания как центра. Для характеристики степени разброса возможных значений дискретной случайной величины относительно ее математического ожидания вводят понятие *дисперсии дискретной случайной величины*:

Дисперсией $D(X)$ (используется также обозначение « σ^2 ») дискретной случайной величины X называется математическое ожидание квадрата отклонения этой величины от ее математического ожидания:

$$D(X) = \sigma^2 = M((X - \mu)^2),$$

На практике дисперсию удобнее вычислить по формуле

$$D(X) = \sigma^2 = M(X^2) - \mu^2,$$

Перечислим основные свойства дисперсии.

1. Дисперсия постоянной величины равна нулю:

$$D(C) = 0$$

2. Дисперсия любой случайной величины есть число неотрицательное:

$$D(X) \geq 0$$

3. Дисперсия произведения постоянного множителя k на дискретную случайную величину равна произведению квадрата этого постоянного множителя на дисперсию данной случайной величины:

$$D(kX) = k^2 \cdot D(X).$$

В вычислительном отношении более удобна не дисперсия, а другая мера рассеивания случайной величины X , которая чаще всего и используется – *среднее квадратическое отклонение (стандартное отклонение или просто стандарт)*.

Средним квадратическим отклонением дискретной случайной величины называется квадратный корень из ее дисперсии:

$$\sigma(x) = \sqrt{D(X)}$$

Удобство стандартного отклонения состоит в том, что оно имеет размерность самой случайной величины X , в то время как дисперсия имеет размерность, представляющую квадрат размерности X .

Непрерывные случайные величины

Случайная величина называется **непрерывной** (НСВ), если множество ее возможных значений целиком заполняет некоторый конечный или бесконечный интервал или системы интервалов на числовой оси.

Например, непрерывными случайными величинами являются: температура больного в фиксированное время суток, масса наугад выбранной таблетки некоторого препарата, рост наугад выбранного студента и др.

Непрерывную случайную величину нельзя задать в виде таблицы ее закона распределения, поскольку невозможно перечислить и выписать в определенной последовательности все ее значения, а также потому, что вероятность любого конкретного значения непрерывной случайной величины равна нулю.

Одним из возможных способов задания непрерывной случайной величины является использование с этой целью соответствующей **функции распределения**.

Функция $F(x)$, равная вероятности того, что случайная величина X в результате испытания примет значение, меньшее x , называется **функцией распределения** данной случайной величины:

$$F(x) = P(X < x),$$

Свойства функции распределения:

1. Функция распределения удовлетворяет неравенству:

$$0 \leq F(x) \leq 1$$

2. Функция распределения является неубывающей функцией, т.е. из $x_2 > x_1$ следует $F(x_2) \geq F(x_1)$.

3. Функция распределения стремится к 0 при неограниченном убывании ее аргумента и стремится к 1 при его неограниченном возрастании.

Плотностью распределения вероятностей (плотностью вероятности) $f(x)$ непрерывной случайной величины X называется производная функции распределения $F(x)$ этой величины:

$$f(x) = F'(x)$$

Под основными числовыми характеристиками непрерывной случайной величины понимают, как и в случае дискретной случайной величины, математическое ожидание, дисперсию и среднее квадратическое отклонение.

Математическое ожидание непрерывной случайной величины:

$$M(X) = \mu = \int_{-\infty}^{+\infty} xf(x)dx$$

Дисперсия непрерывной случайной величины:

$$D(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx$$

Среднее квадратическое отклонение, как и для дискретной случайной величины, определяется формулой:

$$\sigma(x) = \sqrt{D(X)}$$

1.6. Нормальный закон распределения

Из известных видов распределения непрерывных случайных величин наиболее часто используют **нормальное распределение**, которое задается **законом Гаусса**. К нормальному закону распределения при весьма часто встречающихся условиях приближаются другие законы. Так, если мы имеем сумму большого числа независимых величин, подчиненных каким угодно законам распределения, то при некоторых общих условиях она будет приближенно подчиняться нормальному закону.

Непрерывная случайная величина называется **распределенной по нормальному закону (закону Гаусса)**, если ее плотность вероятности имеет вид

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

где μ - математическое ожидание; σ^2 - дисперсия; σ - среднее квадратическое отклонение этой величины.

2. Медицинская статистика.

Одной из отраслей статистики является **медицинская статистика**, которая изучает количественную сторону массовых явлений и процессов в медицине.

Статистика здоровья изучает здоровье общества в целом и его отдельных групп, устанавливает зависимость здоровья от различных факторов социальной среды.

Статистика здравоохранения анализирует данные о медицинских учреждениях, их деятельности, оценивает эффективность различных организационных мероприятий по профилактике и лечению болезней.

Требования, предъявляемые к статистическим данным, можно сформулировать в следующих положениях:

- 1) Достоверность и точность материалов.
- 2) Полнота, понимаемая как охват всех объектов наблюдения за весь исследуемый период, и получение всех сведений по каждому объекту в соответствии с установленной программой.
- 3) Сравнимость и сопоставимость, достигаемая в процессе наблюдения единством программы и номенклатур и в процессе обработки и анализа данных – применением унифицированных методических приемов и показателей.
- 4) Срочность и своевременность получения, обработки и представления статистических материалов.

Изучение состояния здоровья населения, влияние на него некоторых факторов проводится путем специальных статистических исследований.

Статистические исследования помогают определить не только размер, уровень изучаемого явления, но и определяющие его закономерности. Статистические исследования могут быть проведены при изучении рождаемости, смертности, заболеваемости населения. В зависимости от

программы можно изучать уровень и общие закономерности изучаемого явления или влияния на эти явления различных факторов (социальных, гигиенических, биологических).

Предметом статистического исследования может быть также организация медицинской помощи населению, характеристика деятельности лечебно-профилактических учреждений, изучение влияния внешней среды на здоровье человека.

Объектом любого статистического исследования является статистическая совокупность.

Статистическая совокупность – это группа, состоящая из относительно однородных элементов, взятых вместе в известных границах времени и пространства. Примером статистической совокупности может быть: население региона, отдельные его группы, кадры, учреждения здравоохранения и др.

Статистическая совокупность состоит из единиц наблюдения.

Единица наблюдения – каждый первичный элемент статистической совокупности, наделенный признаками сходства. Например : житель города N., родившийся в данном году, заболевший гриппом и т.д.

Признаки сходства служат основанием для объединения единиц наблюдения в совокупность. Объем статистической совокупности составляет общая численность единиц наблюдения.

Учетные признаки – признаки, по которым различают единицы наблюдения в статистической совокупности. По своему характеру учетные признаки подразделяются на атрибутивные (описательные) – выраженные словесно и количественные – выраженные числом.

Различают два вида статистической совокупности: генеральную и выборочную.

Генеральная совокупность – состоит из всех единиц, которые могут быть к ней отнесены с учетом цели исследования.

Пример: рабочие всех предприятий машиностроения города N, прошедшие медосмотр в 2005 году.

Выборочная совокупность - часть генеральной совокупности, отобранная специальным методом.

Пример: рабочие 2-х из 4-х предприятий машиностроения, прошедшие медосмотр в 2005 году.

Выборочная совокупность должна быть репрезентативна по количеству и качеству по отношению к генеральной совокупности.

Репрезентативность – представительность выборочной совокупности по отношению к генеральной совокупности.

Репрезентативность количественная – достаточная численность единиц наблюдения выборочной совокупности.

Репрезентативность качественная – соответствие (однотипность) признаков, характеризующих единицы наблюдения выборочной совокупности по отношению к генеральной. Иными словами, выборочная

совокупность должна быть по качественной характеристике возможно ближе к генеральной совокупности.

Репрезентативность достигается правильно проведенным отбором единиц наблюдения, при котором любая единица всей совокупности в целом имела бы равновеликую возможность попасть в выборочную совокупность.

К выборочному методу обращаются в тех случаях, когда необходимо провести углубленное исследование, соблюдая экономию сил, средств, времени. Выборочный метод при правильном его применении дает достаточно верные результаты, пригодные для их использования в практических и научных целях.

2.1. Этапы статистического исследования

Существует ряд методов отбора единиц для выборочной совокупности, из которых наиболее часто используются следующие способы: случайный, механический, типологический, серийный, парно-сопряженный, комбинированный.

Случайный отбор характерен тем, что все единицы генеральной совокупности имеют равные возможности попасть в выборку (по жребию, по начальной букве фамилии или дню рождения, по таблице случайных чисел).

Механический отбор, когда из всей (генеральной) совокупности берется механически отобранная, например, каждая пятая (20%) или каждая десятая (10%) единица наблюдения.

Типологический отбор (типичная выборка) позволяет производить выбор единиц наблюдения из типичных групп всей генеральной совокупности. Для этого сначала внутри генеральной совокупности все единицы группируются по какому-нибудь признаку в типичные группы (например, по возрасту). Из каждой такой группы производят отбор (случайным или механическим способом) необходимого числа единиц таким образом, чтобы соотношение размеров возрастных групп в выборочной совокупности сохранялось таким же, как и в генеральной совокупности.

Серийный отбор предусматривает выбор из генеральной совокупности не отдельных единиц, а выбор серий. Для этой цели 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 возможность быть отобранной в выборочную совокупность. В каждой отобранной серии обследуются все составляющие ее единицы наблюдения.

Парно-сопряженный отбор или метод уравнивания при формировании выборочной совокупности предусматривает максимальное сходство единиц наблюдения в обеих группах кроме изучаемого фактора. Для этого каждой единице наблюдения в исследуемой группе подбирают копию, то есть пару, в контрольной группе. Такой способ позволяет сформировать группы равные по численности и однородные по одному или

нескольким признакам и более четко определить влияние исследуемого фактора.

Организация статистического исследования

Статистическое исследование состоит из четырёх этапов:

1. Составление программы и плана.
2. Сбор материала.
3. Разработка данных.
4. Анализ, выводы, предложения, внедрение в практику.

I этап – составление программы и плана исследования. *План исследования* представляет собой организационные элементы работы, т.е. кто, где и когда проводит исследование. В первую очередь необходимо определить объект исследования. Под объектом исследования понимают статистическую совокупность, состоящую из отдельных предметов или явлений – единиц наблюдения. Объектом исследования могут быть население региона, отдельные его группы, кадры, учреждения здравоохранения и др. При составлении плана необходимо определить способы формирования совокупности.

Программа статистического исследования представляет собой рабочую гипотезу с конкретными целями и задачами исследования. Она состоит из двух компонентов:

1. Программа сбора материала.
2. Программа разработки.

Программа сбора материала – представляет выбор или разработку учётного документа с необходимым набором признаков, подлежащих регистрации. В соответствии с целями и задачами исследования, в качестве учётного документа могут использоваться как типовые учётно-статистические документы (карта выбывшего из стационара, статистический талон и другие), так и самостоятельно разработанные учётные документы, включающие в себя весь перечень интересующих вопросов (анкеты, карты и т.д.)

Программа разработки – это составление макетов таблиц с учётом структурных признаков изучаемого явления. Таблицы строятся по определённым принципам и подразделяются на простые, групповые и комбинационные. Каждая таблица должна иметь чёткое и краткое название, определяющее её содержание. В таблице различают подлежащее (то, о чём говорится) и сказуемое (то, что разъясняет подлежащее). Статистическое подлежащее – основной признак изучаемого явления, как правило, располагается по горизонтальным строкам таблицы. Статистическое сказуемое – признак, характеризующий подлежащее, располагается в вертикальных графах таблицы. Оформление таблицы заканчивается итогами по графам и строкам.

2.2. Виды статистических таблиц.

Простой называется таблица, в которой представлена только количественная характеристика подлежащего (пример табл.1).

Таблица 1.

Число коек в больницах Н-ской области на 01.01.2015г.

Наименование больниц	Число коек
Больница №1	
Больница №2	
Больница №3	
Всего	

Простые таблицы легко составляются, но их сведения мало пригодны для анализа, поэтому их используют, в основном, для статистической отчётности.

Групповой называется таблица, в которой представлена связь подлежащего только с одним из признаков сказуемого (пример табл. 2).

Таблица 2.

Распределение больных по полу и возрасту, лечившихся в больницах Н-ской области в 2015 г.

Наименование больниц	Пол		О/П	Возраст			Всего
	Ж	М		0-14 лет	15-49 лет	50 лет и старше	
Больница №1							
Больница №2							
Больница №3							
Всего							

Групповая таблица может содержать неограниченное число признаков в сказуемом, но с подлежащим они сочетаются только попарно:

- больница и лечившиеся по полу,
- больница и лечившиеся по возрасту.

Комбинационной называется таблица, данные которой характеризуют связь подлежащего с комбинацией признаков сказуемого (пример табл. 3).

Таблица 3.

Распределение больных, лечившихся в Н-ской больнице, по нозологическим формам, полу и возрасту за 2005-2015гг.

Нозологические формы	Возраст в годах												Всего		
	до 30			39-40			40-49			50 и старше					
	М	Ж	О/П	М	Ж	О/П	М	Ж	О/П	М	Ж	О/П	М	Ж	О/П
Брюшной тиф															
Дизентерия															
Скарлатина															
Дифтерия															
Корь															
Итого															

Комбинационными таблицами пользуются для проведения углубленного изучения связей между отдельными признаками одного явления, или между несколькими однородными явлениями, отличающимися только по одному признаку.

II этап – сбор материала. На этом этапе осуществляется статистическое наблюдение и сбор материала. Статистическое наблюдение включает в себя регистрацию и заполнение учётных документов в соответствии с программой и планом исследования.

III этап – разработка материала. Разработка собранного материала состоит из нескольких этапов:

1. Контроль и шифровка.
2. Группировка.
3. Сводка в таблицы.
4. Вычисление статистических показателей.
5. Графические изображения.

IV этап - анализ, выводы, предложения, внедрение в практику. На завершающем этапе исследования осуществляется анализ полученных результатов, сопоставление их с критериями и данными аналогичных исследований, обобщение материала. На основании проделанной работы формулируются выводы и предложения по внедрению в практику.

**3. Цель деятельности студентов на занятии:
Студент должен знать:**

1. Что изучает Теория вероятности.
2. Что основные определения теории вероятности.
3. Основные формулы теории вероятности.
4. Что изучает Медицинская статистика.
5. Этапы статистического исследования.

Студент должен уметь:

1. Вычислять вероятность случайного события.
2. Находить Сумму и произведение вероятностей.
3. Находить Математическое ожидание, Дисперсию и Среднее квадратическое отклонение дискретной случайной величины.
4. Строить статистические таблицы

4. Содержание обучения:

1. Элементы теории вероятности.
 - 1.1. Основные понятия теории вероятности.
 - 1.2. Классическое определение вероятности.
 - 1.3. Основные теоремы.
 - 1.4. Формула полной вероятности.
 - 1.5. Случайные величины.
 - 1.6. Нормальный закон распределения.
2. Медицинская Статистика.
 - 2.1. Этапы статистического исследования.
 - 2.2. Виды статистических таблиц.

5. Перечень вопросов для проверки уровня знаний:

1. Теория вероятности. Предмет и методы.
2. Математическая статистика. Предмет и методы.
3. Формула Вероятности случайного события.
4. Формула полной вероятности.

6. Перечень вопросов для проверки конечного уровня знаний:

1. Дайте определение теории вероятности.
2. Что такое испытание, случайное событие. Свойства случайного события.
3. Классическое определение вероятности.
4. Сформулируйте теоремы сложения и умножения.
5. Формула полной вероятности.
6. Случайная величина, виды случайных величин.
7. Дайте определения математического ожидания, дисперсии, среднего квадратического отклонение дискретной случайной величины.

8. Медицинская статистика, предмет и метод.
9. Дайте определения генеральной и выборочной совокупностей.
10. Методы отбора единиц для выборочной совокупности.
11. Этапы статистического исследования.
12. Виды статистических таблиц.

7. Практическая часть

ВАРИАНТЫ РЕШЕНИЯ ТИПОВЫХ ЗАДАЧ

План и программа темы: "Изучение мнения пациентов о культуре медицинского обслуживания в амбулатории города Н."

Цель исследования: оценить качество амбулаторного обслуживания на основании мнения пациентов для разработки практических рекомендаций по улучшению организации медицинской помощи.

Задачи исследования:

1. Оценить удовлетворённость пациентов работой врача.
2. Выяснить мнение о его достоинствах и недостатках при оказании медицинской помощи.
3. Оценить работу регистратуры.
4. Выяснить трудности, с которыми встречаются пациенты при получении медицинской помощи.

План исследования:

1. Объект исследования – пациенты амбулатории города Н.
2. Тип исследования по времени – единовременное наблюдение (на момент опроса).
3. Объём исследования – выборочное наблюдение.

Программа наблюдения:

1. Единица наблюдения – пациент, обратившийся в амбулаторию на момент обследования.
2. Учётный документ – анкета по изучению амбулаторного обследования населения.
3. Учётные признаки – 23 вопроса анкеты.

Программа разработки – составление макетов статистических таблиц.

Примеры каждого вида таблиц

1. Простая

Распределение ответов пациентов амбулатории на вопрос: "Как Вы чаще всего поступаете, когда обнаруживаете у себя признаки болезни, если эта болезнь, по Вашему мнению, не опасна для Вас?"

Вариант ответа	Кол-во ответов
1. Ничего не предпринимаю, жду когда болезнь пройдет сама. 2. Использую в лечении советы родственников, лечусь домашними средствами. 3. Спрашиваю советы у других людей. 4. Сразу иду к врачу. 5. Иду к врачу не сразу, а спустя некоторое время. 6. Нет ответа	
Всего ответивших	

2. Групповая.

Распределение ответов пациентов на вопрос: "Удовлетворяет ли Вас работа врача?" в зависимости от возраста и пола.

Вариант ответа	Возраст в годах			Всех возрастов	Пол		О/П
	До 20	20-49	50 и старше		М	Ж	
1. Да 2. Нет 3. Не всегда							
Всего ответов							

3. Комбинационная.

Распределение ответов пациентов на вопрос: "Удовлетворяет ли Вас работа регистратуры?" по возрастно-половым группам.

Вариант ответа	Возраст в годах												Всего			
	до 20			20-29			30-49			50 и ст.			М	Ж	О/П	
	М	Ж	О/П	М	Ж	О/П	М	Ж	О/П	М	Ж	О/П				
1. Да 2. Нет 3. Не всегда																
Всего ответов																

Задача 1. Составьте программу и план к теме исследования «Характеристика детского травматизма в городе Н.»

Цель исследования: профилактика детского травматизма.

Задачи исследования: 1. Изучить динамику детского травматизма.

2. Определить структуру травм у детей по локализации и степени тяжести.

Задача 2. Составьте программу и план к теме исследования «Характеристика заболеваемости рабочих завода Н.»

Цель исследования: разработать профилактические мероприятия по снижению уровня заболеваемости на заводе Н.

Задачи исследования: 1. Изучить динамику заболеваемости на заводе за 5 лет.

2. Изучить заболеваемость в основных цехах.

3. Определить структуру заболеваемости.

Задача 3. Абонент забыл последнюю цифру номера телефона и поэтому набирает её наугад. Определить вероятность того, что ему придётся звонить не более чем в 3 места.

Задача 4. Абонент забыл последние 2 цифры телефонного номера, но помнит, что они различны и образуют двузначное число, меньшее 30. С учетом этого он набирает наугад 2 цифры. Найти вероятность того, что это будут нужные цифры.

8. Самостоятельная работа студентов:

По учебнику изучите основные понятия теории вероятности и медицинской статистики.

9. Хронокарта учебного занятия:

1. Организационный момент – 5 мин.
2. Текущий контроль знаний – 30 мин.
3. Разбор темы – 20 мин.
4. Практическая работа – 30 мин.
5. Подведение итогов занятия – 10 мин.

10. Перечень учебной литературы к занятию:

1. Кобринский Б.А., Зарубина Т.В. «Медицинская информатика», М., Издательский дом «Академия», 2009.
2. Жижин К.С. «Медицинская статистика», Высшее образование, 2007.

Тема 2: «Методы математической статистики. «Основная структура пакета Statistica. Формирование таблиц. Ввод данных. Сохранение файлов данных и открытие существующих»

1. Научно-методическое обоснование темы:

Эффективная работа специалиста-медика на сегодняшний день подразумевает знание не только медицинских дисциплин, но также и хорошее владение навыками автоматизированной обработки медицинских данных.

Практически всем специалистам в области медицины необходим навык по осуществлению статистической обработки данных, которая должна производиться с применением современных программно-технических средств. Почти нет ни одной врачебной должности в больнице (в клинике), в поликлинике, занимая которую, врачу не пришлось бы выполнять какие-нибудь статистические работы. Поэтому, представляется важным, чтобы лечащие врачи знали, как взяться за дело, умели собирать и обрабатывать верные цифры, годные для сравнения и сопоставления. Не только врачи-клиницисты, но и врачи-экспериментаторы, врачи гигиенисты и организаторы здравоохранения не в меньшей мере нуждаются в умении правильно применять статистические методы исследования и правильно толковать и использовать результаты этих исследований.

Нейронные сети - исключительно мощный метод моделирования, позволяющий воспроизводить чрезвычайно сложные зависимости. В частности, нейронные сети нелинейны по своей природе (смысл этого понятия подробно разъясняется далее в этой главе). На протяжении многих лет линейное моделирование было основным методом моделирования в большинстве областей, поскольку для него хорошо разработаны процедуры оптимизации. В задачах, где линейная аппроксимация неудовлетворительна (а таких достаточно много), линейные модели работают плохо. Кроме того, нейронные сети справляются с "проклятием размерности", которое не позволяет моделировать линейные зависимости в случае большого числа переменных.

Простота в использовании. Нейронные сети учатся на примерах. Пользователь нейронной сети подбирает представительные данные, а затем запускает алгоритм обучения, который автоматически воспринимает структуру данных. При этом от пользователя, конечно, требуется какой-то набор эвристических знаний о том, как следует отбирать и подготавливать данные, выбирать нужную архитектуру сети и интерпретировать результаты, однако уровень знаний, необходимый для успешного применения нейронных сетей, гораздо скромнее, чем, например, при использовании традиционных методов статистики. Нейронные сети привлекательны с интуитивной точки зрения, ибо они основаны на примитивной биологической модели нервных систем. В будущем развитие таких нейро-биологических моделей может привести к созданию действительно мыслящих компьютеров. Между тем уже

"простые" нейронные сети, которые строит система ST Neural Networks , являются мощным оружием в арсенале специалиста по прикладной статистике.

Материал данной методической разработки предназначен для изучения основ статистического анализа студентами медицинских вузов в рамках дисциплины «Медицинская информатика» с использованием пакета Statistica 6.0.

2. Краткая теория:

1. Основные понятия и определения

В самых различных областях (медицина, налоговая служба, экономика, геология и т.д.) объекты описываются набором величин, то есть являются многомерным. А обработка таких данных очень трудоемкая работа. Но в данном математическом пакете процесс анализа данных превращается в увлекательное исследование с использованием компьютерных технологий. Вы открываете диалоговые окна, создаете новые файлы с данными, запускаете вычислительные процедуры, строите графики, просматриваете результаты обработки, выводите их на печать. Если раньше каждый шаг исследований, начиная от представления данных, проверки, группировки, сортировки, графической интерпретации, подготовка программ обработки до просмотра результатов, был достаточно трудной задачей. Теперь достаточно пару щелчков мыши, чтобы огромные объемы данных чрезвычайно быстро преобразовались, обработались и появились на экране в виде графиков, диаграмм и таблиц. Тем более рабочие окна русифицированной версии пакета STATISTICA 6.0 6.0 похожи на рабочие окна Microsoft Office Excel, где можно сохранять Рабочие книги (Workbook.stw) с результатами проделанных расчетов и анализов, чтобы в случае необходимости вернуться к любому этапу проделанного анализа.

Начнем наше знакомство с достаточно интересным математическим пакетом STATISTICA 6.0 6.0, запустив его.

2. Создание электронной таблицы.

СПОСОБ 1. Математический пакет STATISTICA 6.0 6.0 автоматически открывает пустую электронную таблицу, которая появится на экране. В заголовке окна электронной таблицы автоматически отображается имя файла и его размер. Размер таблицы по умолчанию принят 10 на 10 (10 переменных с именами VAR1, VAR2,...VAR10 и 10 пронумерованных случаев).

Произведем настройку размеров таблицы (создадим столько переменных и случаев (регистров), сколько необходимо). Нажмите кнопку Variables (Переменные) на панели инструментов и выберите команду Удалить (Delete). В диалоговом окне Delete Variables (Удаление переменных) укажите диапазон удаляемых переменных и нажмите кнопку.

Аналогичные действия проводятся, чтобы добавить число переменных. Выберите команду Добавить (Add). В диалоговом окне Add Variables (Добавить переменные) укажите: количество переменных, которое необходимо добавить (How many), и после какой переменной (After) необходимо добавить новую переменную; также можно сразу ввести название переменной (Name).

Для удаления лишних случаев (регистров) воспользуйтесь кнопкой Cases (Случаи) и командой Delete (Удалить). В появившемся окне Delete Cases (Удалить случаи).

Задайте диапазон удаляемых случаев, нажмите кнопку. Для увеличения числа случаев (регистров) воспользуйтесь командой Add (Добавить). В появившемся окне Add Cases (Добавить случаи).

Задайте необходимое число случаев (How many) и после какого случая вставить (Insert after case). Нажмите кнопку. Теперь электронная таблица содержит столько переменных и случаев, сколько необходимо.

СПОСОБ 2. В основном рабочем окне пакета STATISTICA 6.0 6.0 в строке меню из пункта Файл выберите команду Новый (New). В появившемся диалоговом окне Create New Document (Создание нового документа) необходимо задать Число переменных (Число столбцов в таблице) и Число регистров (Число случаев (строк) в таблице). Нажмите кнопку в правом нижнем углу окна. Из пункта Файл выберите команду Сохранить, введите название данной таблице и сохраните.

3. Подготовка таблицы к вводу данных, заголовок таблицы и имена переменных.

Дважды щелкните мышью на белом поле в таблице под словами: Data: <Название таблицы>. В этой ячейки можно задать название таблицы. Таблица почти готова к вводу данных, однако нужно ввести имена переменных, которые отражают смысл записей, и специфицировать их. Дважды щёлкните на имени переменной VAR1 в электронной таблице. На экране появится окно спецификации переменной Variable 1. В поле Name (Имя) напишите название переменной. Длина имени не должна превышать 8 символов. Также можно указать формат переменной (Display format). Нажмите кнопку. То же сделайте для остальных переменных. Теперь таблица готова к тому, чтобы ввести в неё данные.

4. Ввод данных в электронную таблицу.

Если данных не много, то можно вводить их с клавиатуры, но большие таблицы данных лучше копировать с Microsoft Excel. Замечание: Не стоит забывать, что в пакете STATISTICA 6.0 6.0 можно производить непосредственные расчеты в ячейках таблицы, также как в Microsoft Excel.

5. Сохранение файла данных.

Для сохранения созданного файла нажмите мышью на панели управления кнопку Сохранить либо наберите на клавиатуре CTRL+S. Созданный файл сохранится и всегда будет доступен.

Пакет имеет идеальную структуру, модулей у него большое количество, а размещение на платформе windows обеспечивает связь пакета практически со всеми windows- приложениями.

Несомненным достоинством пакета относят большой выбор методов обработки данных: от достаточно простых приемов описательной статистики до сложных методов многомерного анализа и планирования эксперимента. И он имеет широкие графические возможности и хорошую методическую базу на русском языке.

Стартовое окно пакета служит для выбора одного из статистических модулей и называется Statistica 6.0 Module swicher (переключатель модулей пакета «Статистика»).

Обязательный набор модулей:

- Основные статистики и таблицы (Basic Statistic and Tables);
- Непараметрические критерии и подгонка распределений (Nonparametric Statistic&Disuribution Fitting);
- Одно- и многофакторный дисперсионный анализ (ANOVA/MANOVA);
- Линейная регрессия (Linear Regression);
- Канонический дискриминантный анализ (Canonical Analysis);
- Кластерный анализ (Cluster Analysis);
- Управление данными (Data Management);
- Дискриминантный функциональный анализ (Discriminant Function Analysis);
- Факторный анализ (Factor Analysis).

Наиболее часто используемые кнопки при начале анализа данных:

- Переключиться в...(Switch To...);
- Список для выбора (Customize List);
- Закончить выбор и переключиться в (End & Switch To);
- Отказаться (Cancel).

Стандартная панель:

- Файл;
- Правка;
- Анализ;
- Параметры;
- Окно;
- Помощь.

Особенностью данного пакета является широкий спектр двух- и трехмерных графических форм.

Процедура запуска пакета:

Из главного меню создаем документ, кстати он может быть различного типа: файл исходных данных и файл отчета. После запуска открывается стартовое окно, выбирается одна из пяти возможностей, из которых основными являются две: ввести данные (Type in data) и открыть существующий файл (Open an existing file). Начнем с модуля «Основные статистики и таблицы». Обычно при открытии пакета «выскакивает» уже готовая таблица с данными из какого-то примера, находящихся в изобилии каталоге(Examples).

Чтобы создать свой личный файл, нажимаем кнопку File(Файл) и в меню – New Data(Файл новых данных) присвоим ему имя ИМЯ (автоматически к нему будет присоединено расширение «sta...»).

Сначала мы вводим данные в отдельные столбцы таблицы. Размер матрицы данных в нашем пакете изначально ограничен десятью случаями(Cases) и 10-ю переменными(Variables).

Для увеличения количества случаев:

- Нажмем кнопку случаи(Cases)
- Выберем добавить (Add), в поле Количество добавляемых случаев(Number of Cases to Add) введем число нужное, а в поле Вставить после случая(Insert after Cases) – соответствующее число.

Таблица данных приобретет другие размеры. Также варьируется и количество переменных. Для удаления используется команда Удалить(Delete) и в полях От переменной(From variable) До переменной(To variable) введем соответственно значения.

Переменные в пакете Statistica 6.0 for Windows изначально имеют VAR1, VAR2...Эти обозначения можно менять по своему усмотрению, щелкнув по названию переменной. Откроется окно, в котором вы сможете произвести перенастройку. **Напоминаем: название переменной не может содержать более 8 символов**, и лучше- английское написание.

Гораздо лучше уже после обработки данных перед выводом на печать русифицировать название таблиц, графиков. Так как наши данные имеют числовой тип и представляют двузначные числа в поле Width(Общая ширина), надо ввести значение 2. Столбец может автоматически сузиться и не будет соразмерен названию переменной. Для этого полю переменной надо задать адекватную ширину. В поле Decimal(Количество десятичных знаков после запятой) для обеих переменных зададим 0. В поле Data (Дата) зададим формат вводимых значений.

В поле Long Name(Полное-длинное-имя) можно ввести Label(Метку) переменной, в этом варианте ее «длинное» название будет вводиться при попытке вывода результатов. Сюда можно вводить и формулы.

В пакете можно оперировать метками отдельных значений переменной, задавать текстовые названия для числовых результатов измерений. Для этой цели в диалоговом окне нажмем кнопку Text Values(текстовые значения) и введем текстовые названия, числовые значения и их метки.

Текстовые значения не должны выходить за пределы восьми символов, а длина меток числовых значений-40 символов. Новая строка вводится клавишей Insert. В таблице показываются сами значения, а чтобы перейти к их текстовым названиям, нажимаем на панели инструментов кнопку Text Value(Текстовые названия). Автоматически появляются не числовые значения, а их названия. Повторное нажатие кнопки возвращает нас к числовым значениям. Сохраним введенные данные в файле. Это можно сделать несколькими способами:

Активизировать на стандартной панели значок, на котором изображена дискета.

Или рядом, слева активизировать значок Файл(File)- Сохранить(Save).

В первом случае при первом сохранении открывается диалоговое окно для определения маршрута размещения файла и определения его имени. После этой процедуры имя файла появляется на площади окна.

Внесение дальнейших изменений в файл системой отслеживается автоматически: новый вариант файла будет помещен под тем же именем и в той же папке, по тому же маршруту размещения.

6. Нейронные сети.

Идея нейронных сетей в статистике разрабатывается не первое десятилетие. Как и любой другой статистический продукт, нейронные сети имеют свои ограничения. Их нецелесообразно использовать там, где достаточно односложные ответы можно получить, применяя дисперсионный, регрессионный, кластерный или факторный анализы, не говоря уже о первичной обработке данных – описательных приемах. Но в ситуациях, когда надо разгрести «завал» данных, получение осмысленного результата из которого проблематично, - тут-то нейронные сети и могут облегчить участь исследователя.

Основная идея нейронных сетей в том, что они механически повторяют структуру действительного нейрона (нервной клетки) мозга человека: эта клеточная структура имеет несколько отростков – входные - дендриты (их может быть несколько) и один выходной – аксон. Нейрон начинает передавать информацию через синапсы (узлы связи) другим нейронам только в случае, когда возбужден, или иначе, переполнен, информацией. Дозируя информацию, можно регулировать активность нейрона.

И на этом фоне удивительной чертой нейронных сетей является их способность к обучению, чего начисто лишены основные методы статистического анализа. В данном варианте реализуется действительная структура человеческого мозга: во-первых, принимать решение, опираясь на функцию памяти о прошлом опыте; во-вторых, действовать по ассоциации, используя обрывочные сведения о предмете анализа.

Систем обучения нейронных сетей достаточно много: У. Маккалоха, Д. Хебба, М. Минского, Дж. Хопфилда...

1. Откроем пакет, войдем в модуль Нейронные сети.
2. Через (Файл новый) выберем команду (Сеть).
3. На экране появится (Создать сеть)- (Create Network).
4. В поле (Тип) выбираем (Многослойный персептрон) и делаем установки: Вход=1, Выход=1.
5. Зададим число слоев сети равное трем, т.е. выбираем трехслойный персептрон.
6. Временное окно(Steps) пометим «12», это будет отвечать ежемесячной регистрации заболеваемости дизентерией с присущей сезонной составляющей.
7. Горизонт (Lookahead) пометим «1».
8. На экране монитора справа появится схема персептрона. Поскольку модель определена, необходимо сеть обучить. Для этого надо задать 66 обучающих (Training) и контрольных (Verification) наблюдений.
9. Далее перемешиваем данные (Shuffle), поскольку мы анализируем временный ряд, а порядок данных очень важен в соответствии с временными промежутками, то исключается использование кнопкой Сгруппировать (Group Sets).
10. Откроем командой Запуск окно Проекция временного ряда(Time Series Projection). Ряд можно построить целиком или с какого-то интересующего нас момента. Графически кривая идет достаточно круто вверх, число прогнозируемых случаев не имеет тенденции к снижению, колеблясь в пределах 1000,0 – 1020,0 на 10000 населения.
11. Проверим качество работы обученной сети, открыв окно Статистики регрессии, включим Запуск(Run).
12. Для того, чтобы получить прогноз на один шаг вперед в меню Запуск(Run) выберем команду (Single Case), откроется соответствующее диалоговое окно, где надо ввести номер наблюдения, для которого строится прогноз, ждем вверху справа кнопку (Run), в строке Output получим искомый результат прогноза. Введя для примера 13, получим прогнозируемый уровень 1011,231. Оценивая значимость входов, определим, что из четырех выбранных факторов ведущими (по величине объясненной дисперсии) является детская заболеваемость: 0,567 и оперативная эффективность: 0,451. Точность прогноза на среднем уровне, что объясняется малым объемом обучающей выборки, и тем, что велика вероятность, что процесс по годам нестационарен.

3. Цель деятельности студентов на занятии:

Студент должен знать:

1. Почему специалисту-медику необходимы навыки проведения статистических исследований в своей работе.
2. Основную структуру пакета Statistica 6.0 6.0.
3. Процесс ввода данных в таблицу Spreadsheet, сохранение файлов и открытие существующих.

Студент должен уметь:

1. Владеть базовыми средствами для создания, редактирования, форматирования таблиц в пакете Statistica 6.0 6.0. В необходимой мере владеть навыком по осуществлению статистической обработки данных с применением современных программно-технических средств, в частности, с помощью возможностей данного пакета.
2. Эффективно использовать пакет Statistica 6.0 6.0. для более глубокого статистического исследования, с показателями корреляции, регрессии и описательной статистики.
3. Проводить анализ временных рядов с использованием пакета Statistica 6.0.(нейронные сети).

4. Содержание обучения:

1. Основные понятия и определения
2. Создание электронной таблицы.
3. Подготовка таблицы к вводу данных, заголовков таблицы и имена переменных.
4. Ввод данных в электронную таблицу.
5. Сохранение файла данных.
6. Нейронные сети.

5. Перечень вопросов для проверки исходного уровня знаний:

1. Что такое статистическая гипотеза?
2. Роль компьютера в медико-биологической статистике.
3. Методы проверки статистических гипотез.
4. Какую структуру имеет пакет Statistica 6.0?
5. Особенности пакета Statistica 6.0.
6. Как вводятся данные в таблицу пакета Statistica 6.0?
7. Как происходит статобработка данных?

6. Перечень вопросов для проверки конечного уровня знаний:

1. Как создать свой личный файл в пакете Statistica 6.0?
2. Для чего служит стартовое окно пакета?

3. Какую структуру имеет пакет Statistica 6.0?
4. Особенности пакета Statistica 6.0.
5. Как вводятся данные в таблицу пакета Statistica 6.0? Понятие Нейронные сети.
6. Понятие персептрона.
7. Построение временного ряда.

7. Практическая часть:

Задание №1. Получены исходные данные числа положительных находок кишечной палочки в смывах с инвентаря и посуды столовых двух полевых станов «А» «В»:

№	A	B	№	A	B	№	A	B
1	24	35	3	35	24	5	33	22
2	43	34	4	19	20	6	26	28

Введем эти исходные данные в матрицу данных пакета в отдельные столбцы. Переменные в этом случае получают ремарку типа var1... var3 и т.д.

Мы можем изменять названия по ходу обработки данных.

Для задания характеристик переменной надо, подведя курсор к ее названию, дважды щелкнуть левой кнопкой мыши. Открыть диалоговое окно и выполнить следующие действия:

1. Введем имя переменной (оно не должно содержать более 8 символов!) и лучше присваивать переменным англоязычные имена: Stan1, Stan2.

2. Зададим тип данных, кнопкой Type (Тип данных):

- Numeric (числовой) – для него важно иметь Width (Общая длина), или количество десятичных знаков после запятой;
- String (Текстовый) – для него задается только общая длина;
- Date (Тип даты) – с конкретным форматом результата.

Так как наши данные числовые (двузначные числа), мы задаем для них числовой тип с длиной «два» и «нулевым» количеством знаков после запятой.

3. Вводим Labels (Метки) – они отражают комментарии, поясняющие смысл переменной. Для переменной Stan1 введем метку «Полевой стан 1», а для переменной Stan2 – «Полевой стан 2». Метки будут «всплывать» как подсказки, стоит только подвести курсор к названию переменной и секунду подождать.

Теперь вводим данные в таблицу с клавиатуры, а затем сохраняем файл.

Задание №2. Дан динамический ряд числа заболеваний дизентерией за год. Предварительно была сформирована помесечная таблица

анализируемых реальных данных, она была сохранена на диске в конкретной папке. И выделены из 11 входящих четыре основные: 1) детская заболеваемость дизентерией, 2) число порывов на водоводной сети, 3) качество профилактической работы персонала в очагах, 4) оперативная эффективность. Необходимо получить прогноз динамики заболеваемости и определить, какой из четырех отобранных факторов является ведущим.

месяц	Заболеваемость детей дизентерией на 1000 чел.	Число порывов водоводной сети, абс.	Качество проф. работы в очагах, баллы	Оперативная эффективность, %	Число случаев дизентерии(в расчете на 10000 чел.)
1.	20,1	31	1	60	300,1
2.	31,1	22	2	67	423,1
3.	41,1	20	1	67	321,4
4.	43,1	19	2	68	444,4
5.	50,1	13	3	87	600,9
6.	53,1	12	1	87	789,9
7.	55,1	14	1	87	988,8
8.	60,1	21	2	89	1012,1
9.	60,1	22	3	89	1112,4
10.	66,6	23	3	90	1123,3
11.	67,1	30	3	91	1109,8
12.	67,8	18	3	88	899,2

8. Хронокарта учебного занятия:

1. Организационный момент – 5 мин.
2. Текущий контроль знаний – 30 мин.
3. Разбор темы – 20 мин.
4. Практическая работа – 30 мин.
5. Подведение итогов занятия – 10 мин.

9. Самостоятельная работа студентов.

Изучить основные понятия математической статистики. Овладеть основными принципами работы в программе Statistica 6.0 6.0. Изучить использование многомерных статистических методов в статистической обработке медицинских данных.

10. Перечень учебной литературы к занятию:

1. Кобринский Б.А., Зарубина Т.В. «Медицинская информатика», М., Издательский дом «Академия», 2009.
2. Жижин К.С. «Медицинская статистика», Высшее образование, 2007.

Тема 3: «Методы математической статистики. Подготовка данных к статистическому анализу. Элементы организации медико-статистического исследования. Статистическая совокупность. Статистические величины. Вычисление статистических величин»

1. Научно-методическое обоснование темы:

Изучая массовые общественные явления, статистика в своих выводах опирается на числовые данные, полученные в конкретных условиях места и времени. Результаты статистического наблюдения регистрируются прежде всего в форме первичных *абсолютных величин*. Абсолютная величина отражает уровень развития явления.

Производная величина – показатель, получаемый в результате преобразования абсолютной величины на основе сопоставления ее с другой абсолютной величиной. Она выражается отношением или разностью абсолютных величин. Основными видами производных величин, применяемых в биомедицинской статистике, являются относительные величины (статистические коэффициенты) и средние величины.

Абсолютные величины характеризуют, например, численность населения, число рождений, единичные случаи некоторых инфекционных заболеваний, их хронологические колебания. Они необходимы для организационно-плановых построений в здравоохранении (например, планирование необходимого количества коек), а также для расчета производных величин.

Сама по себе абсолютная величина не дает полного представления об изучаемом явлении, не показывает его структуру, соотношение между отдельными частями, развитие во времени. В ней не выявлены соотношения с другими абсолютными показателями. Эти функции выполняют определенные на основе абсолютных величин относительные показатели.

Средняя величина – это обобщающий показатель, характеризующий типический уровень явления. Он выражает величину признака, отнесенную к единице совокупности.

Средняя всегда обобщает количественную вариацию признака, т.е. в средних величинах погашаются индивидуальные различия единиц совокупности, обусловленные случайными обстоятельствами. В отличие от средней абсолютная величина, характеризующая уровень признака отдельной единицы совокупности, не позволяет сравнивать значения признака у единиц, относящихся к разным совокупностям. Таким образом, возникает необходимость расчета средних величин как обобщающей характеристики совокупности.

2. Краткая теория:

1. Относительные величины (статистические коэффициенты)

В подавляющем большинстве случаев, ряды абсолютных чисел не пригодны для сравнения, выявления связей и закономерностей, качественных

особенностей изучаемых процессов. Поэтому вычисляют относительные величины, виды, которых зависят от того, что сопоставляется:

- явление со средой, из которой оно происходит;
- составные элементы одного и того же явления;
- независимые явления, сравниваемые между собой.

Различают следующие виды относительных величин:

- Интенсивные коэффициенты (относительные величины частоты).
- Экстенсивные коэффициенты (относительные величины распределения или структуры).
- Коэффициенты (относительные величины) соотношения.
- Коэффициенты (относительные величины) наглядности.

1.1. Интенсивные коэффициенты.

Интенсивные коэффициенты – характеризуют силу, частоту (степень интенсивности, уровень) распространения явления в среде, в которой оно происходит и с которой оно непосредственно связано.

$$\text{Интенсивный показатель} = \frac{\text{Явление}}{\text{Среда}} \cdot 100(1000; 10000 \dots \text{ и т. д.})$$

Среда, в этом случае, есть основная статистическая совокупность, в которой происходят изучаемые процессы. В демографической и медицинской статистике в качестве среды чаще всего рассматривается население.

Выбор числового основания (100;1000;10000... и т.д.) зависит от распространенности явления - чем реже встречается изучаемое явление, тем большее основание выбирается, чтобы не было коэффициентов меньше единицы, которыми неудобно пользоваться. Например, на 1000 рассчитываются основные демографические показатели, первичная заболеваемость; на 100.000 - инфекционная заболеваемость, уровень заболеваемости туберкулезом, нервно - психической патологией и др.

Примерами интенсивных коэффициентов могут служить коэффициенты рождаемости, смертности, заболеваемости, инвалидности. Для детального анализа явления рассчитываются специальные (групповые) показатели (по полу, возрасту и т.д.). Например, вычисление повозрастного показателя (ПП):

$$\text{ПП} = \frac{\text{Численность явления в данном возрасте}}{\text{Численность среды в данном возрасте}} \cdot 100; 1000 \dots \text{ и т. д.}$$

Вычисление интенсивных показателей производится следующим образом. Например: население города Н. в 2005 г. составило 1.318.600 человек. В течение года умерло 22.944 человек. Для вычисления коэффициента смертности необходимо составить и решить следующую пропорцию:

$$\frac{1,318,600 - 22,944}{1000 - x}$$

$$x = \frac{22,944 * 1000}{1,318,600} = 17,4\%$$

При вычислении интенсивных коэффициентов мы всегда имеем дело с **двумя самостоятельными, качественно различными** совокупностями, одна из которых характеризует среду, а вторая - явление (например, число больных и число умерших); нельзя считать, что больные «распределились на выздоровевших и умерших», умершие - это новое (в данном случае необратимое) явление, самостоятельная совокупность.

Примеры применения интенсивных коэффициентов:

- определение уровня, частоты, распространенности того или иного явления;
- сравнение ряда различных совокупностей по степени частоты того или иного явления (например, сравнение уровней рождаемости в разных странах, сравнение уровней смертности в разных возрастных группах);
- выявление динамики изменений частоты явления в наблюдаемой совокупности (например, изменение распространенности инфекционных заболеваний населения страны за несколько лет).

1.2. Коэффициенты соотношения

Коэффициенты соотношения - характеризуют численное соотношение двух, не связанных непосредственно между собой, независимых совокупностей, сопоставляемых только логически. **Техника вычисления показателей соотношения** аналогична технике вычисления интенсивных показателей:

$$\text{Показатель соотношения} = \frac{\text{Явление А}}{\text{Явление В}} * 1; 100 (1000 \dots \text{и т. д.})$$

Если интенсивные коэффициенты характеризуют частоту явлений, происходящих в данной среде (т.е. непосредственно связанных с этой средой), то коэффициенты соотношения указывают на числовое соотношение двух явлений, непосредственно между собой не связанных.

Примером таких относительных чисел может служить показатель обеспеченности населения врачами (число врачей на 10.000 населения). Коэффициенты соотношения находят широкое применение при характеристике различных видов медицинской деятельности (например, показатели применения лечебно-диагностических методов).

1.3. Экстенсивные коэффициенты

Экстенсивные коэффициенты - характеризуют распределение явления на его составные части, его внутреннюю структуру или отношение частей к целому (удельный вес).

При вычислении экстенсивных показателей мы имеем дело только с **одной** статистической совокупностью (только с больными, только с умершими) и поэтому, как бы детально ни дифференцировался их внутренний состав, понятие о частоте явления получить нельзя, так как отсутс-

твует среда, т.е. основной фон. Большинство экстенсивных коэффициентов выражается в процентах, реже - в промилле или долях единицы.

Вычисление экстенсивных коэффициентов производится следующим образом. Например: в 2005 г. население города Р. составило 1.318.600 человек, в том числе мужчин – 605.300 человек. Если принять все население города Р. за 100% , то доля мужчин составит:

$$\begin{array}{l} 1,318,600 - 100\% \\ 605,300 - x \end{array} \qquad x = \frac{605,300 * 100}{1,318,600} = 45,9\%$$

Экстенсивными коэффициентами можно характеризовать структуру рождаемости (распределение родившихся по полу, росту, весу); структуру смертности (распределение умерших по возрасту, полу и причинам смерти); структуру заболеваемости (распределение больных по нозологическим формам); состав населения по полу, возрасту и социальным группам и др.

Характерной чертой экстенсивных коэффициентов является их взаимосвязанность, вызывающая определенный автоматизм сдвигов, т.к. их сумма всегда составляет 100%. Например, при изучении структуры заболеваемости удельный вес какого-нибудь отдельного заболевания может возрасти в следующих случаях:

- 1) при подлинном его росте, т.е. при увеличении интенсивного показателя;
- 2) при одном и том же его уровне, если число других заболеваний в этот период снизилось;
- 3) при снижении уровня данного заболевания, если уменьшение числа других заболеваний происходило более быстрыми темпами.

Экстенсивные коэффициенты дают представление об удельном весе того или иного заболевания (или класса болезней) только в данной группе населения и только за определенный период.

1.4. Коэффициенты наглядности

Коэффициенты наглядности – применяются с целью более наглядного и доступного сравнения рядов абсолютных, относительных или средних величин. Они представляют технический прием преобразования цифровых показателей.

Этот коэффициент получают путем преобразования ряда величин по отношению к одной из них – **базисной** (любой, не обязательно начальной). Эта базисная величина принимается за 1; 100; 1000 и т.п., а остальные величины ряда, при помощи обычной пропорции, пересчитываются по отношению к ней.

Рождаемость в России за 2012 и 2015 гг. (на 1000 нас.)

Годы	Интенсивный коэффициент	Коэффициенты наглядности (уровень рождаемости в 1997 году принят за 100%)
2012	8,6	100,0%
2015	8,3	96,5%

Коэффициенты наглядности могут быть применимы для демонстрации тенденций динамических сдвигов и изменений в изучаемом процессе (в сторону увеличения или уменьшения).

2. Средние величины. Меры оценки разнообразия признака в совокупности и типичности средних величин

Средние величины представляют собой второй тип производных величин, находящих широкое применение в медицинской статистике. *Средняя величина является сводной, обобщающей характеристикой статистической совокупности по определенному изменяющемуся количественному признаку* (средний рост, средний вес, средний возраст умерших). *Средняя величина отражает общее определяющее свойство всей статистической совокупности в целом, заменяя его одним числом с типичным значением данного признака.* Средняя величина нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и характеризует постоянное свойство явлений.

В медицине средние величины могут использоваться для характеристики физического развития, основных антропометрических признаков (морфологических и функциональных: рост, вес, динамометрия и др.) и их динамики (средние величины прироста или убыли признака). Разработка этих показателей и их сочетаний в виде стандартов имеет большое практическое значение для анализа здоровья населения (в особенности детей, спортсменов). Эпидемиологи рассчитывают среднее число заболеваний в очаге, распределение очагов по срокам и средние сроки производства дезинфекции.

В демографических и медико-социальных исследованиях рассчитываются: средняя продолжительность предстоящей жизни, средний возраст умерших, средняя численность населения и т.д.

В экспериментально-лабораторных исследованиях также используются средние величины: температура, число ударов пульса в минуту, уровень артериального давления, средняя скорость или среднее время реакции на тот или иной раздражитель, средние уровни содержания биохимических элементов в крови и др.

И статистические коэффициенты, и средние величины представляют собой вероятностные величины, однако между ними существуют значительные различия:

1) Статистические коэффициенты характеризуют признак, встречающийся только у некоторой части совокупности (так называемый альтернативный признак), который может наступить, но может и не наступить (рождение, смерть, заболевание). Средние величины характеризуют, признаки, присущие всей совокупности, но в разной степени (вес, рост, дни лечения).

2) Статистические коэффициенты применяются для измерения качественных (атрибутивных или описательных) признаков, а средние - для варьирующих количественных признаков, где речь идет об отличиях в числовых размерах признака, а не о факте его наличия или отсутствия.

Основное достоинство средних величин их типичность - средняя сразу дает общую характеристику явления. В связи с этим можно выделить два основных требования для вычисления средних величин:

- однородность совокупности;
- достаточное число наблюдений.

Любое распределение случайной величины, не обязательно подчиняющееся определенному закону распределения вероятностей, характеризуется параметрами распределения: средняя величина (M), среднее квадратическое отклонение (σ), коэффициент вариации (Cv) и др.

Например, при изучении распределения 10 больных по срокам лечения, мы получим ряд числовых значений: 38, 13, 17, 20, 14, 18, 25, 32, 23, 25 - **неупорядоченный ряд**.

Рассчитать параметры распределения можно, пользуясь и таким рядом. Однако охарактеризовать ряд несколькими параметрами еще недостаточно, необходимо исследовать, есть ли в статистическом ряду какая-либо устойчивая закономерность. Но, пользуясь неупорядоченным рядом, возможную закономерность обнаружить сложно, поэтому строят ранжированные ряды.

Ряд, в котором дается распределение единиц изучаемой совокупности по значениям варьирующего признака, называется вариационным. Другими словами – **вариационный ряд** – ряд однородных величин, расположенных в возрастающем или убывающем порядке, где варианты (группы вариантов) отличаются друг от друга на определенную величину, называемую интервалом (i).

Таким образом, ряд распределения больных по срокам лечения можно представить следующим образом:

V (дни)	13	14	17	18	20	22	23	25	32	38
p	1	1	1	1	1	1	1	1	1	1

Меняющийся, варьирующий признак изучаемого явления (рост, вес и др.), его числовое значение называется **вариантой** (V).

Числа случаев наблюдения данного признака, указывающие сколько раз встречается данная варианта, называются **частотами** (p).

Вариационные ряды могут быть:

1) в зависимости от изучаемого явления:

- дискретные (прерывные) – образуются на основе прерывно меняющихся признаков, значения которых выражаются только в целых числах (частота пульса, количество студентов в группе и т.д.);

- интервальные (непрерывные) – образуются обычно на основе признаков, которые могут принимать любые значения и выражаются любым числом (рост, вес и т.д.)

2) в зависимости от числа наблюдений:

- простые – варианта представлена одним числовым значением;

- сгруппированные – варианты группируются по определенному признаку. Например, при изучении физического развития может производиться группировка по весу: 40-44 кг; 45-49 кг. и т.д.

3) в зависимости от порядка расположения вариант:

- возрастающие – варианты располагаются в порядке возрастания;

- убывающие – варианты располагаются в порядке убывания.

Отдельный вариационный ряд может одновременно включать в себя несколько характеристик. Например, простой, убывающий, прерывный; или – сгруппированный, возрастающий, непрерывный.

Виды средних величин, которые обычно используются в медицинской статистике, - это медиана, мода, средняя арифметическая. Другие виды средних: средняя гармоническая, средняя квадратическая, средняя кубическая, средняя геометрическая и другие - применяются лишь в специальных исследованиях.

Медиана (Me) - это срединная, центральная варианта, делящая вариационный ряд пополам на две равные части.

Например, если число наблюдений составляет 33, медианой будет варианта, занимающая 17-е ранговое место, так как в обе стороны от нее находится по 16 наблюдений.

В ряде с четным числом наблюдений в центре находятся две величины. Если они одинаковы по своему значению, не возникает затруднений в приближенном определении медианы, если же числовые значения двух величин различны, то за медиану принимается их полусумма.

Мода (Mo) – это чаще всего встречающаяся или наиболее часто повторяющаяся величина признака. При приближенном нахождении моды в простом (не сгруппированном) ряде, она определяется как варианта с наибольшим количеством частот.

Отличие медианы и моды от средней арифметической заключается в том, что при упрощенном, ориентировочном определении эти величины легко и быстро найти по их положению в вариационном ряду (*позиционные средние*), кроме того, они не зависят от значений крайних вариант или от степени рассеяния ряда.

2.1. Средняя арифметическая величина (M).

Чаще всего используется в медицинской статистике **средняя арифметическая величина** (M - от латинского Media). Средняя арифметическая может быть **простая** и **взвешенная**.

Примером **средней арифметической простой** может служить результат измерения веса, например, 6 человек:

V (кг)	59	60	61	62	63	64	$\Sigma = 369$
p	1	1	1	1	1	1	$\Sigma p = n = 6$

Расчет производится по формуле:

$$M = \frac{\Sigma V}{n}$$

Сумма этих измерений, деленная на число наблюдений, и дает среднюю величину веса:

$$M = \frac{369}{6} = 61,5 \text{ кг.}$$

Таким образом, средняя арифметическая простая получается как сумма величин (вариант), деленная на их число. Среднюю арифметическую простую можно вычислить лишь в тех случаях, когда каждая величина (варианта) представлена единичным наблюдением, т. е. когда частоты равны единице.

Если частоты вариант больше единицы, простая средняя неприменима - здесь надо вычислять **среднюю арифметическую взвешенную**, которая получается как сумма произведений вариант на соответствующие частоты, деленная на общее число наблюдений:

$$M = \frac{\Sigma V * p}{n}$$

Например: частота пульса (число ударов в минуту) у 18 студентов после проведения атропиновой пробы составила: 86, 92, 100, 96, 90, 102, 88, 92, 80, 92, 96, 100, 86, 84, 102, 90, 86, 92.

V(уд/мин)	80	84	86	88	90	92	96	100	102	
p	1	1	3	1	2	4	2	2	2	$\Sigma p = n = 18$
Vp	80	84	258	88	180	358	192	200	204	$\Sigma Vp = 1644$

$$M = \frac{1644}{18} = 91,3 \text{ уд/мин.}$$

Средняя арифметическая простая - это частный случай средней арифметической взвешенной, поэтому формула средней арифметической взвешенной может использоваться и для расчета средней арифметической простой. В последнем случае частоты равны единице и умножение излишне.

Все три средние величины (*Mo*, *Me*, *M*) совпадают (либо практически очень близки) в симметричном вариационном ряду: средняя арифметическая соответствует середине ряда (в симметричном ряду отклонения в сторону увеличения и в сторону уменьшения вариантов соответственно уравниваются); медиана (как центральная величина) также соответствует середине ряда; мода (как наиболее насыщенная величина) приходится на наивысшую точку ряда, также находящуюся в его центре. Поэтому для всех симметричных рядов нет необходимости вычислять другие средние величины, кроме средней арифметической.

Свойства средней арифметической величины:

1. Средняя величина является обобщающей характеристикой статистической совокупности по определенному изменяющемуся количественному признаку, отражает общее определяющее свойство всей статистической совокупности в целом, заменяя его одним числом с типичным значением данного признака. Средняя величина нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и характеризует постоянное свойство явлений.

2. Сумма отклонений вариант от средней арифметической величины равна 0.

3. В строго симметричном вариационном ряду средняя арифметическая занимает срединное положение и равна *Mo*, *Me*.

Средние арифметические величины, взятые сами по себе без дополнительных приемов оценки, часто имеют ограниченное значение, так как они не отражают степени рассеяния (разнообразия) ряда. Одинаковые по размеру средние величины могут быть получены из рядов с различной степенью рассеяния. Средние - это величины, вокруг которых рассеяны различные варианты, и *чем ближе друг к другу отдельные варианты, тем меньше рассеяние ряда, тем типичнее средняя величина.*

Приближенным методом оценки разнообразия ряда может служить определение амплитуды. **Амплитуда** - разность между наибольшим и наименьшим значением вариант:

$$A = V_{max} - V_{min}$$

Но амплитуда не учитывает промежуточные значения вариант внутри ряда, кроме того, ее размеры могут зависеть и от числа наблюдений.

2.2. Среднее квадратическое отклонение (σ).

Основной мерой оценки разнообразия ряда является **среднее квадратическое отклонение (σ).**

Вычисление точного значения среднего квадратического отклонения производится по формуле:

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

Если число наблюдений меньше 30 (малая выборка), то расчет производится по формуле:

$$\sigma = \pm \sqrt{\frac{\sum d^2 p}{n - 1}}$$

Для вычисления сигмы необходимо:

- 1) определить отклонения (d) от средней ($V - M$);
- 2) возвести отклонения в квадрат (d^2);
- 3) перемножить квадраты отклонений на частоты ($d^2 p$);
- 4) суммировать произведения квадратов отклонений на частоты;
- 5) разделить эту сумму на число наблюдений;
- 6) извлечь из частного квадратный корень.

При помощи сигмы можно установить степень типичности средней, пределы рассеяния ряда, пределы колебаний вокруг средней отдельных вариантов. Чем меньше сигма, тем меньше рассеяние ряда, тем точнее и типичнее получается вычисленная для этого ряда средняя величина.

Применение сигмы дает возможность оценки и сравнения разнообразия нескольких *однородных* рядов распределения, так как σ - величина именная, выражается абсолютным числом в единицах изучаемой совокупности (см, кг, мг/л и т.д.). В этом случае принимаются во внимание абсолютные размеры сигмы. Например, при сравнении двух рядов распределения по признаку веса, при условии, что средние будут близки по уровню, но сигма в одном ряду будет $\pm 5,6$ кг., а в другом $\pm 2,1$ кг. - второй ряд менее рассеян, и его средняя более типична.

2.3. Коэффициенту изменчивости (вариации) C_v (V).

При оценке разнообразия неоднородных рядов (например, таких признаков как вес и рост), непосредственное сравнение размеров сигмы невозможно. В этом случае, для установления степени относительного разнообразия рядов, прибегают к производной величине - **коэффициенту изменчивости (вариации)**, который является относительной величиной, выражается в % и обозначаемому буквой C_v (V).

Коэффициент изменчивости получается из процентного отношения сигмы к средней:

$$C_v = \frac{\sigma}{M} * 100\%.$$

Например, при изучении физического развития студентов – мужчин 1 курса получены следующие показатели: M (вес) = 67,5 кг.; M (рост) = 178,1 см. Соответственно σ = $\pm 2,8$ кг. и $\pm 6,2$ см. Среднее квадратическое

отклонение по росту более чем в 2 раза превышает сигму по весу. Коэффициент вариации C_v равен:

$$C_v(\text{по весу}) = \frac{2,8 \text{ кг}}{67,5 \text{ кг}} * 100\% = 4,1\%$$

$$C_v(\text{по росту}) = \frac{6,2 \text{ см}}{178,1 \text{ см}} * 100\% = 3,5 \%$$

Коэффициент вариации по росту меньше, чем по весу, то есть рост оказался более устойчивым признаком, чем вес.

Различают три степени разнообразия коэффициентов вариации:

- до 10% - слабое разнообразие;
- 10 – 20 % - среднее разнообразие;
- более 20 % - сильное разнообразие.

Этот же метод вычисления коэффициента разнообразия пригоден и при анализе однородных рядов, у которых средние величины очень разнятся по размеру, а также для оценки изолированного, единичного ряда.

3. Пример вычисления средней арифметической (M); среднего квадратического отклонения (σ); коэффициента вариации (C_v):

Длительность лечения ангины у 45 больных составила: 20, 20, 19, 16, 19, 16, 14, 13, 15, 13, 12, 13, 13, 3, 12, 11, 12, 11, 10, 12, 11, 10, 11, 8, 7, 11, 11, 10, 10, 10, 9, 8, 8, 9, 5, 5, 6, 9, 5, 5, 9, 6, 7, 7, 14, и 15 дней.

Первый этап: Строим вариационный ряд, с учетом частоты встречаемости каждой варианты; даем характеристику ряда; находим произведения вариант на соответствующую частоту, суммируем полученные произведения и рассчитываем среднюю арифметическую:

Первый этап			Второй этап		
Длительность лечения (в днях) V	Число больных p	Vp	$d (V-M)$	d^2	d^2p
20	2	40	9	81	162
19	2	38	8	64	128
16	2	32	5	25	50
15	2	30	4	16	32
14	2	28	3	9	18
13	4	52	2	4	16
12	4	48	1	1	4
11	6	66	0	0	0
10	5	50	-1	1	5
9	4	36	-2	4	16
8	3	24	-3	9	27
7	3	21	-4	16	48
6	2	12	-5	25	50

5	3	15	-6	36	108
3	1	3	-8	64	64
$\Sigma p = n = 45$		$\Sigma Vp = 495$			$\Sigma d^2 p = 728$
Ряд простой, убывающий, прерывный					

$$M = \frac{\Sigma V * p}{n} = \frac{495}{45} = 11 \text{ дней.}$$

Второй этап: рассчитываем $d (V-M)$; d^2 ; $d^2 p$.

Третий этап: рассчитываем среднее квадратическое отклонение (σ); коэффициент вариации (C_v):

$$\sigma = \pm \sqrt{\frac{\Sigma d^2 p}{n}} = \pm \sqrt{\frac{728}{45}} = \pm \sqrt{16,2} = \pm 4,02 \text{ дня}$$

$$C_v = \frac{\sigma}{M} * 100\% = \frac{4,02}{11} * 100\% = 36,5\%$$

Заключение: Средняя длительность лечения ангины в поликлинике составила 11 дней. Средняя является недостаточно типичной для данного ряда, о чем свидетельствует коэффициент вариации, равный 36,5% (большая степень разнообразия признака).

3. Цель деятельности студентов на занятии:

Студент должен знать:

1. Что изучает Медицинская статистика.
2. Этапы статистического исследования.
3. Что такое относительные величины.
4. Что такое средние величины.

Студент должен уметь:

1. Составлять план статистического исследования.
2. Вычислять относительные величины.
3. Вычислять средние величины.

4. Содержание обучения:

1. Относительные величины (статистические коэффициенты)
 - 1.1. Интенсивные коэффициенты.
 - 1.2. Коэффициенты соотношения.
 - 1.3. Экстенсивные коэффициенты
 - 1.4. Коэффициенты наглядности
2. Средние величины. Меры оценки разнообразия признака в совокупности и типичности средних величин
 - 2.1. Средняя арифметическая величина (M).
 - 2.2. Среднее квадратическое отклонение (σ).
 - 2.3. Коэффициенту изменчивости (вариации) Cv (V).
3. Пример вычисления средней арифметической (M); среднего квадратического отклонения (σ); коэффициента вариации (Cv):

5. Перечень вопросов для проверки уровня знаний:

1. Чем обусловлена необходимость вычисления относительных величин?
2. В чем отличие интенсивных коэффициентов от коэффициентов соотношения?
3. С какой целью применяются показатели наглядности, и что они характеризуют?
4. Назовите область применения средних величин в медицине.
5. Какие различия существуют между средними величинами и статистическими коэффициентами?
6. Что должны характеризовать средние величины в статистике?
7. Какие виды средних величин обычно используются в медицинской статистике?

6. Перечень вопросов для проверки конечного уровня знаний:

1. Перечислите виды относительных величин.
2. Что характеризуют интенсивные показатели?
3. Что характеризуют показатели соотношения?
4. Что характеризуют экстенсивные показатели?
5. Как рассчитывают показатели наглядности?
6. Дайте определение средних величин.
7. Дайте определение вариационного ряда, варианты и частоты встречаемости варианты.
8. Назовите виды вариационных рядов.
9. Что такое мода?
10. Что такое медиана?
11. Как вычисляется средняя арифметическая (простая, взвешенная)?
12. Как определяется амплитуда ряда?

13. Как определяется среднее квадратическое отклонение?
14. Как определяется коэффициент вариации?
15. Назовите степени разнообразия коэффициента вариации.

7. Практическая часть

Задача 1. Пользуясь приведенными данными, рассчитайте интенсивные и экстенсивные показатели, если численность населения города П. составляет – 1.308.400 человек.

Из них в возрасте: 0 - 14 лет - 223.600 человек
 15 - 49 лет - 647.800 тыс. человек
 50 лет и старше - 437.000 тыс. человек

Родилось (за год) - 9684 человек. Умерло (за год) - 22.508 человек.

Задача 2. Пользуясь приведенными данными, рассчитайте интенсивные и экстенсивные показатели, если численность населения города С. составляет – 2.181.300 человек. Из них: городское население – 1.201.200 человек; сельское население - 980.100 человек

Задача 1. У 10 рабочих, имевших в течение пяти лет контакт со свинцом, определяли его содержание в моче. Концентрация свинца составила (в мг/л): 68, 70, 78, 75, 71, 81, 74, 73, 77, 76.

Задача 2. В больнице проанализировано 35 «Медицинских карт» лиц, перенесших катаральную форму ангины. Сроки лечения составили (в днях): 3, 5, 6, 7, 4, 5, 3, 3, 5, 6, 5, 7, 6, 5, 4, 9, 5, 6, 4, 3, 7, 5, 6, 7, 4, 3, 5, 5, 4, 3, 6, 7, 4, 6, 5.

8. Самостоятельная работа студентов:

По учебнику изучите основные понятия теории вероятности и медицинской статистики.

9. Хронокарта учебного занятия:

1. Организационный момент – 5 мин.
2. Текущий контроль знаний – 30 мин.
3. Разбор темы – 20 мин.
4. Практическая работа – 30 мин.
5. Подведение итогов занятия – 10 мин.

10. Перечень учебной литературы к занятию:

1. Кобринский Б.А., Зарубина Т.В. «Медицинская информатика», М., Издательский дом «Академия», 2009.
2. Жижин К.С. «Медицинская статистика», Высшее образование, 2007.

Тема 4-5: «Меры оценки разнообразия признака в совокупности и типичности средних величин»»

1. Научно-методическое обоснование темы:

Ранее упоминалось о применении выборочного метода наблюдения. Под выборочным методом в статистике понимается такой метод наблюдения, при котором для отыскания типичных черт характеристик какой-либо совокупности изучаются не все единицы этой совокупности, а лишь часть их. Как бы тщательно ни производилась выборка, какой репрезентативной ни была бы выборочная совокупность (отобранная часть наблюдений), она неизбежно будет отличаться от всей генеральной (общей) совокупности. Таким образом, полного тождества достичь не удастся, и некоторая неточность встречается неизбежно. Однако имеются методы установления степени различий числовых характеристик обеих совокупностей и пределов возможных колебаний показателей при данном числе наблюдений. Число наблюдений играет значительную роль - чем больше число наблюдений, тем точнее отображается генеральная совокупность и тем меньше размеры ошибки.

Так называемые средние ошибки являются мерой точности и достоверности любых статистических величин. *Под достоверностью статистических показателей* (синонимы: существенность, значимость, надежность) *понимают доказательность, то есть право на обобщение явления, правомерность распространения выводов и на другие аналогичные явления. Или - степень их соответствия отображаемой ими действительности.* Достоверными результатами считаются те, которые не искажают и правильно отражают объективную реальность.

2. Краткая теория:

1. Оценка достоверности результатов исследования

Оценить достоверность результатов исследования означает определить, с какой вероятностью возможно перенести результаты, полученные на выборочной совокупности, на всю генеральную совокупность.

В большинстве медицинских исследований врачу приходится, как правило, иметь дело с частью изучаемого явления, а выводы по результатам такого исследования переносить на все явление в целом - на генеральную совокупность.

Оценка достоверности результатов исследования предусматривает определение:

- 1) ошибок репрезентативности (средних ошибок средних арифметических и относительных величин) - m ;
- 2) доверительных границ средних (или относительных) величин;
- 3) достоверности разности средних (или относительных) величин (по критерию t - Стьюдента).

1.1. Определение средней ошибки средней (или относительной) величины (ошибка репрезентативности – t).

Теория выборочного метода, наряду с обеспечением репрезентативности, практически сводится к оценке расхождений между числовыми характеристиками генеральной и выборочной совокупности, т. е. к определению средних ошибок и так называемых доверительных границ или интервалов. *Средняя ошибка позволяет установить тот интервал, в котором заключено действительное значение производной величины при данном числе наблюдений, т. е. средняя ошибка всегда является конкретной.*

Ошибка репрезентативности является важнейшей статистической величиной, необходимой для оценки достоверности результатов исследования. Эта ошибка возникает в тех случаях, когда требуется по части охарактеризовать явление в целом. Эти ошибки неизбежны. Они «вытекают» из сущности выборочного исследования. Генеральная совокупность может быть охарактеризована по выборочной совокупности только с некоторой погрешностью, измеряемой ошибкой репрезентативности.

Ошибки репрезентативности не тождественны обычным представлением об ошибках: методических, точности измерения, арифметических и др.

По величине ошибки репрезентативности определяют, насколько результаты, полученные при выборочном исследовании, отличаются от результатов, которые могли бы быть получены при проведении сплошного исследования без исключения всех элементов генеральной совокупности.

Это единственный вид ошибок, учитываемых статистическими методами, которые не могут быть устранены, если не проведено сплошное исследование.

Ошибки репрезентативности можно свести к достаточно малой величине, т.е. к величине допустимой погрешности. Делается это путем увеличения числа наблюдений (n).

Каждая средняя величина - M (средняя длительность лечения, средний рост, средняя масса тела и др.), а также относительная величина - P (уровень летальности, заболеваемости и др.) должны быть представлены со своей средней ошибкой - m .

Средняя арифметическая величина выборочной совокупности (M) имеет ошибку репрезентативности, которая называется **средней ошибкой средней арифметической** (m_M) и определяется по формуле:

$$m_M = \pm \frac{\sigma}{\sqrt{n}}$$

Как видно из этой формулы, между размерами сигмы (отражающей разнообразие явления) и размерами средней ошибки существует прямая

связь. Между числом наблюдений и размерами средней ошибки существует обратная связь (пропорциональная не числу наблюдений, а квадратному корню из этого числа). Следовательно, уменьшение величины этой ошибки при определении степени разнообразия (σ) возможно путем увеличения числа наблюдений. При числе наблюдений менее 30 в знаменателе следует взять $(n - 1)$.

$$m_M = \pm \frac{\sigma}{\sqrt{n - 1}}$$

На этом принципе основан метод определения достаточного числа наблюдений для выборочного исследования.

Относительные величины (P), полученные при выборочном исследовании, также имеют свою ошибку репрезентативности, которая называется средней ошибкой относительной величины и обозначается m_p .

Для определения средней ошибки относительной величины (P) используется следующая формула:

$$m_p = \pm \sqrt{\frac{P * q}{n}}$$

Где: P - относительная величина.;

q – разность между основанием, на которое рассчитана относительная величина и самой относительной величиной. Если показатель выражен в процентах, то $q = 100 - P$; если P - в промиллях, то $q = 1000 - P$, если P - в продцимиллях, то $q = 10.000 - P$, и т.д.;

n - число наблюдений. При числе наблюдений менее 30 в знаменатель следует взять $(n - 1)$.

$$m_p = \pm \sqrt{\frac{P * q}{n - 1}}$$

Каждая средняя арифметическая или относительная величина, полученная на выборочной совокупности, должна быть представлена со своей средней ошибкой. Это дает возможность рассчитать доверительные границы средних и относительных величин, а также определить достоверность разности сравниваемых показателей (результатов исследования).

1.2. Определение доверительных границ.

Определяя для средней арифметической (или относительной) величины два крайних значения: минимально возможное и максимально возможное, находят пределы, в которых может быть искомая величина *генерального параметра*. Эти пределы называют доверительными границами.

Доверительные границы – границы средних (или относительных) величин, выход за пределы которых вследствие случайных колебаний имеет незначительную вероятность.

Вероятность попадания средней или относительной величины в доверительный интервал называется **доверительной вероятностью**.

Доверительные границы **средней арифметической генеральной совокупности** определяют по формуле:

$$M_{ген} = M_{выб} \pm t \cdot m_M$$

Доверительные границы относительной величины в генеральной совокупности определяют по следующей формуле:

$$P_{ген} = P_{выб} \pm t \cdot m_p$$

Где: $M_{ген}$ и $P_{ген}$ - значения средней и относительной величин, полученных для генеральной совокупности;

$M_{выб}$ и $P_{выб}$ - значения средней и относительной величин, полученных для выборочной совокупности;

m_M и m_p - ошибки репрезентативности выборочных величин;

t - доверительный критерий, который зависит от величины безошибочного прогноза, устанавливаемого при планировании исследования.

Произведение $t \cdot m$ (Δ) - предельная ошибка показателя, полученного при данном выборочном исследовании.

Размеры предельной ошибки зависят от коэффициента t , который избирает сам исследователь, исходя из заданной вероятности безошибочного прогноза.

Величина критерия t связана с вероятностью безошибочного прогноза (P) и числом наблюдений в выборочной совокупности (табл. 1.).

Таблица 1.

Зависимость доверительного критерия t от степени вероятности безошибочного прогноза P (при $n > 30$)

Степень вероятности безошибочного прогноза (P %)	Доверительный критерий t
95,0	2
99,0	2,6
99,9	3,3

Для большинства медико-биологических и социальных исследований достоверными считаются доверительные границы, установленные с вероятностью безошибочного прогноза = 95% и более.

Чтобы найти критерий t при числе наблюдений (n) < 30, необходимо пользоваться специальной таблицей Н.А. Плохинского (табл. 2), в которой слева показано число наблюдений – единица ($n - 1$), а сверху (P) – степень вероятности безошибочного прогноза.

При определении доверительных границ сначала надо решить вопрос о том, с какой степенью вероятности безошибочного прогноза необходимо представить доверительные границы средней или относительной величины. Избрав определенную степень вероятности, соответственно этому находят величину доверительного критерия t при данном числе наблюдений. Таким образом, доверительный критерий устанавливается заранее, при планировании исследования.

Таблица 2.

Значение критерия t для трех степеней вероятности (по Н.А.Плохинскому)

$n = n-1$	P	95%	99%	99,9%
1		12,7	63,7	37,0
2		4,3	9,9	31,6
3		3,2	5,8	12,9
4		2,8	4,6	8,6
5		2,6	4,0	6,9
6		2,4	3,7	6,0
7		2,4	3,5	5,3
8		2,3	3,4	5,0
9		2,3	3,3	4,8
10		2,2	3,2	4,6
11		2,2	3,1	4,4
12		2,2	3,1	4,3
13		2,3	3,0	4,1
14-15		2,1	3,0	4,1
16-17		2,1	2,9	4,0
18-20		2,1	2,9	3,9
21-24		2,1	2,8	3,8
25-29		2,0	2,8	3,7

Любой параметр (средняя или относительная величина) может оцениваться с учетом доверительных границ, полученных при расчете.

Например: требуется определить доверительные границы среднего уровня пепсина у больных гипертертиозом с 95% вероятностью безошибочного прогноза. Если известно, что:

$$n = 49;$$

$$M_{\text{выб}} = 1\text{г}\%;$$

$$m_M = \pm 0,05\text{г}\%$$

1.Определение доверительных границ средней величины в генеральной совокупности:

$$M_{\text{ген}} = M_{\text{выб}} \pm t \cdot m_M = 1\text{г}\% \pm 2 \cdot 0,05\text{г}\%$$

$$M_{\text{ген}} = \begin{matrix} 1\text{г}\% + 0,1\text{г}\% = 1,1\text{г}\% \\ 1\text{г}\% - 0,1\text{г}\% = 0,9\text{г}\% \end{matrix}$$

Заключение: установлено с вероятностью безошибочного прогноза 95%, что средний уровень пепсина в генеральной совокупности у больных гипертериозом находится в пределах от 1,1 г% до 0,9 г%.

Как видно, доверительные границы зависят от размера доверительного интервала.

Анализ доверительных интервалов указывает, что при заданных степенях вероятности и $n > 30$ - t имеет неизменную величину и при этом доверительный интервал зависит от величины ошибки репрезентативности.

С уменьшением величины ошибки суживаются доверительные границы средних и относительных величин, полученных на выборочной совокупности, т.е. уточняются результаты исследования, которые приближаются к соответствующим величинам генеральной совокупности. Если ошибка большая, то получают для выборочной величины большие доверительные границы, которые могут противоречить логической оценке искомой величины в генеральной совокупности. В подобном случае надо искать резервы сокращения размаха доверительных границ в размере величины ошибки репрезентативности.

Доверительные границы $M_{выб}$ и $P_{выб}$ зависят не только от средних ошибок этих величин, но и от избранной исследователем степени вероятности безошибочного прогноза. При большой степени вероятности размах доверительных границ увеличивается.

1.3. Определение достоверности разности средних (или относительных) величин (по критерию t - Стьюдента).

В медицине и здравоохранении по разности параметров оценивают средние и относительные величины, полученные для разных групп населения по полу, возрасту, а также групп больных и здоровых и т.д. Во всех случаях при сопоставлении двух сравниваемых величин возникает необходимость не только определить их разность, но и оценить ее достоверность.

Достоверность разности величин, полученных при выборочных исследованиях, означает, что вывод об их различии может быть перенесен на соответствующие генеральные совокупности.

Достоверность разности выборочной совокупности измеряется доверительным критерием, который рассчитывается по специальным формулам для средних и относительных величин.

Формула оценки достоверности разности сравниваемых средних величин:

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

Для относительных величин:

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

Где: $M_1; M_2 ; P_1; P_2$ - параметры, полученные при выборочных исследованиях;

$m_1; m_2$ - их средние ошибки;

t - критерий достоверности (Стьюдента).

Разность статистически достоверна при $t \geq 2$, что соответствует вероятности безошибочного прогноза, равной 95% и более.

Для большинства исследований, проводимых в медицине и здравоохранении, такая степень вероятности является вполне достаточной.

При величине критерия достоверности $t < 2$ степень вероятности безошибочного прогноза составляет $P < 95\%$. При такой степени вероятности нельзя утверждать, что полученная разность показателей достоверна с достаточной степенью вероятности. В этом случае необходимо получить дополнительные данные, увеличив число наблюдений.

Иногда при увеличении численности выборки разность продолжает оставаться не достоверной. Если при повторных исследованиях разность остается недостоверной, можно считать доказанным, что между сравниваемыми совокупностями не обнаружено различий по изучаемому признаку.

Например: требуется определить, достоверны ли различия в уровне пепсина в желудочном соке больных гипертериозом и здоровых лиц. Обследуются на пепсин две группы: 49 больных гипертериозом и 50 здоровых людей (контрольная группа). Результаты представлены в таблице 3.

Таблица 3.

Сравнение среднего уровня пепсина в желудочном соке больных гипертериозом и здоровых лиц

Сравниваемые группы	N	M (г%)	m (г%)	t	Уровень вероятности безошибочного прогноза (P)
Больные гипертериозом	49	1,0	$\pm 0,3$	10,0	< 99,9
Здоровые (контрольная группа)	50	4,0	$\pm 0,1$		

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

$$t = \frac{4 - 1}{\sqrt{0,3^2 + 0,1^2}}$$

Заключение: при гипертериозе наблюдается снижение уровня пепсина, что подтверждается с большой степенью вероятности безошибочного прогноза ($P > 99,9\%$). Следовательно, снижение уровня пепсина может быть

использовано в качестве одного из симптомов для подтверждения диагностики гипертериоза.

Подобным же образом оценивают достоверность разности сравниваемых относительных величин.

Указанная методика оценки достоверности и разности результатов исследования позволяет проводить только сравнение групп по парам, при обязательном наличии обобщающих параметров - средних арифметических или относительных величин и их средних ошибок.

2. Динамические ряды

Динамическим рядом называется совокупность однородных статистических величин, показывающих изменение явления на протяжении определенного промежутка времени.

Числа, из которых состоит динамический ряд, называют **уровнями** ряда. Уровень – это элемент динамического ряда.

Различают три основных типа динамических рядов в зависимости от составляющих его величин:

1. Динамические ряды, построенные из **абсолютных величин** (например, численность населения в различные годы) – **простой динамический ряд**.

2. Динамические ряды, построенные из **относительных величин** (демонстрирующие, например, изменения коэффициентов смертности) - **сложный (производный) динамический ряд**, так как такие ряды получаются из сочетания двух простых рядов (например, численности населения и числа смертей по годам).

3. Динамические ряды, построенные из **средних величин** (демонстрирующие, например, показатели физического развития – рост, вес и др.) – **сложный (производный) динамический ряд**, так как средние величины относятся к производным величинам.

Динамические ряды в зависимости от сроков, которые они отражают, делятся на: моментные и интервальные.

Моментный ряд состоит из величин, характеризующих размеры явления на определенные даты - моменты (например, на конец года – 31 декабря 2004 года). Уровни моментного ряда **не подлежат дроблению**.

Интервальный ряд - ряд чисел, строящийся из величин, учтенных не на одну дату, а за определенный отрезок (интервал) времени. Интервальный ряд можно разделить на дробные периоды, а можно укрупнить интервалы.

Анализ динамического ряда определяется показателями, характеризующими интенсивность его изменений и называемыми **коэффициентами динамики**, к которым относятся:

1) **Абсолютный прирост** или **убыль** (абсолютный размер разности уровней) - разность между последующим и предыдущим уровнем (дает возможность анализировать скорость происходящих изменений в ее абсолютном выражении).

2) **Темп прироста** или **убывания** - процентное отношение абсолютного прироста (или снижения) к предыдущему уровню.

3) **Темп роста** или **снижения** - процентное отношение последующего уровня к предыдущему.

4) Для анализа динамического ряда используются также **показатели наглядности**, хотя следует помнить, что для коэффициента наглядности не обязательны взаимосвязанные динамические изменения.

Динамические ряды характеризуют изменение показателей здоровья - уровень и темп снижения заболеваемости, демографические сдвиги (рождаемости, общей и младенческой смертности), изменения физического развития. Пример анализа динамического ряда (табл. 4.):

Таблица 4.

Динамика рождаемости населения Н-ской области
за 2010 - 2014 гг.

Год	Число родившихся на 1000 населения	Абсолютный прирост	Темп прироста	Темп роста	Показатель наглядности
2010	7,7	-	-	-	100,0%
2011	7,9	0,2	2,6%	102,6%	102,6%
2012	7,8	-0,1	-1,3%	98,7%	101,3%
2013	7,5	-0,3	-3,8%	96,2%	97,4%
2014	7,3	-0,2	-2,7%	97,3%	94,8%

Расчет показателей динамического ряда:

1) Абсолютный прирост: $7,9 - 7,7 = 0,2$
 $7,8 - 7,9 = -0,1$ и т.д.

2) Темп прироста: $0,2 : 7,7 \cdot 100\% = 2,6\%$
 $-0,1 : 7,9 \cdot 100\% = -1,3\%$ и т.д.

3) Темп роста: $7,9 : 7,7 \cdot 100\% = 102,6\%$
 $7,8 : 7,9 \cdot 100\% = 98,7\%$ и т.д.

4) Показатель наглядности: уровень 2010 г. принимаем за 100%
 $7,9 : 7,7 \cdot 100\% = 102,6\%$
 $7,8 : 7,7 \cdot 100\% = 101,3\%$ и т.д.

3. Прямой метод стандартизации

Стандартизация – это метод, позволяющий устранить (элиминировать) влияние качественной или количественной неоднородности сравниваемых групп (совокупностей) на величину интенсивных показателей. Например, уровни заболеваемости и смертности обычно наиболее высоки у детей и лиц старших возрастных групп, а наиболее низки у лиц молодого и среднего возраста. Следовательно, та группа населения, в составе которой относительно больше детей и лиц старших возрастов, может иметь более высокий уровень заболеваемости и смертности, даже если эти показатели в отдельных возрастных группах у нее ниже.

Устранение возможного влияния различий в составе сравниваемых совокупностей по какому-либо признаку (возрасту, полу, профессии) на значение интенсивных показателей достигается путем условного уравнивания составов изучаемых групп по данному признаку. Если имеются данные о составе групп (совокупностей) по полу, профессии и другим признакам, а также численности заболевших или умерших в каждой группе и, следовательно, можно определить повозрастные коэффициенты смертности и заболеваемости, то для вычисления стандартизованных коэффициентов прибегают к прямому методу стандартизации.

Стандартизованные показатели - гипотетические величины, они не дают представления об истинном размере явления, а указывают на то, каковы были бы показатели в группах, если бы различие в их составе было исключено.

Этапы расчета стандартизованных показателей.

I этап - расчет интенсивных показателей в отдельных группах по признаку различия (возрасту, полу и т.д.) и по совокупности в целом.

II этап - определение стандарта, то есть одинакового для сравниваемых совокупностей численного состава по данному признаку.

III этап - вычисление ожидаемых абсолютных величин в стандарте на основе групповых интенсивных показателей, получение итоговых чисел по сравниваемым совокупностям в целом путем суммирования ожидаемых величин.

IV этап - вычисление стандартизованных показателей для сравниваемых совокупностей.

Способы получения стандарта.

1. Сумма изучаемых групп.
2. Полусумма изучаемых групп.
3. Численный состав одной из групп.
4. Численный состав по литературным данным.

В качестве стандарта (например, при элиминировании различий в возрастном составе) можно принять возрастное распределение одной из сравниваемых групп, средний возрастной состав сравниваемых групп, либо возрастное распределение третьей группы, особенно такой, в состав которой входят сравниваемые группы. Стандарт следует выбирать каждый раз

применительно к конкретно изучаемому материалу и в связи с задачами, стоящими перед исследователями.

Пример расчета стандартизованных показателей: Известен возрастной состав населения и есть информация для расчета повозрастных коэффициентов смертности населения от злокачественных новообразований (в каждой возрастной группе). Методика вычисления стандартизованных коэффициентов прямым методом складывается из четырех последовательных этапов (табл.6.1).

Первый этап. Вычисление «повозрастных» коэффициентов смертности от злокачественных новообразований (отдельно для каждой возрастной группы).

Второй этап. Выбор стандарта осуществляется произвольно. В этом примере за стандарт взят возрастной состав населения в городе А.

Третий этап. Расчет «ожидаемых» чисел. Определяется сколько бы человек умерло от злокачественных новообразований в каждой возрастной группе населения города Б, при имеющихся повозрастных показателях смертности от злокачественных новообразований в этом городе, но при возрастном составе города А (стандарт).

Например, в возрастной группе «до 30 лет»:

$$\frac{6 - 100.000}{X_1 - 350.000} \quad X_1 = \frac{6 * 350.000}{100.000} = 21,0$$

или в возрастной группе 40 - 49 лет:

$$\frac{140 - 100.000}{X_3 - 95.000} \quad X_1 = \frac{140 * 95.000}{100.000} = 133,0$$

и так далее.

Четвертый этап. Расчет стандартизованных коэффициентов. Сумму «ожидаемых» чисел (1069,0) получаем из общей численности населения города А (700.000). Умерших от злокачественных новообразований на 100.000 населения приходится:

$$\frac{1069 - 700.000}{X_{ст} - 100.000} \quad X_1 = \frac{1069 * 100.000}{700.000} = 152,7$$

Заключение: если бы возрастной состав населения города Б был бы такой же, как в городе А (стандарт), то смертность населения от злокачественных новообразований в городе Б была бы существенно выше (152,7 против 120,2 на 100.000 населения).

Таблица 5.

Стандартизация коэффициентов смертности от злокачественных новообразований в городах А и Б (прямым методом)

Возрастные группы	Первый этап						Третий этап	
	Город А			Город Б			Смертность от злокачественных новообразований	«ожидаемое» число при фактической по возрастной смертности в г.Б и возрастном составе г.А (стандарт) на 100.000 населения
	Численность населения (тыс.)	Число умерших	Смертность от злокачественных новообразований	Численность населения (тыс.)	Число умерших	Смертность от злокачественных новообразований		
До 29 лет	350	14	4,0	750	45	6,0	$6 \cdot 350.000$ ----- = 21,0 100.000	
30 – 39 лет	100	25	25,0	120	36	30,0	$30 \cdot 100.000$ ----- = 30,0 100.000	
40 – 49 лет	95	114	120,0	125	175	140,0	$140 \cdot 95.000$ ----- = 133,0 100.000	
50 – 59 лет	75	240	320,0	95	361	380,0	$380 \cdot 75.000$ ----- = 285,0 100.000	
60 лет и >	80	544	680,0	110	825	750,0	$750 \cdot 80.000$ ----- = 600,0 100.000	
Всего	700	937	133,9	1200	1442	120,2	1069,0	

Четвертый этап:

$$\frac{1069 - 700.000}{X_{гр} - 100.000}$$

$$X_1 = \frac{1069 * 100.000}{700.000} = 152,7$$

3. Цель деятельности студентов на занятии:

Студент должен знать:

1. Что такое достоверность статистических показателей.
2. Что такое динамические ряды.
3. Что такое методы стандартизации.

Студент должен уметь:

1. Определять средние ошибки средней величины.
2. Определять доверительные границы.
3. Использовать прямой метод стандартизации.

4. Содержание обучения:

1. Оценка достоверности результатов исследования.
 - 1.1. Определение средней ошибки средней (или относительной) величины (ошибка репрезентативности – t).
 - 1.2. Определение доверительных границ.
 - 1.3. Определение достоверности разности средних (или относительных) величин (по критерию t - Стьюдента).
2. Динамические ряды
3. Прямой метод стандартизации

5. Перечень вопросов для проверки уровня знаний:

1. Что такое «достоверность».
2. Что включает в себя понятие «оценка достоверности результатов»?
3. Как определяются ошибки репрезентативности производных величин?
4. Что такое доверительные границы производных величин?

6. Перечень вопросов для проверки конечного уровня знаний:

1. Что влияет на доверительные границы?
2. Что обозначают термины «уровень вероятности безошибочного прогноза»?
3. Что такое критерий достоверности?
4. Как определить достоверность разности производных величин?
5. Опишите формулу оценки достоверности разности сравниваемых средних и относительных величин:
6. Перечислите три основных типа динамических рядов в зависимости от составляющих его величин.
7. Перечислите этапы расчета стандартизованных показателей.

7. Практическая часть

Задача 1. Провести анализ динамического ряда, используя данные таблицы.

Годы	Смертность от болезней нервной системы (на 100.000 населения)	Абсолютный прирост	Темп прироста	Темп роста	Показатель наглядности
1998	9,6				
1999	9,5				
2000	14,9				
2001	15,2				
2002	20,7				
2003	20,0				
2004	24,7				

Задача 2. Провести анализ динамического ряда, используя данные таблицы.

Годы	Умерло детей до 1 года (на 1000 родившихся)	Абсолютный прирост	Темп прироста	Темп роста	Показатель наглядности
1997	14,2				
1998	18,3				
1999	16,3				
2000	15,7				
2001	16,1				
2002	16,2				

Задача 3. Рассчитайте стандартизованные показатели, если известно распределение больных и умерших в двух больницах в зависимости от срока госпитализации при аппендиците (в абсолютных числах).

Срок госпитализации (в днях)	Больница № 1		Больница № 2	
	Число больных	Число умерших	Число больных	Число умерших
1 – 3	400	1	100	-
3 – 4	150	3	200	2
5 – 6	50	6	300	10
Всего	600	10	600	12

За стандарт принять сумму составов больных по срокам госпитализации в больницах № 1 и № 2.

Задача 4. Рассчитайте стандартизованные показатели, если известно распределение рабочих и длительно и часто болеющих лиц (ДЧБ) основных и вспомогательных цехов завода Н. по возрасту (в абсолютных числах).

Возраст	Основные цеха	Вспомогательные цеха
---------	---------------	----------------------

(в годах)	Число рабочих	Из них: ДЧБ	Число рабочих	Из них: ДЧБ
20 – 29	100	5	200	21
30 – 39	300	45	150	25
40 и старше	200	40	50	10
Всего	600	90	400	56

За стандарт принять полусумму составов рабочих по возрасту в основных и вспомогательных цехах.

8. Самостоятельная работа студентов:

По учебнику изучите основные понятия теории вероятности и медицинской статистики.

9. Хронокарта учебного занятия:

1. Организационный момент – 5 мин.
2. Текущий контроль знаний – 30 мин.
3. Разбор темы – 20 мин.
4. Практическая работа – 30 мин.
5. Подведение итогов занятия – 10 мин.

10. Перечень учебной литературы к занятию:

1. Кобринский Б.А., Зарубина Т.В. «Медицинская информатика», М., Издательский дом «Академия», 2009.
2. Жижин К.С. «Медицинская статистика», Высшее образование, 2007.

Тема 6: «Обзор графических возможностей системы *Statistica*»

1. Научно-методическое обоснование темы:

Эффективная работа специалиста-медика на сегодняшний день подразумевает знание не только медицинских дисциплин, но также и хорошее владение навыками автоматизированной обработки медицинских данных.

2. Краткая теория:

Двумерные графики

Трёхмерные последовательные графики

Трёхмерные статистические графики

Пиктографики

Матричные графики

Категоризованные графики

Расширение графических возможностей с помощью *STATISTICA Visual Basic*

Многомерные дихотомии

Система *STATISTICA* включает широкий спектр графических методов для визуального представления результатов исследований. Все графические средства системы *STATISTICA* обеспечивают возможность выбора встроенного аналитического интерактивного метода анализа и содержат большой набор программ настройки, позволяющих пользователю интерактивно управлять отображением информации на экране. Гибкие средства управления одновременно несколькими графиками дают возможность пользователю объединять различные графические изображения и осуществлять динамический обмен между приложениями (используя средства OLE).

В основном, графики применяются для следующих целей:

- объяснение структуры исходных данных и результатов исследования человеку, незнакомому с методами прикладной статистики
- выяснение природы распределения случайных величин, подгонка распределения, визуальная проверка статистических гипотез

В программном пакете *STATISTICA* предлагаются разнообразные графические методы, с помощью которых пользователь может запрашивать или самостоятельно организовывать построение графиков. Эти методы, включающие большой набор графиков различных типов, таких как пользовательские, статистические и специализированные, они дополняют друг друга, обеспечивая высокий уровень взаимосвязи между числовыми данными (начальными, промежуточными и конечными) и их графическим представлением. Например, в качестве элемента автоматического вывода в программе статистической обработки может быть введен запрос на построение специализированных графиков. Кроме того, фактически любые

данные (и/или метки), которые создаются в процессе работы программы, могут быть выведены в графическом виде с помощью встроенных сервисных программ.

Двумерные графики

2М графики представляют наиболее общие методы визуализации данных. Построение двумерных графиков на плоскости "XY" в *STATISTICA* дает множество способов интерпретации значений как координат точек в двумерном пространстве. Они представляют собой, с одной стороны, стандартный метод графического представления исходных данных (например, различные диаграммы рассеяния, гистограммы и др.), а с другой стороны, стандартный способ графического анализа (например, категоризованные нормальные вероятностные графики, вероятностные графики с исключенным трендом и др.).

Построить графики перечисленных типов можно из меню *Графика*.

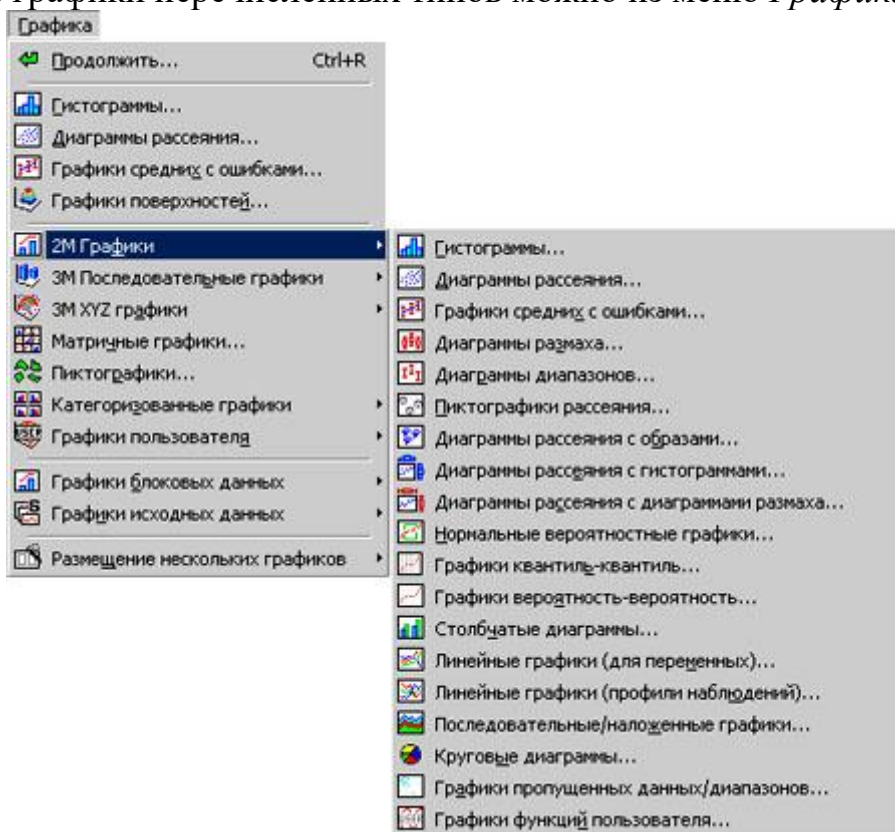


Рис. 1. Пункт меню *Графика/2М Графики*

Приведём несколько примеров построения двумерных графиков.



Рис. 2. Сечения графиков плотностей двумерного нормального распределения (нулевой вектор мат. ожидания, ковариационная матрица с 1 на главной диагонали) плоскостью $z=0.05$

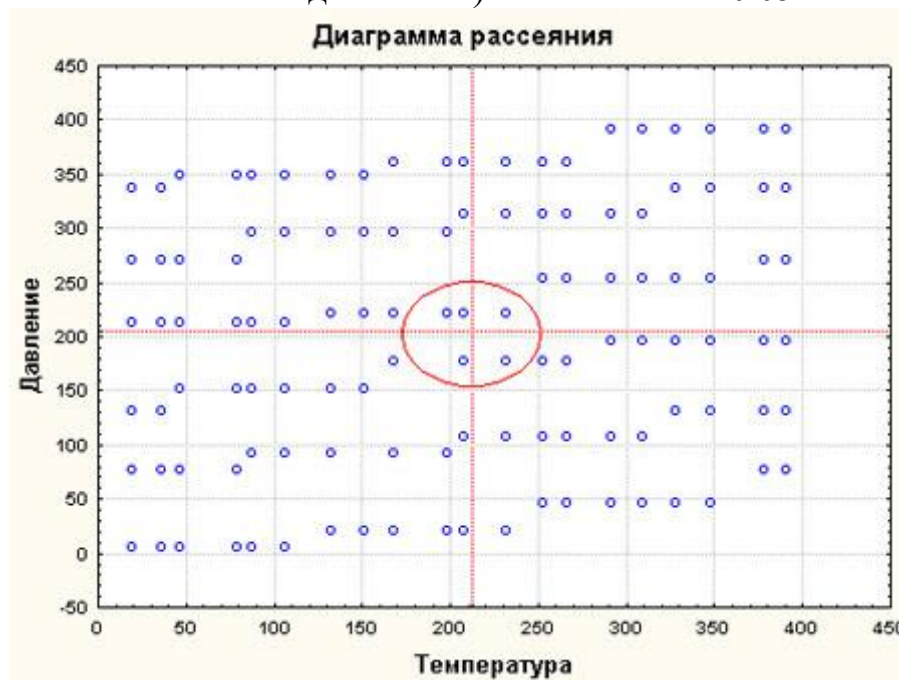


Рис. 3. Диаграмма рассеяния. Выделены точки с параметрами, максимально близкими к заданным.

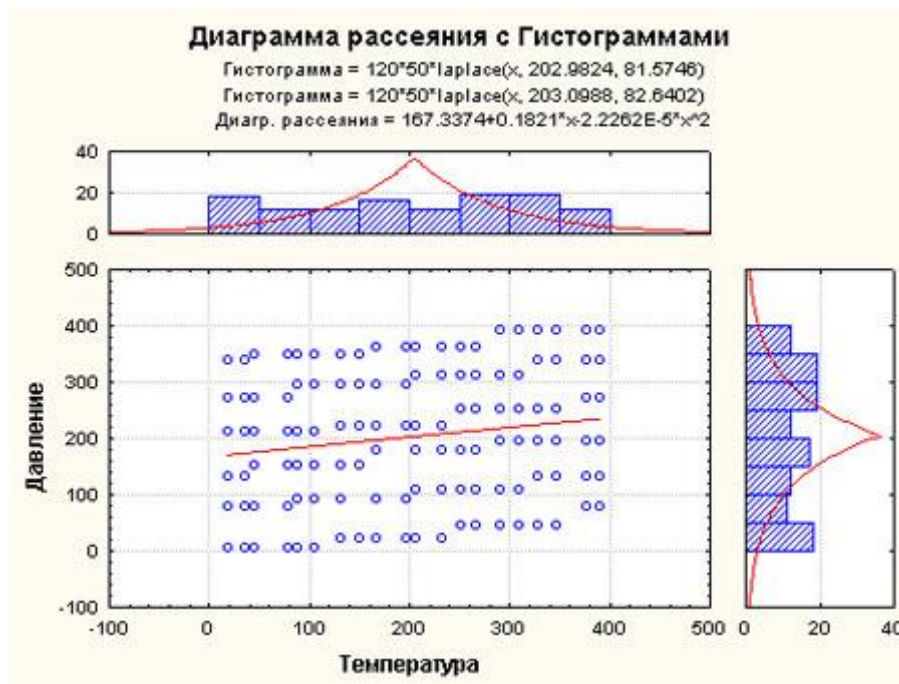




















Рис. 4. Диаграмма рассеяния с гистограммами. Построена диаграмма рассеяния двух переменных и их гистограммы с подгонкой распределения Лапласа.

К 2М Графикам относятся:

-  2М Гистограммы
-  Нормальные вероятностные графики
-  2М Диаграммы рассеяния
-  Графики квантиль – квантиль
-  Графики средних с ошибками
-  Графики вероятность – вероятность
-  2М Диаграммы размаха
-  2М Столбчатые диаграммы
-  2М Диаграммы диапазонов
-  2М Линейные графики
-  2М Линейные графики (профили наблюдений)
-  Пиктографики рассеяния
-  Диаграммы рассеяния с образами
-  Последовательные/наложенные графики
-  2М Диаграммы рассеяния с гистограммами
-  2М Диаграммы рассеяния с диаграммами размаха
-  Круговые диаграммы
-  Графики пропущенных данных/диапазонов

Трёхмерные последовательные графики

3М последовательные графики предлагают различные методы визуализации в трехмерном пространстве простых последовательностей значений, диапазонов значений или статистической информации о группах или подмножествах (частоты, распределения).

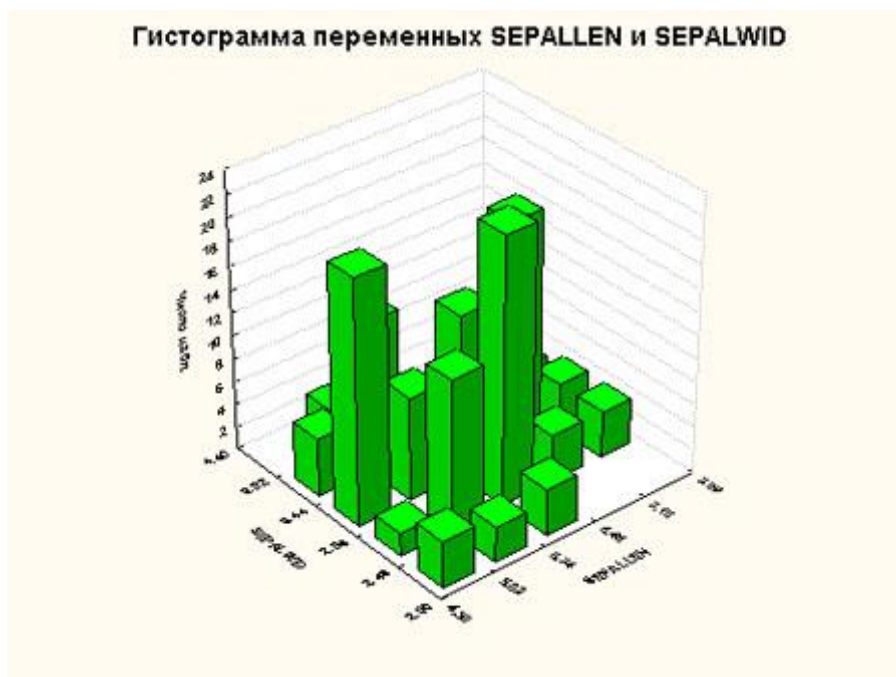






Рис. 5. Трёхмерная гистограмма переменных.

3М последовательные графики включают:

-  3М Диаграммы исходных данных
-  3М Диаграммы диапазонов
-  3М Гистограммы двух переменных
-  3М Диаграммы размаха

Трёхмерные статистические графики

Статистические XYZ графики содержат опции для создания различных типов 3М графиков, где расположение точек данных определяется тремя (а в случае тернарных графиков, четырьмя) координатами.

Независимо от типа создаваемого 3М XYZ графика и метода его создания (т.е. независимо от того, откуда были взяты данные для графика и как они вычислялись), к нему могут быть применены все средства настройки графиков системы *STATISTICA* для изменения вида графика или соединения его с другими графиками или документами. Также, все встроенные аналитические средства, вызываемые из самого графика (такие как функции подгонки, сглаживание, закрасивание и т.д.), можно применить к любому графику, независимо от источника данных и метода создания графика.

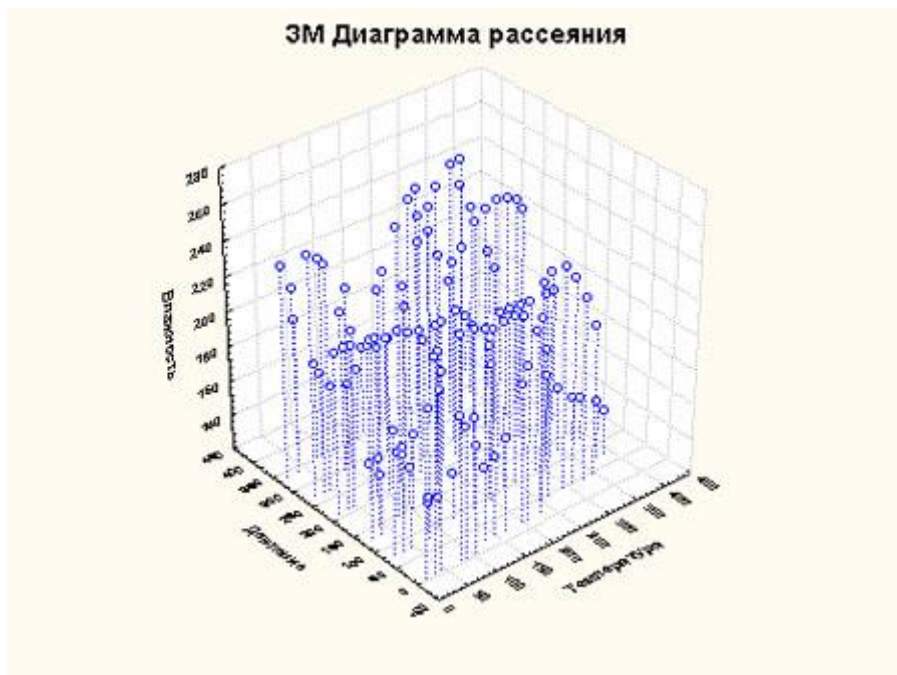







Рис. 6. Трёхмерная диаграмма рассеяния.

Ниже перечислены разновидности трёхмерных графиков:

-  3М Диаграммы рассеяния
-  3М Диаграммы рассеяния с образами
-  3М Графики поверхностей
-  3М Карты линий уровня
-  Тернарные графики

Пиктографики

Пиктографики обеспечивают графическое представление данных, где наблюдения или отдельные испытания отображаются в виде символов со многими элементами. Главная идея пиктографиков состоит в представлении отдельных единиц наблюдения в виде определенных графических объектов; при этом значения переменных ставятся в соответствие определенным характеристикам или параметрам объектов (обычно одно наблюдение равно одному объекту). Это соответствие таково, что общий вид объектов меняется как функция конфигурации значений. Таким образом, наблюдатель может "идентифицировать" уникальные для каждого набора значений наглядные характеристики объектов. Исследование таких пиктограмм может помочь обнаружить определенные группы простых зависимостей и взаимосвязей между переменными.

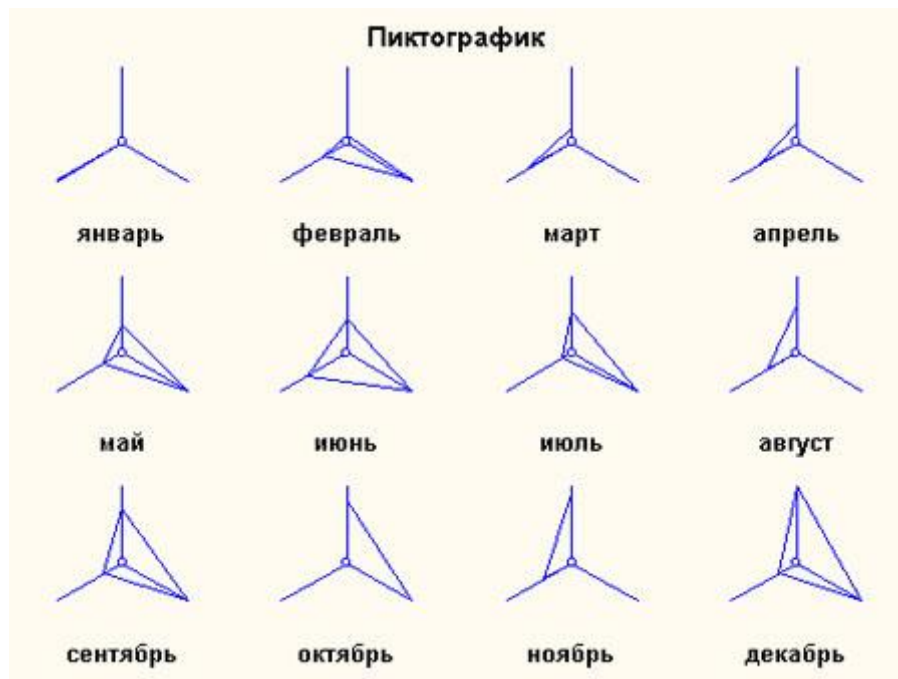


Рис. 7. Пиктографик типа *Лучи* для трёх переменных, содержащих некоторые данные для каждого месяца года.

Матричные графики

Матричные графики отражают зависимости между несколькими переменными в виде матрицы графиков X-Y. Наиболее часто используемым типом матричных графиков является матрица диаграмм рассеяния, которую можно считать графическим эквивалентом корреляционной матрицы.

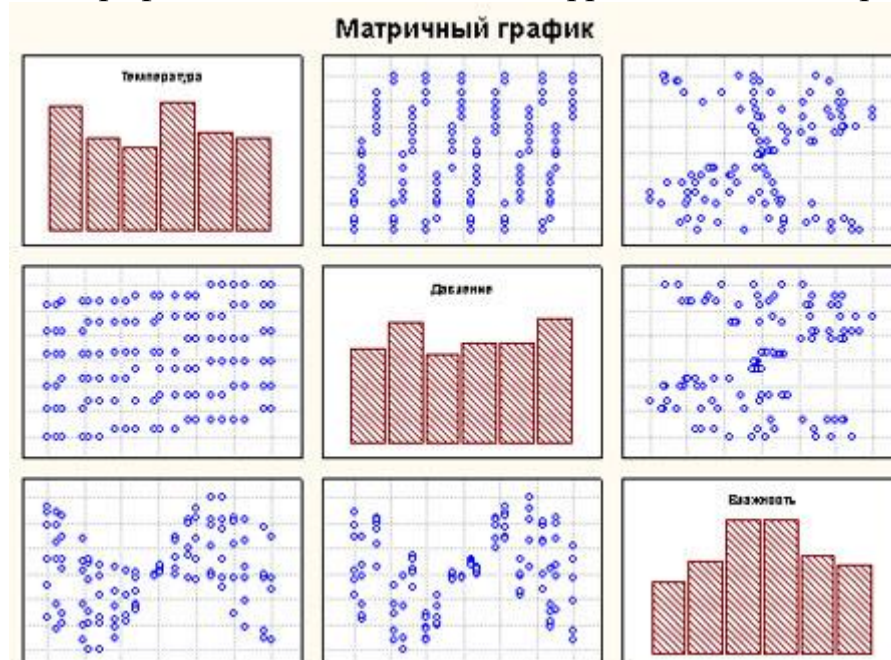


Рис. 8. Матричный график трёх переменных.

Категоризованные графики

Один из наиболее важных аналитических методов использует разделение множества данных на категории, чтобы сравнивать структуру данных в получившихся подмножествах. Этот способ известен под несколькими названиями (например, классификация, группировка, категоризация, разбиение, расслоение) и используется как для разведочного анализа данных, так и для проверки гипотез. Пример: зависимость возраста и риска инфаркта может различаться для женщин и мужчин. Зависимость между приёмом лекарств и снижением уровня холестерина может присутствовать только у женщин с низким кровяным давлением и только в возрасте тридцати-сорока лет. Индексы мощности процесса или гистограммы мощности могут различаться для временных промежутков, когда управление осуществляется разными операторами. Наклоны регрессии могут различаться для разных экспериментальных групп.

Многие вычислительные средства, основывающиеся на группировке, предназначены для количественного описания различий, выявляемых с помощью группировки (например, в дисперсионном анализе). Однако графические средства дают особые преимущества, которых нельзя достичь, используя лишь вычислительный метод: они могут выявлять закономерности, которые трудно поддаются количественному описанию (например, сложные взаимодействия, исключения, аномалии), и предоставляют уникальные обширные возможности аналитического исследования или "добычи данных".

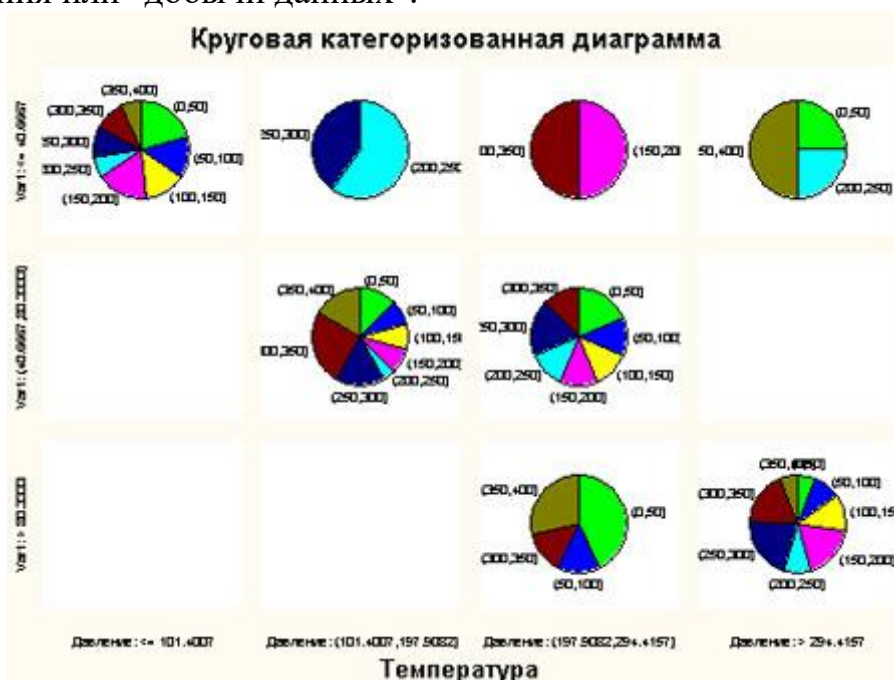














Рис. 9. Категоризованные круговые диаграммы.

Виды категоризованных графиков:

-  2М Категоризированные гистограммы
-  Графики вероятность – вероятность
-  2М Категоризированные диаграммы рассеяния
-  Категоризированные линейные графики
-  Категоризированные графики средних с ошибками
-  Категоризированные круговые диаграммы
-  Категоризированные диаграммы размаха
-  Категоризированные графики пропущенных данных/диапазонов
-  Категоризированные нормальные вероятностные графики
-  3М Категоризированные графики
-  Категоризированные графики квантиль – квантиль
-  Категоризированные тернарные графики

Расширение графических возможностей с помощью *STATISTICA Visual Basic*

STATISTICA Visual Basic (SVB) - язык программирования, интегрированный в *STATISTICA*, позволяющий получать доступ практически ко всем функциональным возможностям пакета программными средствами. С помощью SVB вы можете создавать свои вычислительные процедуры, разрабатывать собственные графики и статистические функции.

Многомерные дихотомии

Многомерные дихотомии (греч. διχο - на две части + τομή - сечение) часто возникают в практических задачах: в маркетинговых исследованиях, в анализе анкетных данных, данных соцопросов, медицине (наличие у пациента тех или иных признаков), контроле качества и т.д.

Дихотомия кодирует нулем и единицей ответы типа «нет» и «да».

В многомерной дихотомии имеется набор нулей и единиц, соответствующий многим вопросам и ответам.

Например, предположим, что мы проводим исследование-опрос предпочтений покупателей в супермаркете. В анкету мы включаем вопросы, на которые можно отвечать либо «да», либо «нет». В итоге, после проведения опроса, всех покупателей можно разделить на группы, согласно набору ответов на вопросы.

Если, например, вопросов в анкете было всего два (двумерная дихотомия), то всего может получиться 4 группы покупателей («да», «да»; «да», «нет»; «нет», «да»; «нет», «нет»).

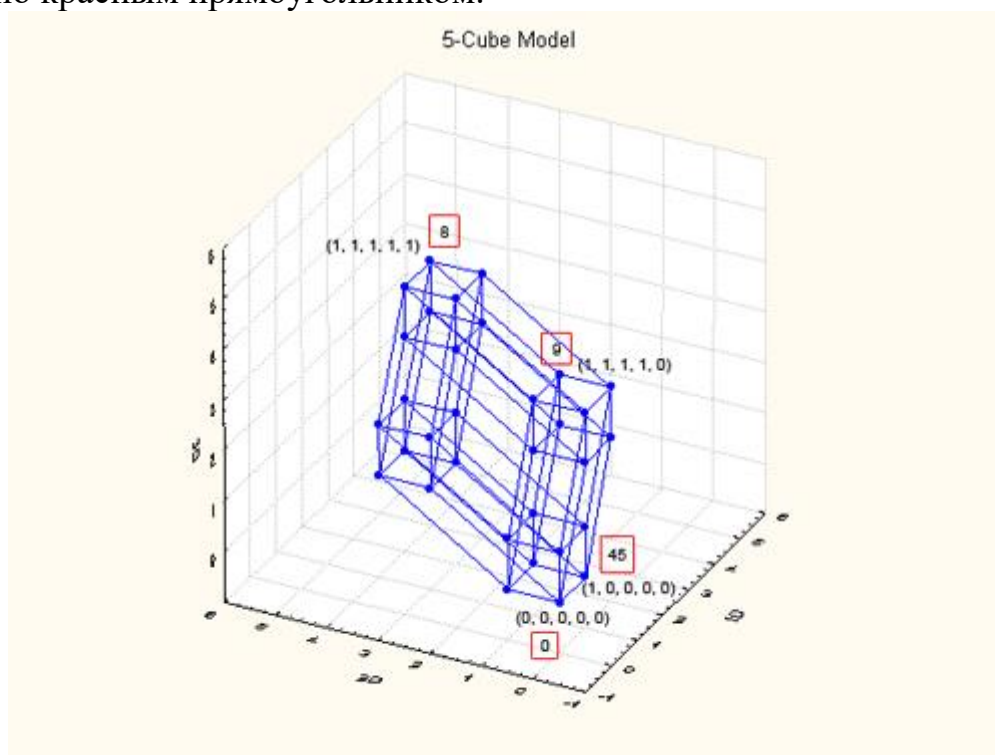
Подобное представление данных довольно понятно для малых размерностей дихотомий. Однако если количество вопросов в анкете возрастет до пяти или больше, то целостное представление о структуре данных начинает пропадать.

Поэтому важно уметь визуально представить многомерные дихотомийные данные.

Увидеть многомерные данные сложно - один из способов - с помощью многомерных кубов.

На рисунке ниже приведена модель пятимерного куба, описывающая опрос с анкетой, состоящей из пяти вопросов. Каждая группа опрошенных представляется в виде вершины с бинарным кодом (самая нижняя вершина имеет код (0,0,0,0,0), а самая верхняя - (1,1,1,1,1)).

Здесь «0» (или «1») обозначают ответ «нет» (или «да») на каждый из пяти вопросов. В метках вершин рядом с кодом группы, указано число, соответствующее числу людей относящихся к данной группе. На рисунке оно обведено красным прямоугольником.



3. Цель деятельности студентов на занятии:

Студент должен знать:

4. Почему специалисту-медику необходимы навыки проведения статистических исследований в своей работе.
5. Основную структуру пакета Statistica 6.0.
6. Графические возможности системы Statistica 6.0.

Студент должен уметь:

1. Владеть базовыми средствами для создания, редактирования, форматирования таблиц в пакете Statistica 6.0. В необходимой мере владеть навыком по осуществлению статистической обработки данных с применением современных

программно-технических средств, в частности, с помощью возможностей данного пакета.

2. Эффективно использовать графические возможности пакета Statistica 6.0 для более глубокого статистического исследования, с показателями корреляции, регрессии и описательной статистики.

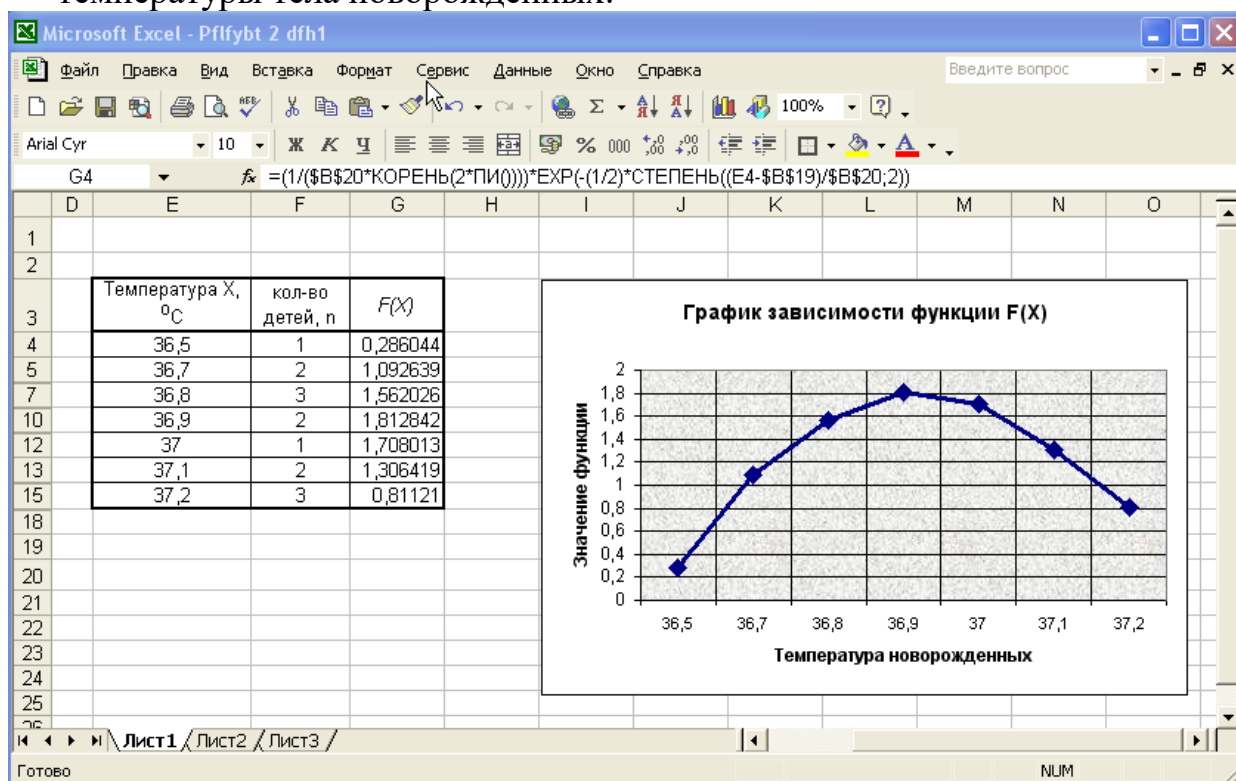
4. Содержание обучения:

Теоретическая часть:

1. Формирование таблиц. Ввод числовых и текстовых данных;
2. Использование графических возможностей пакета.

Практическая часть:

Построить сгруппированный вариационный ряд, вычислить для каждого значения признака (температуры новорожденных) значение функции, (надо заметить, что в рассматриваемом примере ячейка B19 содержит значение среднего по совокупности, а в ячейка B20 – значение стандартного отклонение). Построить график зависимости функции от температуры тела новорожденных:



4. **Вывод:** Данное распределение признака (температуры новорожденных) является нормальным, все значения распределены симметрично относительно среднего на одно и два стандартных отклонения. Можно сказать, что новорожденных с низкой и высокой температурой намного меньше, чем с нормальной, и в основном у новорожденных нормальная температура.

5. Перечень вопросов для проверки исходного уровня знаний:

1. Элементы организации медико-статистического исследования.
2. Статистическая величина.
3. Подготовка данных к статистическому анализу.

6. Перечень вопросов для проверки конечного уровня знаний:

1. Как построить графики в пакете Statistica 6.0?
2. Виды категоризованных графиков.
3. Пиктографики.
4. Матричные графики.

7. Хронокарта учебного занятия:

1. Организационный момент – 5 мин.
2. Текущий контроль знаний – 30 мин.
3. Разбор темы – 20 мин.
4. Практическая работа – 30 мин.
5. Подведение итогов занятия – 10 мин.

8. Самостоятельная работа студентов.

Разобрать один из видов графики в Statistice «Японские свечи».

9. Перечень учебной литературы к занятию:

1. Кобринский Б.А., Зарубина Т.В.
«Медицинская информатика», М., Издательский дом «Академия», 2009.
3. Жижин К.С. «Медицинская статистика», Высшее образование, 2007.
4. Лекция по данной теме.

Тема 7-8: «Оптимизация диагностики. МИС диагностических служб»

1. Научно-методическое обоснование темы:

Диагностика заболеваний, особенно на ранних стадиях развития или при отсутствии симптомов заболевания, имеет важное значение для профилактики, диагностики и лечения. «Тяжелую болезнь в начале легче вылечить, но трудно распознать. Когда же она усиливается, ее легче распознать, но уже труднее вылечить» (Н. Макиавелли).

Для более быстрого, безошибочного, определения заболевания используются новейшие разработки, «Информационные медицинские системы диагностических служб».

Система диагностической службы предназначена для автоматизации работы сотрудников диагностических кабинетов, упрощения формирования заключений, архивированию изображений и моментального доступа к архивам исследований.

КМИС поддерживает широкий спектр видов диагностики, включая рентгенографию, рентгеноскопию, флюорографию, томографию, ультразвуковые исследования, функциональную и эндоскопическую диагностику и т.д.

2. Краткая теория:

Комплексная медицинская информационная система

В подсистеме диагностических служб КМИС (Комплексная медицинская информационная система) предусмотрены следующие основные документы:

- Функциональная диагностика, включая ЭКГ, ВЭМ, суточное мониторирование ЭКГ, спирометрию;
- Нейрофункциональная диагностика;
- Ультразвуковая диагностика;
- Эндоскопическая диагностика;
- Рентгенологическая диагностика, включая рентгенографию, флюорографию, рентгеноскопию, томографию, маммографию и т.д.

При создании нового документа система автоматически помещает в него паспортную информацию о пациенте и дату выполнения. В любом документе осуществляется хранение описания исследования, заключения и статистический блок для кодирования вида исследования, в том числе учет условных единиц труда (УЕТ). Структура документов универсальна и позволяет системе автоматически цитировать заключения или описания во вторичных документах, например - эпикризах, выписках и т.д. Все

документы диагностической службы подключены к подсистеме статистики, за счет чего анализ нагрузки или другие отчеты формируются в единой программе статистики очень просто и быстро. Предусмотрена возможность составления собственных отчетов. Любой документ диагностической службы может быть распечатан или отправлен по электронной почте. Для исключения модификации ответов после их утверждения используются возможности подсистемы безопасности - электронная цифровая подпись и ограничение доступа.

1 Печать 2 Закрыть 3 Открыть 4 Архивирование 5 Шаблон 6 Диаграмма 7

ТЕСТИРОВАНИЕ ВЛАДИМИР ИВАНОВИЧ (29.11.1950) Возраст 55
ВЭМ (06.06.2006) [Показать сведения](#)

Дата выполнения 06.06.2006

Цель направления Ишемическая болезнь сердца
 Каким ЛПУ направлен ПОЛИКЛИНИКА №777

Исследование Вид: ВЭМ, Название: Эргоспирометрия
 Выполнено в ЛПУ - Демонстрационное ЛПУ
 Код (ФИО) врача Прохоров Прохор Прохорович - врач клинической лабораторной диагностики /
 Условные единицы труда врача - 12, медсестры - 12
 Количество исследований 1
 Стоимость исследования 0
 Статус патологии: Вновь выявленная Подтвержденная Норма
 Номер исследования 323

Демонстрационное ЛПУ
ВЭМ № 323

Первично
 Диагноз направления: ИБС
 АД 150 / 100

Описание
 Выполнена 3 ступенчатая проба на ВЭМ продолжительностью 9 минут, максимальной мощностью 100 ватт и проделанной работе 4050 кгм.
 Проба прекращена из-за усталости, отказ больного от проведения пробы при достижении ЧСС 118 в минуту (должная ЧСС 154 в минуту), АД 199 / 90 мм рт. ст.
 Особенности Регулярная терапия ББ, иАПФ, аспирин, симвастатин. Исходно АД 150/100.Редкая НЭС начиная с нагрузки 75 вт с исчезновением в ВП. Неспецифическая депрессия ST в отведениях II, III, aVF, v5 0,6 - 0,8 мм на высоте нагрузки.

Заключение:
 Проба неинформативна при средней толерантности к физической нагрузке и коротком (5 мин) восстановительном периоде. Ишемические изменения не выявлены.

Подпись: Прохоров П.П.

Для этого документа сохранено связанное архивное изображение или видео-файл
 Для того, чтобы открыть архив, воспользуйтесь этой ссылкой:

Электронная цифровая подпись. Документ подписан и не может быть изменен.
 Документ подписал(а): ЭЦП тестового пользователя
 Дата и время подписи: 08.06.2006 / 10:20

История изменения

Пример ультразвукового ВЭМ-исследования. Электронный документ диагностической службы содержит описательную часть, заключение, раздел статистических показателей, электронно-цифровую подпись.

Т.к. основная информация хранится в формализованных бланках, предоставляется возможность построения графиков динамики изменения определенных показателей пациента. Информация из подсистемы диагностических служб доступна для всех основных программ системы, в том числе создание эпикризов или выписок, профосмотра, диспансерного наблюдения и т.д.

Журнал отделения функциональной диагностики F1-Помощь

Основной список
Неподписанные
Функциональная
Сегодня!
По датам
Нагрузка
Нейрофизиологическая
Сегодня!
По датам
Нагрузка
Панель управления

1 Изменить документ 2 Открыть БД 3 Открыть 4 Экспорт в Excel 5

Год	ЭЦП	Арх.	Название исследования	Док-тов	Ис-ний	У.Е.Т
2004				1	1	1.70
12 Декабрь				1	1	1.70
Прохоров Прохор Прохорович - врач клинической лабораторной				1	1	1.70
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)			ЭКГ в покое стандартная 12 отведений (07.12.2004)	1	1	1.70
2006				1	1	1.70
03 Март				1	1	1.70
Прохоров Прохор Прохорович - врач клинической лабораторной				1	1	1.70
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)			ЭКГ в покое стандартная 12 отведений (14.03.2006)	1	1	1.70
2008				3	3	18.70
10 Октябрь				1	1	6.80
Харламов Харламп Харламович - врач функциональной диагностики				1	1	6.80
БОРИСОВ БОРИС БОРИСОВИЧ (25.12.1907)			Проба с калием (01.10.2008)	1	1	6.80
11 Ноябрь				1	1	5.10
Харламов Харламп Харламович - врач функциональной диагностики				1	1	5.10
ТЕСТОВАЯ МАРЬЯ ИВАНОВНА (01.01.2006)			Проба с другими лекарственными препаратами (25.11.2008 - 08:56)	1	1	5.10
12 Декабрь				1	1	6.80
Харламов Харламп Харламович - врач функциональной диагностики				1	1	6.80
ТЕСТОВАЯ МАРЬЯ ИВАНОВНА (01.01.2006)			Многократная ЭКГ в палате (11.12.2008 - 11:28)	1	1	6.80
2009				1	1	1.70
03 Март				1	1	1.70
Егоров Егор Егорович - врач УЗИ				1	1	1.70
ТЕСТОВАЯ МАРЬЯ ИВАНОВНА (01.01.2006)			ЭКГ в покое стандартная 12 отведений (20.03.2009)	1	1	1.70
				6	6	23.80

Лучевая диаг-ка Эндоскопическая диаг-ка
Версия 3.3.0.0

Д
ля
архи
ви
вани
я
резул
тато
в
иссл
едов
аний
разра
бота
на
спец
иаль

ная подсистема архивирования, при помощи которой в специальных базах данных могут быть сохранены результаты исследования в различных форматах:

- Графические изображения в наиболее распространенных форматах (JPEG, GIF, TIFF, BMP, DICOM и т.д.);
- Видеофайлы (AVI, MPEG, ASF и т.д.);
- Аудиофайлы (WAV, MP3).

С целью снижения нагрузки на информационную сеть возможно отдельное хранение графических или видеоархивов, в т.ч. на отдельном сервере. При этом переключение пользователя из основной БД к выделенной БД архива на другом сервере осуществляется автоматически при помощи гиперссылок.

Автоматизация диагностической службы позволяет врачу сосредоточиться на выполнении и оценке исследований, упростить формирование заключений и архивирование результатов исследований, получать быстрый доступ к архивам выполненных ранее исследований. Подключение к финансово-экономической системе и подсистеме статистики позволяют получать всю необходимую информацию о нагрузке, учитывать платные услуги и т.д.

КМИС автоматически формирует журналы для различных кабинетов диагностической службы, в том числе журнал отделения функциональной диагностики, журнал отделения лучевой диагностики и журнал эндоскопической диагностики. Эти журналы позволяют быстро получить необходимые сведения о выполненных обследованиях, найти неподписанные протоколы, получить сведения о выполнении функции врачебной должности.

С целью дополнительно расширения возможностей КМИС в части автоматизации диагностических подразделений, в составе системы имеются 2 дополнительных решения:

- *Радиологическая информационная система (РИС)* с возможностью интеграции с PACS-системами
- *Лабораторная информационная система (ЛИС)* с возможностью подключения лабораторных анализаторов.

Радиологическая информационная система (РИС)

«Радиологическая информационная система КМИС» (РИС) – это специальный набор программного обеспечения внутри КМИС, позволяющий эффективно автоматизировать работу подразделений лучевой диагностики ЛПУ в рамках внедрения единой электронной медицинской карты (ЭМК) пациента.

Основная задача РИС - поддержка деятельности специалистов отделения лучевой диагностики и медицинской визуализации. РИС позволяет эффективно и очень быстро автоматизировать ввод, хранение и обработку протоколов диагностического обследования, а также обеспечивает возможность просмотра изображений из электронной медицинской карты в любом рабочем месте врача. При использовании совместной работы РИС и PACS обеспечивается интеграция между КМИС и диагностическим оборудованием по протоколу DICOM, а также хранение и обработка медицинских изображений, включая 3D-реконструкцию и удаленный web-доступ к архивам медицинских изображений.

Функциональные возможности

«Радиологическая информационная система КМИС» поддерживает следующие функциональные возможности:

- Регистрация пациентов (в подсистеме регистратуры и приемного отделения)
- Запись пациентов на исследования (через электронные календари КМИС)
- Автоматизированное формирование протоколов исследования с использованием гибко настраиваемого справочника шаблонов
- Поддержка различных печатных форм, включая возможность создавать собственные печатные формы ЛПУ
- Встроенный учет услуг, включая учет форм оплаты (ОМС, ДМС, договора, различные бюджеты и т.д.), стоимости и себестоимости, УЕТ, исполнителей и места выполнения исследования и т.д.

- Учет лучевой нагрузки, включая учет поглощенной дозы за 1 снимок и автоматический расчет суммарной поглощенной дозы за исследование.
- Автоматизированное ведение единого листа лучевой нагрузки пациента, куда скапливается вся информация о полученной в ходе обследования и лечения пациента дозы облучения.
- Автоматическое формирование журнала работы отделения лучевой диагностики с формированием оперативной информации о выполняемых исследованиях, расчете нагрузки на отделение лучевой диагностики в различных периодах, автоматическом формировании журнала записи рентгенологических исследований за период (форма №050/у, утвержденная приказом Минздрава СССР №1030 от 04.10.1980 г.).
- Возможность учета контрольного обследования, которое необходимо выполнить пациенту
- Учет выявленной патологии
- Учет применяемого контраста
- Автоматическое ведение электронной флюоротеке с функцией автоматизированного планирования необходимого флюорографического обследования прикрепленного контингента, составления планов по флюоротеке и т.д.
- Встроенная функция архивирования изображений, аудио- и видеофайлов
- Встроенная возможность интеграции с PACS-серверами (поддерживается FUJIFILM Synapse PACS или Siemens PACS на выбор, возможна заказная интеграция с другими решениями).
- Интеграция с подсистемой «Статистика КМИС» и «Финансово-экономической системой КМИС» для централизованного формирования всей необходимой статистической информации и взаиморасчетам за выполненные исследования.
- Возможность создания собственных учетных форм, включая сложные формализованные бланки исследований с аналитической обработкой и расчетными параметрами.
- Возможность формирования собственных статистических отчетов произвольной сложности и вида
- Интеграция с внешними информационными системами для передачи информации из протоколов исследований в другие БД
- Экспорт медицинских протоколов в другие МИС на основе CDA (стандарт HL7)

- Интеграция с системой информационного обмена КМИС.РИР для обмена медицинскими документами с другими ЛПУ
- Полностью кроссплатформенная реализация, включая работу клиентской части в Microsoft Windows, Linux и Mac OS X.

Описание решения

Основа работы РИС – это ведение протокола диагностического обследования. Протоколы создаются непосредственно в электронной медицинской карте (ЭМК) пациента и содержат следующую информацию:

- Статистический учетный блок, включая кодирование выполненной услуги, исполнителя, условия направления на исследование и т.д.
- Описательная часть, включая описание и заключение, данные о выявленной патологии, информация о применяемом контрастном веществе, полученной лучевой нагрузке и т.д.
 - Блок архивирования для хранения изображений или интеграции с хранилищем медицинских изображений на PACS-сервере.

Печать | закрыть | Открыть | Архивирование | журнал лучевой диагностики

Компьютерная томография брюшной полости (19.07.2012)
 ТЕСТ АННА ИВАНОВНА (16.12.1955) Возраст 56 лет

Оказанные медицинские услуги по данному документу

Дата выполнения: 19.07.2012
 Статус патологии: вновь выявленная подтвержденная Норма

ГБУЗ Центральная районная больница №10

Описание:
 1-е исследование: Ксенентикс 300-100,0 в/венно, болюс.
 На серии КТ-томограмм получены изображения органов брюшной полости и забрюшинного пространства. Печень несколько увеличена в размерах за счет правой доли (КВР 18,7см), контуры ровные, четкие, плотность паренхимы 62-65ед Н. Внутривенные желчные протоки и сосуды не расширены. В проекции VII сегмента печени, в субкапсулярных отделах определяется единичный кальцификат диаметром 0,3см. Желчный пузырь не увеличен. Рентгеноконтрастных конкрементов в его просвете не выявлено. Селезенка обычной формы, величины и положения, однородной структуры. Поджелудочная железа не увеличена, дольчатой структуры. Паранкретическая клетчатка не изменена. Оба надпочечника расположены обычно, структурно не изменены. Почки обычной формы, величины и положения. Паренхима обеих почек без особенностей, ЧЛС не изменена. Забрюшинные лимфатические узлы не увеличены. Аорта и нижняя полая вена не изменены. Дегенеративно-дистрофические изменения позвоночника на уровне сканирования.

Заключение:
 Без патологических изменений

Связь с флюороотеккой	<input type="checkbox"/> Учет в флюороотекке
Количество снимков	1
Контраст	<input checked="" type="checkbox"/> Использовался контраст
Способ введения контраста	Внутривенно
Вид примененного контраста	Омнипак 300
Количество введенного препарата	100

Подпись: Аюляян А.А.

Для этого документа сохранено связанное архивное изображение или видео-файл
 Для того, чтобы открыть архив, воспользуйтесь этой ссылкой:
 Номер архивного диска: 02566

Документ подписан электронной цифровой подписью и не может быть изменен.
 Документ подписал(а): DevTeam
 Дата и время подписи: 19.07.2012 / 11:12

Протокол диагностического исследования в радиологической информационной системе КМИС

В составе системы предусмотрены гибкие справочники, с помощью которых можно настроить РИС под любые индивидуальные особенности работы ЛПУ:

- Справочник видов исследований
- Справочник сотрудников диагностической службы (врачи, средний медперсонал)
- Причины направлений
- Справочник выявленной патологии
- Дополнительные манипуляции, выполняемые во время исследования (апликации, коагуляции и т.д.)
- Медицинское оборудование, применяемое для исследований
- Контрастные вещества
- Виды анестезии
- Виды доступов
- Радиофармпрепараты
- Справочник типовых шаблонов для быстрого формирования протокола КМИС поставляется с уже встроенным справочником исследований и типовых протоколов для следующих видов лучевой диагностики:
- Ангиография (включая различные ее виды, коронарографию и т.д.)
- Компьютерная томография
- Магнитно-резонансная томография
- Маммография
- Рентгеновская томография
- Рентгенография
- Рентгеноскопия
- Спиральная компьютерная томография
- Флюорография
- Ультразвуковая диагностика (более 50 различных видов исследований)

Все протоколы исследований автоматически собираются в удобном «Журнале отделения лучевой диагностики», с помощью которого удобно

систематизировать информацию о выполненных исследованиях, получать оперативные сводки о нагрузке и т.д.

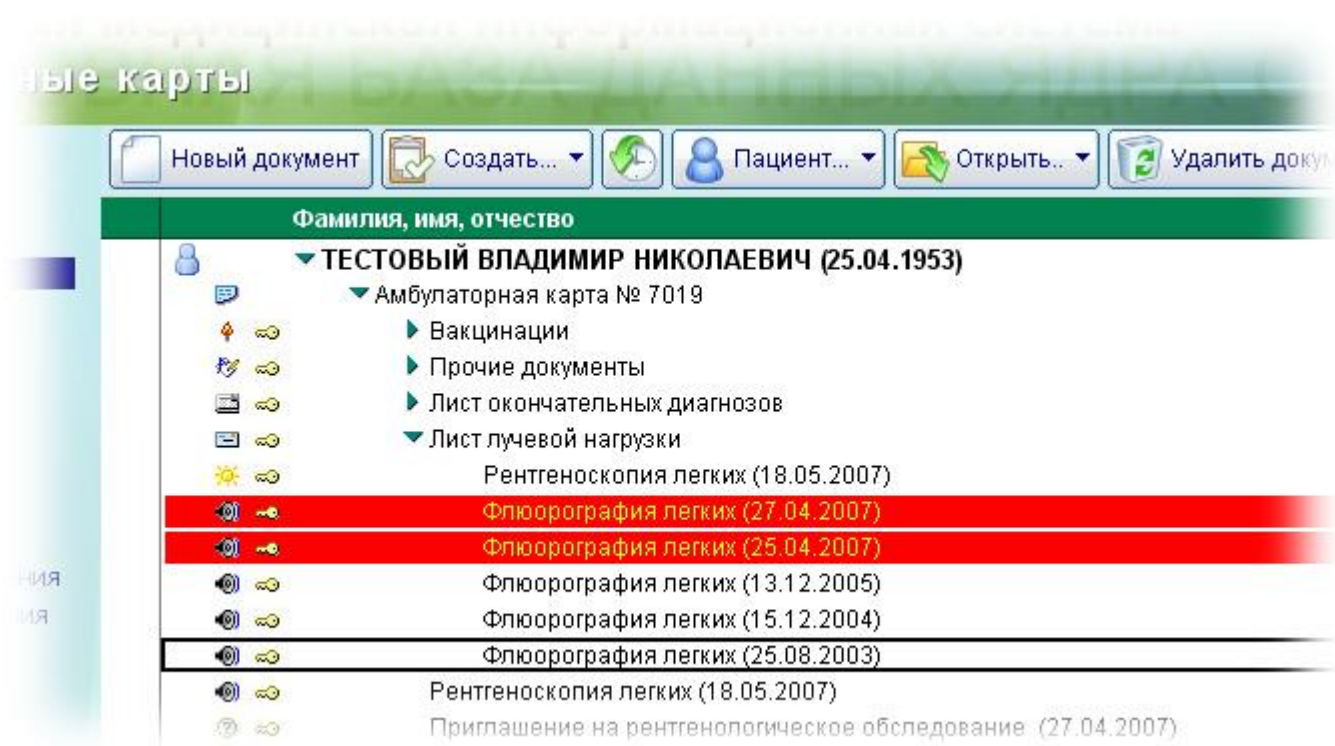
ФИО пациента	Дата	Название исследования	ЭЦП	Арх.	Снимков	Доза
БОРИСОВ БОРИС БОРИСОВИЧ (25.12.1907)	09.01.2008	Флюорография легких			1	0,8
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)	15.07.2003	Флюорография легких			1	0,8
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)	07.12.2004	Флюорография легких			1	0,8
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)	16.03.2006	Флюорография легких			1	0,8
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)	23.01.2007	Флюорография легких			1	0,6
ДЛЯТЕСТОВ ВИКТОР АЛЕКСАНДРОВИЧ (17.02.1952)	12.03.2007	Флюорография легких			1	0,6
ПРОВЕРКИТЕСТОВ ВЛАДИМИР ВИКТОРОВИЧ (02.09.1953)	24.12.2003	Флюорография легких			1	0,8
ПРОВЕРКИТЕСТОВ ВЛАДИМИР ВИКТОРОВИЧ (02.09.1953)	07.09.2005	Флюорография легких			1	0,8
СИДОРОВ СИДОР СИДОРОВИЧ (22.11.1977)	23.04.2008	Рентгенография живота			1	0,3
СИДОРОВ СИДОР СИДОРОВИЧ (22.11.1977)	30.04.2008	Флюорография легких			1	0,8
ТЕСТ ИВАН АДРЕЕВИЧ (11.10.1978)	20.10.2003	Флюорография легких			1	0,8
ТЕСТ ИВАН АДРЕЕВИЧ (11.10.1978)	20.04.2005	Флюорография легких			1	0,8
ТЕСТОВАЯ ГАЛИНА ВЛАДИМИРОВНА (27.11.1958)	17.12.2003	Флюорография легких			1	0,8
ТЕСТОВАЯ ГАЛИНА ВЛАДИМИРОВНА (27.11.1958)	21.02.2006	Флюорография легких			1	0,8
ТЕСТОВАЯ МАРЬЯ ИВАНОВНА (01.01.2006)	24.10.2008	Флюорография легких			1	0,8

Работа врача-рентгенолога значительно упрощается за счет средств автоматизации, предусмотренных в КМИС

Администраторы системы могут не только гибко настраивать справочники РИС, но и создавать собственные учетные и отчетные формы самого произвольного содержания и сложности, свободно интегрировать протоколы обследования с внешними информационными системами или наоборот, импортировать в КМИС результаты обследования из унаследованных информационных систем.

РИС поддерживает автоматическое ведение в электронной амбулаторной карте пациента «Листа лучевой нагрузки», к которому подключены все рентгенологические документы - флюорография, рентгенография, рентгеноскопия, маммография и др. В автоматическом режиме накапливается информация о лучевой нагрузке на пациента в течение его жизни, данные об исследованиях за год, отметки о выполненной флюорографии, маммографии. Этот раздел позволяет быстро и наглядно

получить список всех рентгенологических исследований, выполненных пациенту в течение жизни. При этом автоматически рассчитывается суммарная лучевая нагрузка на пациента в течение года. Применение электронной флюоротеки позволяет врачу на участке полностью отказаться от бумажной, значительно сократить время, затрачиваемой медсестрой на составление списков должников. Удобный раздел «Лист лучевой нагрузки» позволяет быстро получить представление обо всех рентгенологических исследованиях пациента.



Лист лучевой нагрузки в электронной амбулаторной карте пациента

Для автоматического составления планов по флюорографическому обследованию в "Паспорт поликлиники" встроено специальное приложение "Флюоротека". В нем имеется специальная команда, которая позволяет всего несколькими щелчками мыши сформировать готовый план по флюорографии, выявить пациентов, не выполнивших исследование вовремя. При этом имеется возможность гибкого составления списка пациентов, для которых система сформирует план обследования. Сформированный список можно просмотреть на экране, распечатать или отправить по электронной почте.

Паспорт поликлиники КМИС Информационная система

Настройка F1-Помощь

Открыть БД Открыть Экспорт в Excel Флюоротека Удалить документ

Фамилия, имя, отчество	Дата проведения флюорографии	Плановое проведение флюорографии	Место работы	Участок	Кол - во
Флюорография сделана 6					
ВРАЧ ЛЕЧЕБНОГО УЧРЕЖДЕНИЯ г. МОСКВЫ 1					
СИДОРОВ ИВАН ИВАНОВИЧ (17.02.1952)	12.03.2007	12.03.2008	Металлургический завод ОАО - Транспортный / МАШИНИСТ	Участок №1	1
Евгеньева Евгения Евгеньевна - врач-педиатр 1					
ЗЫКОВА АННА АНДРЕЕВНА (30.10.1989)	10.01.2012	10.01.2013	ОАО АВГУСТ - МЕХАНИЧЕСКИЙ ЦЕХ / СЛЕСАРЬ-РЕМОНТНИК	Участок №5	1
Петров Петр Петрович - врач-терапевт участковый 4					
БОРИСОВ БОРИС БОРИСОВИЧ (25.12.1907)	09.01.2008	09.01.2009		Участок №4	4
Флюорография не сделана 26					
Акопян Акол Акопович - врач-эндоскопист 5					
0000 ИВАН ИВАНОВИЧ (23.12.2004)				Участок №1	5
222 РЕВЕНКО ИВАНОВИЧ (29.08.2011)				Участок №1	
2222 ИВАН ИВАНОВИЧ (31.05.2011)			Металлургический завод ОАО	Участок не указан	
2222 ИВАН ИВАНОВИЧ (31.05.2011)			Металлургический завод ОАО	Участок не указан	
ПЕСТЕЛЬ ИВАН ИВАНОВИЧ (22.04.1956)			ОАО САН ИНБЕВ / ВЕДУЩИЙ ИНЖЕНЕР	Терапевтический участок №1	
Богданов Богдан Богданович - зав. травматологическим отделением 2					
111 ИВАН ИВАНОВИЧ (01.01.1900)				Участок не указан	2

Стационар Статистика
Версия 3.3.0.10

Работа с приложением электронной флюоротеки

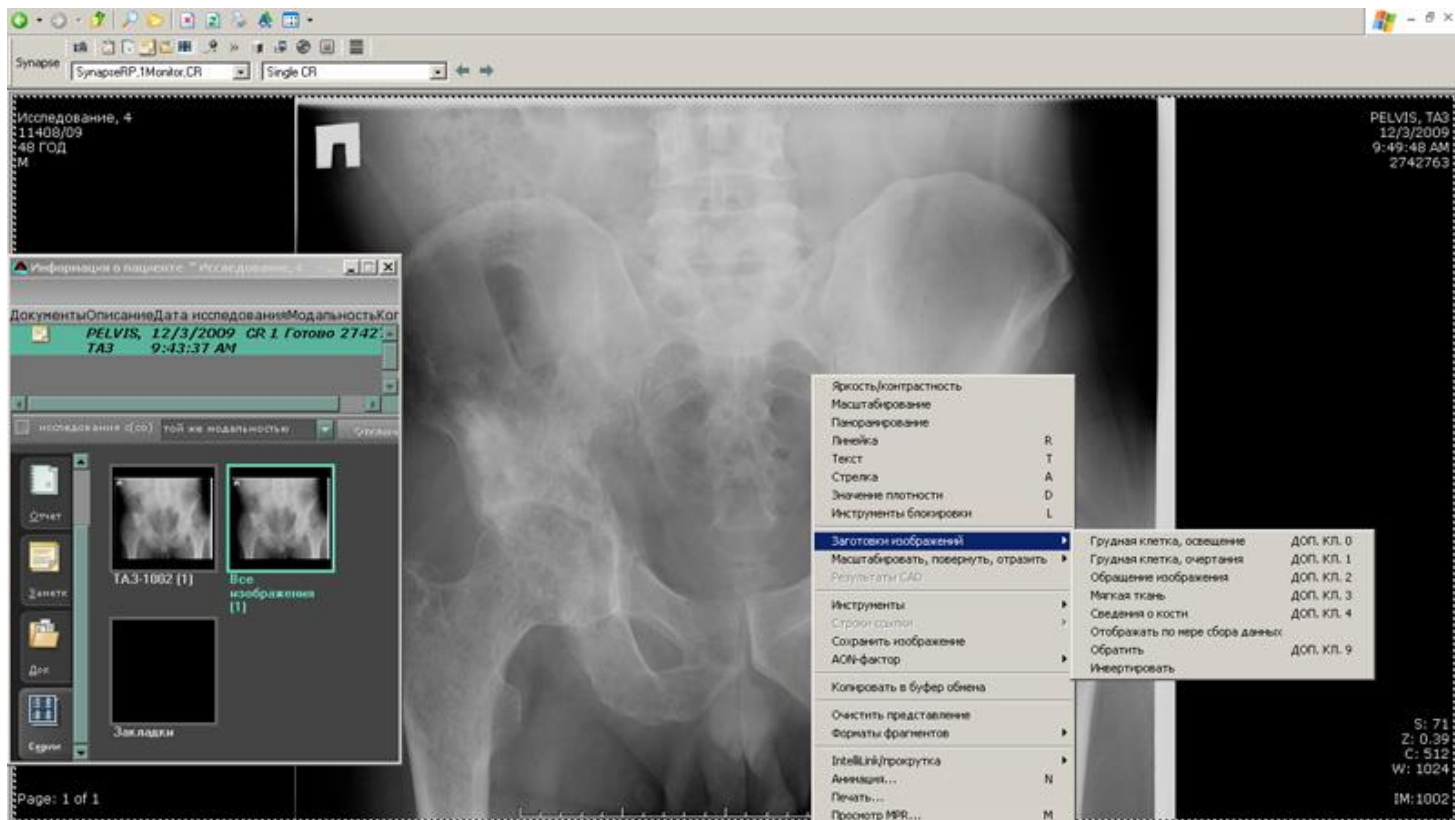
Все рентгенологические документы подключены к подсистеме статистики, что позволяет формировать гибкие и настраиваемые отчеты по работе рентгенологической службы, в том числе создавать отчеты по нагрузке (в том числе лучевой нагрузке), автоматизировать работу по составлению отчетов ДОЗ-3 и т.д.

Для архивирования результатов исследований возможно применение одного различных сценарием:

- Использование встроенной функцией архивирования файлов, при помощи которой в специальных базах данных могут быть сохранены результаты исследования в любых форматах, в том числе как графические изображения (JPEG, GIF, TIFF, BMP и т.д.), видеофайлы (AVI, MPEG, ASF и т.д.), аудиофайлы (WAV, MP3) и т.д.
- Интеграция со специализированным PACS-сервером (в качестве типового решения компания К-МИС применяет FUJIFILM Synapse PACS, но также поддерживается и интеграция с Siemens PACS).

В случае использования встроенной функции архивирования с целью снижения нагрузки на информационную сеть предусмотрено отдельное хранение графических или видеоархивов в выделенных БД, в т.ч. на отдельном сервере. При этом переключение пользователя из основной БД к выделенной БД архива на другом сервере осуществляется автоматически при помощи гиперссылок.

Интеграция со специализированными системами архивирования, передачи и обработки изображений (PACS) позволяет врачам и другим пользователям КМИС, не работающим в диагностическом подразделении, получать доступ к архивным изображениям пациента.



Работа с архивом медицинских изображений при использовании радиологической информационной системы КМИС совместно с FUJIFILM Synapse PACS

Модуль интеграции с PACS позволяет автоматически открывать и обрабатывать изображения прямо из протоколов обследований в КМИС. Для этого пользователю достаточно просто щелкнуть по значку архива в любом протоколе обследования пациента, который получен с использованием медицинского оборудования (томография, ультразвуковые исследования, рентген и т.д.).

Как правило, на рабочей станции пользователей КМИС никакого дополнительного программного обеспечения устанавливать не требуется - вся работа организована с использованием "тонкого клиента" (на базе Интернет-браузера). У пользователей появляется возможность просмотреть результаты конкретного исследования или просмотреть все исследования, проведенные за какой-либо промежуток времени, что может значительно улучшить качество диагностики. Прямо на рабочем месте можно проводить

реконструкцию изображений и делать подсчеты с помощью встроенных инструментов, что может помочь, например, при планировании операций.

Так как КМИС поддерживает интеграцию с различными PACS-решениями, данный функционал зависит от конкретной модели PACS-системы, выбранной в ЛПУ.

Детальное описание возможностей FUJIFILM Synapse PACS, который является базовым и рекомендованным для применения в проектах КМИС, представлено на странице <http://www.kmis.ru/site.nsf/pages/SynapsePACS.htm>

При использовании решения Siemens Syngo Imaging XS доступны следующие функции:

- Поддержка интеграции с медицинским оборудованием посредством стандарта DICOM 3.0
- Предоставление MPR, MIP, MinIP, SSD, VRT
- Сохранение вращающейся трехмерной модели в виде видеоролика
- Количественная оценка ангиографических КТ- и МРТ-исследований с контрастированием
- MPR-эндоскопия: визуализация полых структур
- Просмотр и обработка изображений, полученных методом радионуклидной диагностики
- Маркировка позвоночника, подсчет кардиоторакального индекса, просмотр данных цифровой субтракционной ангиографии
- Встроенное решение для маммографии с поддержкой отчетности
- Запрос и получение изображения из нескольких источников одновременно
- Запись результатов на CD/DVD и Печать изображений
- Использование нескольких мониторов
- Простой русскоязычный интерфейс и некоторые другие

Автоматизация службы лучевой диагностики ЛПУ с помощью радиологической информационной системы КМИС позволяет врачу действительно сосредоточиться на выполнении и оценке исследований, упростить формирование заключений и архивирование результатов исследований, получать быстрый доступ к архивам выполненных ранее

исследований. Подключение к финансово-экономической системе и подсистеме статистики позволяют получать всю необходимую информацию о нагрузке, учитывать платные услуги и т.д. Все это вместе взятое позволяет максимально повысить эффективность работы соответствующих подразделений, внедрить новейшие методы обследования, что в свою очередь – самым благоприятным образом сказывается на качестве медицинской помощи, оказываемой в ЛПУ.

Лабораторная информационная система КМИС

Одно из первых подразделений, где возможен быстрый переход на работу в рамках медицинской информационной системы – это лаборатория. Функционально лаборатория связана с регистратурой, лечащими врачами и процедурным кабинетом, поэтому цель внедрения лабораторной информационной системы – оперативное назначение обследования, быстрое и правильное получение биологических материалов, выполнение исследований и передача результатов в медицинские документы. В тех случаях, когда при внедрении КМИС в лечебном учреждении есть определённый переходный период и приходится дублировать информацию на бумажных и электронных носителях, лаборатория может сразу перейти на электронный документооборот.

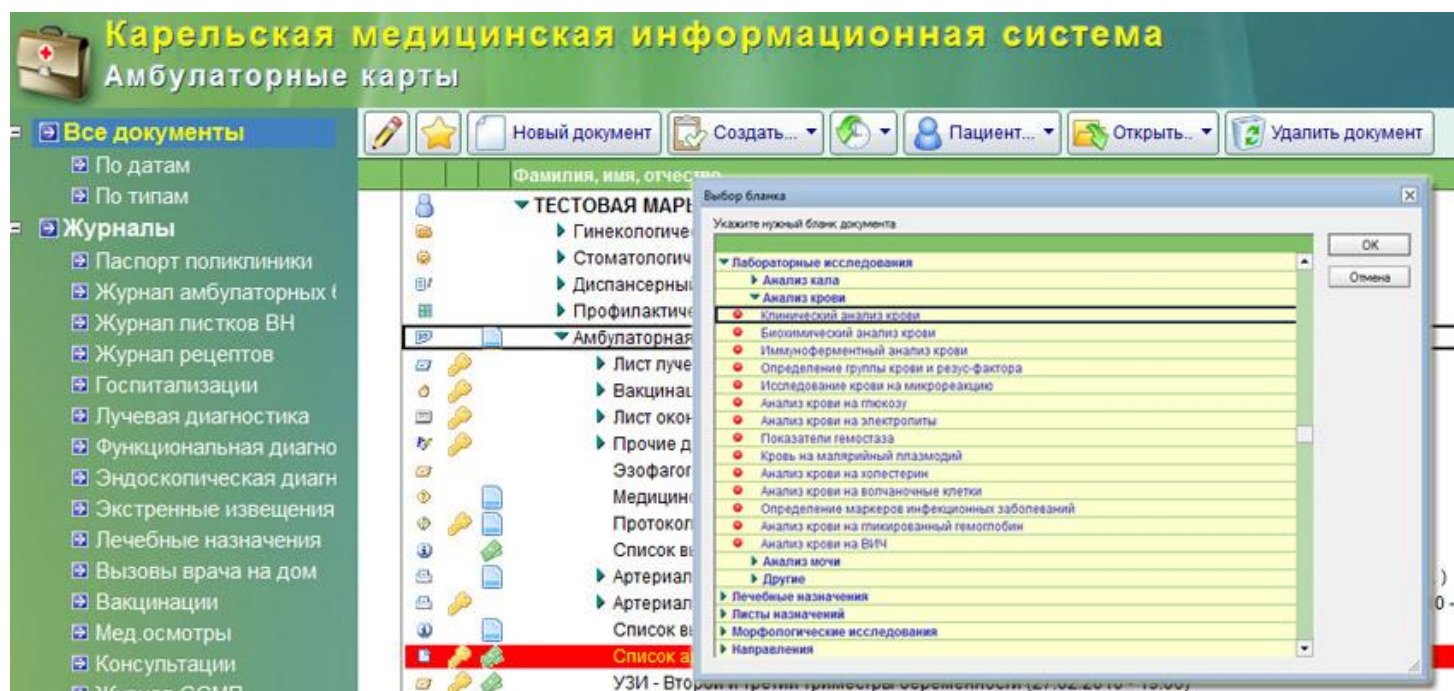
Лабораторная информационная система является одной из наиболее разработанных и широко используемых подсистем КМИС, предоставляя врачам до 80% диагностической информации. Основу лабораторной подсистемы составляет комплекс специального ПО – бланк лабораторного обследования, модуль заказа, лабораторная информационная система, встроенный модуль контроля качества, драйверы лабораторных анализаторов и т.д..

Выполнение лабораторных назначений

Работа пользователей с лабораторной подсистемой осуществляется по следующему алгоритму: врачи (или медсестры) со своих рабочих мест осуществляют заказ необходимого лабораторного исследования. Для этого:

- **Выбирается амбулаторная карта или история болезни пациента.**
- **Дается команда на создание нового назначения.** Система выводит доступный на текущий момент список назначений (с учетом прав доступа пользователя и настроек лабораторной подсистемы).
- **Выбирается нужный бланк** (в КМИС представлено свыше 35 различных бланков, насчитывающих в общей сложности свыше 200 параметров)

Создание бланка заказа лабораторного исследования



В текущей версии ЛИС поддерживаются следующие виды исследований и параметры:

- **Исследование биологических жидкостей**
 - Исследование мочи
 - Копрологическое исследование
 - Исследование кала
 - Исследование мокроты
 - Исследование эякулята
 - Исследование простатического содержимого
 - Исследование синовиальной жидкости
- **Гематологические исследования**
 - Клинический анализ крови
 - Анализ крови на гликированный гемоглобин
 - Анализ крови на метгемоглобин
 - Анализ крови на тельца Гейнца-Эрлиха
 - Анализ крови на карбоксигемоглобин
- **Цитологические исследования**

- Исследование гинекологического /урологического/ мазка
- Мазок мокроты на эозинофилы
- Мазок на энтеробиоз
- **Биохимические исследования**
 - Биохимический анализ крови
 - Анализ крови на глюкозу
 - Анализ крови на электролиты
 - Анализ крови на холестерин
 - Анализ крови на метгемоглобин
 - Анализ крови на холинэстеразу
 - Гамма-глутамилтранспептидаза (ГГТП)
 - Альфа-1-антитрипсин
 - Микроэлементы
- **Коагулогические исследования**
 - Исследование показателей гемостаза
 - Растворимые фибрин-мономерные комплексы (РФМК)
- **Иммунологические исследования**
 - Иммунологическое исследование
 - Определение группы крови и резус-фактора
 - Анализ крови на волчаночные клетки
 - Определение маркеров инфекционных заболеваний
 - Анализ мочи на микроальбуминурию
- **Микробиологические исследования**
 - Исследование крови на микрореакцию
 - Анализ крови на малярийный плазмодий
 - Анализ крови на ВИЧ
 - Анализ мочи на ВК
 - Анализ мазка на грибки

- Анализ мокроты на ВК
- Иммуносерологическое исследование биологического материала
- Мазок из зева и носа на наличие патогенного стафилококка
- Исследования на носительство возбудителей кишечных инфекций
- Микологические исследования
- **Токсикологические исследования**
 - Анализ мочи на содержание металлов
 - Анализ мочи на алкоголь
 - Анализ мочи на копропорфирин
 - Анализ мочи на ртуть

Работа с бланком лабораторного заказа

Система выводит на экран пустой бланк для заказа лабораторного исследования (например, бланк клинического анализа крови). В нем пользователь может выполнять следующие операции:

- **Выбирать нужные параметры.** При этом система выводит список только тех параметров, которые лаборатория в данный момент может выполнить. Если по каким-то причинам определенный параметр не может быть определен (нет реактивов, не работает анализатор), то врач-лаборант может отметить этот факт в системе со своего рабочего места и данный параметр будет скрыт в бланках назначений.
- **Корректировать дату исследования.** При этом система автоматически отслеживает установленные лимиты назначений на выбранную дату и подсказывает пользователю - сколько еще возможных назначений осталось на выбранный день
- **При необходимости выбирать лабораторию** или процедурный кабинет

Выполненное назначение сохраняется в базе данных. При сохранении система проверяет доступность выбранного назначения (с учетом установленных лимитов). Если назначение возможно, бланк заказа записывается в систему и передается на обработку в базу данных лаборатории. Если пользователь пытается выполнить назначение сверх установленного лимита, то система блокирует такое назначение и предлагает пользователю указать другую дату или направить бланк заказа в другую лабораторию.

Заказ исследования "Клинический анализ крови"

ФИО: **ТЕСТОВАЯ МАРЬЯ ИВАНОВНА** Возраст: 4
Дата рождения: 01.01.2006 Пол: ЖЕНЩИНА
Адрес: Карелия Респ, Кондопожский р-н, г Кондопога, ул. Коммунальная д.1 кв.1
Название назначаемого анализа: Клинический анализ крови

Cito Заказ на: 16.09.2010 Доступно: Неограниченно
 Выбрать все Результат распечатать Сделано: 1

Лечащий врач: Петров Петр Петрович - врач-терапевт участковый

Эритроциты
 Средний объем эритроцитов
 Ширина распределения эритроцитов по объему RDW
 Степень анизоцитоза эритроцитов
 Гематокрит
 Тромбоциты
 Тромбокрит
 Средний объем тромбоцитов
 Степень анизцитоза тромбоцитов
 Лейкоциты
 Гемоглобин
 Средний гемоглобин клетки
 Средняя концентрация гемоглобина в одном эритроците
 Цветовой показатель
 Ретикулоциты
 СОЭ
 Лейкоформула
 Свертываемость по Сухареву

Тип оплаты: ДМС
Страховая компания: Макс-М [Страховая программа-Стандартная программа]
Цель назначения: Цель по умолчанию

Отменить Назначить
Сохранить изменения Выход

Program version: 1.619

Java Applet Window

Работа с модулем заказа лабораторного исследования

Работа с листом заказов

По мере выполнения работы система собирает с рабочих мест назначенные исследования. Все они аккумулируются в единой базе данных

ЛИС, с которой осуществляют свою работу сотрудники лаборатории. В ЛИС предусмотрен электронный лист заказов, где хранятся все назначенные пациенту анализы в хронологическом порядке. Предусмотрены несколько представлений информации – по датам, по кодам лаборатории, по кодам назначений и т.д. Имеется возможность распечатать план работы сотрудников лаборатории. Также предусмотрена возможность внести отметку о взятии материала.

При необходимости сотрудники лаборатории могут зарегистрировать новый заказ в ручном режиме.

Код	№	Фамилия	Имя	Отчество	Дата рождения	Возраст	Пол	Адрес фактического проживания	Вид анализа	Назначил	Номер полиса
21	1	ТЕСТОВАЯ	МАРЬЯ	ИВАНОВНА	01.01.2006	4	ЖЕНЩИНА	Карелия Респ, Кондопожский	Копрологическое исследование	DevTeam	
23	1	ТЕСТОВАЯ	МАРЬЯ	ИВАНОВНА	01.01.2006	4	ЖЕНЩИНА	Карелия Респ, Кондопожский	Биохимический анализ крови	DevTeam	
26	2	ТЕСТОВАЯ	МАРИЯ	МИХАЙЛОВНА	08.01.1968	42	ЖЕНЩИНА	МОСКОВСКАЯ ОБЛАСТЬ обл, г	Биохимический анализ крови	DevTeam	45090
28	3	ТЕСТ	АННА	ИВАНОВНА	16.12.1955	55	ЖЕНЩИНА	Хакасия Респ, г Абакан, . МОС	Биохимический анализ крови	DevTeam	

Работа с листом заказов в лабораторной информационной системе

Внесение результатов исследований

Для внесения результатов автоматизации работы сотрудников самой лаборатории в состав системы включен специальный электронный журнал, который позволяет быстро и наглядно получать доступ к собранным в единые списки назначениям, просматривать журнал выполненных исследований и т.д. Данный модуль позволяет быстро внести результаты измерений вручную или запустить диспетчер драйверов анализаторов. Все результаты, вносимые в ЛИС с разных рабочих мест лаборантов, тут же становятся видны в модуле внесения результатов. Здесь же происходит управление референтными значениями, верификация результатов и их автоматическая отправка в КМИС.

Работа с модулем внесения результатов лабораторных исследований

Подключение анализаторов к информационной системе

Лабораторная информационная система КМИС версии 4.0 и выше позволяет подключать драйверы лабораторных анализаторов для автоматического получения результатов исследований на этих приборах в систему. Стоимость драйверов лабораторных анализаторов не входит в стоимость системы - их необходимо приобретать отдельно. В настоящее время поддерживаются следующие анализаторы:

- Abbott Cell-Dyn 1700/3000/3700
- Drew Scientific Excell 10/18/22
- Vitalab Flexor E

- Konelab 20/30/60
- Olympus AU-400/640/2700/5400
- Sysmex CA-510/520/530/540/550/560
- DPC Immulitte One/1000/2000
- ABX Micros 45/60/CRP
- ABX Pentra 60/60C+/ES 60/80/XL80/120/120 RET/120/SPS/120 RET SPS
- ADVIA 120/2120/2120i
- MEK 6400/6410/6420
- Erma PCE 210
- STA Compact
- Clima MC-15
- Biosen
- Clinitek 500

Список поддерживаемого оборудования постоянно расширяется. Более того, мы можем разработать недостающие драйверы анализаторов под заказа.

Получение результатов лабораторных исследований


Оформленный бланк лабораторного исследования содержит следующую информацию:

- Паспортные данные пациента (загружаются автоматически при назначении с рабочих мест врачей и медицинских сестер);
- Вид анализа;
- Дата выполнения;
- ФИО и должность исполнителя
- Отметку о срочности выполнения исследования
- Лабораторию, выполнившую исследование;
- Процедурный кабинет, где произведён забор материала;
- Список определённых показателей и их значение;
- Границы референтных значений;

- Единицы измерения;
- Сигнальные отметки о выходе результатов за пределы нормы, маркировки H (high) или L (low);
- Электронная цифровая подпись.

Печать 2 Закрывать 3 Открыть..

Внимание! Отклонение референтных значений в показателе(ях): креатин ,холестерин ,триглицериды

 **ПЕСТЕЛЬ ИВАН ИВАНОВИЧ (10.01.1956) Возраст 54**
Биохимический анализ крови (14.09.2010) [Показать сведения](#)

Дата и время выполнения: 14.09.2010 16:19:13

Citol (срочное назначение) Включить контроль исполнения в журнале лаборатории


Статистическая информация

Лаборатория: 1
 Исполнитель: Петров Петр Петрович - врач-терапевт участковый / 54545

Результаты

Показатель	Значение	Выход за нормы	Ед. измерения	Нижняя гр.	Верхняя гр.
креатин	453	H	мкмоль/л	44	132
холестерин	43	H	ммоль/л	3.1	8.5
триглицериды	34	H	ммоль/л	1.71	1.71
АСТ					

Прочее
 Подпись: DevTeam

 Электронная цифровая подпись. Документ подписан и не может быть изменен.
 Документ подписал(а): CN=DevTeam/O=KMIS
 Дата и время подписи: 14.09.2010 / 19:19

История изменения

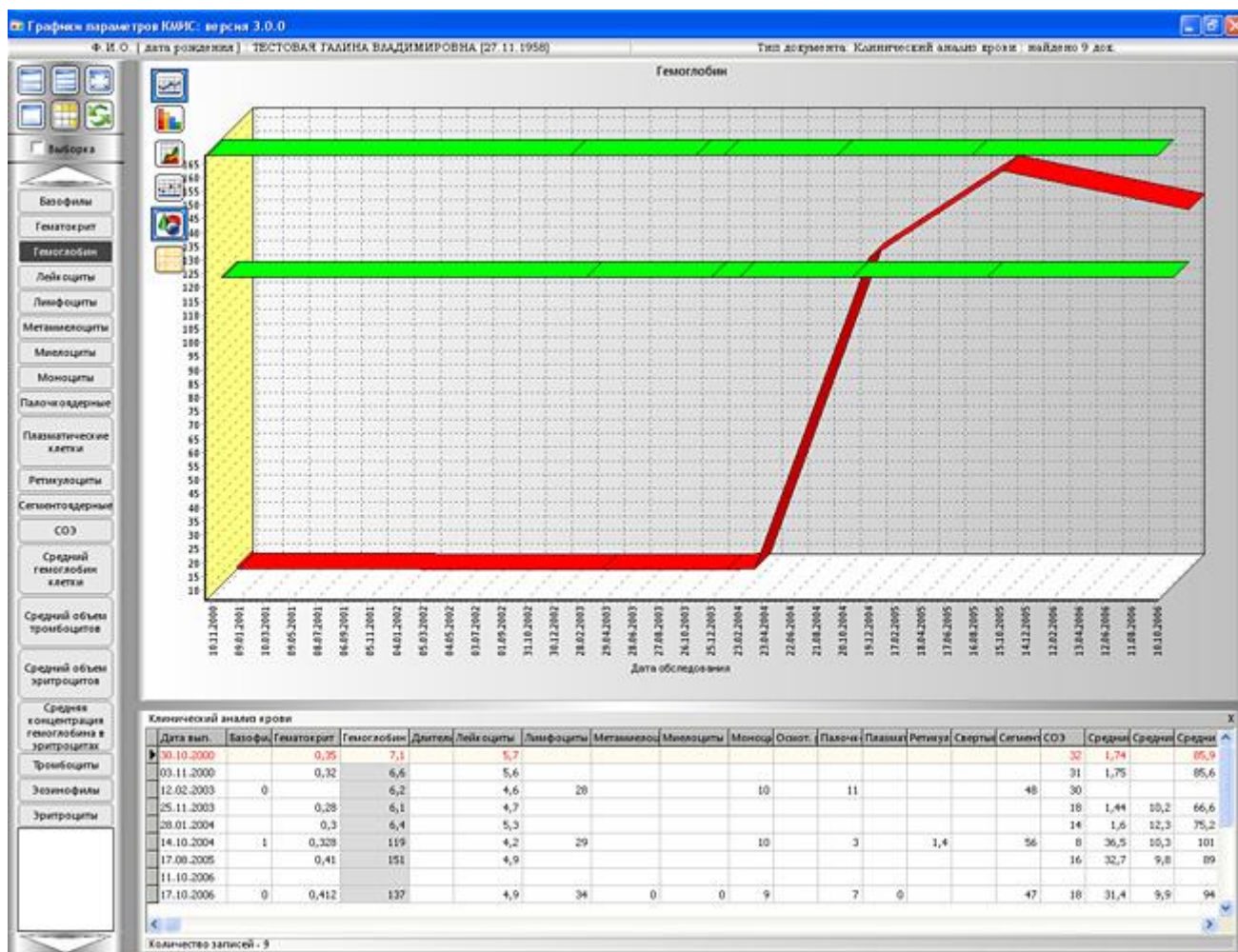
Работа с бланком заказа лабораторного исследования. Система наглядно выводит результаты исследования, референтные значения, единицы измерения и отметки о выходе за границы нормы. Ведется статистический учет срочных исследований и информации об исполнителе

Печать результатов

Поддерживается несколько возможностей печати результатов – как с рабочего места сотрудника лаборатории (в ЛИС), так и ответа из электронной медицинской карты.

Возможность построения графиков динамического изменения

Интеграция в подсистему динамических показателей позволяет просмотреть изменение любых параметров конкретного пациента. В программе предусмотрена возможность построения диаграмм при помощи специальной команды (рисунок ниже).



По числовым параметрам, хранящимся в электронных документах, система может построить графики их изменений в динамике

Кроме того, сформированные диаграммы можно:

- Экспортировать в документы формата Microsoft Word/Excel, PDF, OpenDocument Format, TIFF, RTF и т.д.;
- Напечатать диаграмму из программы;
- Изменить вид диаграммы;
- Сохранить в буфер обмена, а затем, например, отправить по электронной почте.

Поддержка сбора и анализа статистики

В ЛИС встроены отчеты о нагрузке и финансово-экономических показателях выполненной работы, которые свободно экспортируются в документы Microsoft Office, Open Office, PDF и т.д.

Встроенная система внутрилабораторного контроля качества

В систему встроена поддержка отраслевого стандарта «Правила проведения внутрилабораторного контроля качества количественных методов клинических лабораторных исследований с использованием контрольных материалов» ОСТ 91500.13.0001-2003 для обеспечения повседневных внутрилабораторных процедур контроля качества, направленных на выявление недопустимых случайных и систематических погрешностей на аналитическом этапе клинических лабораторных исследований.

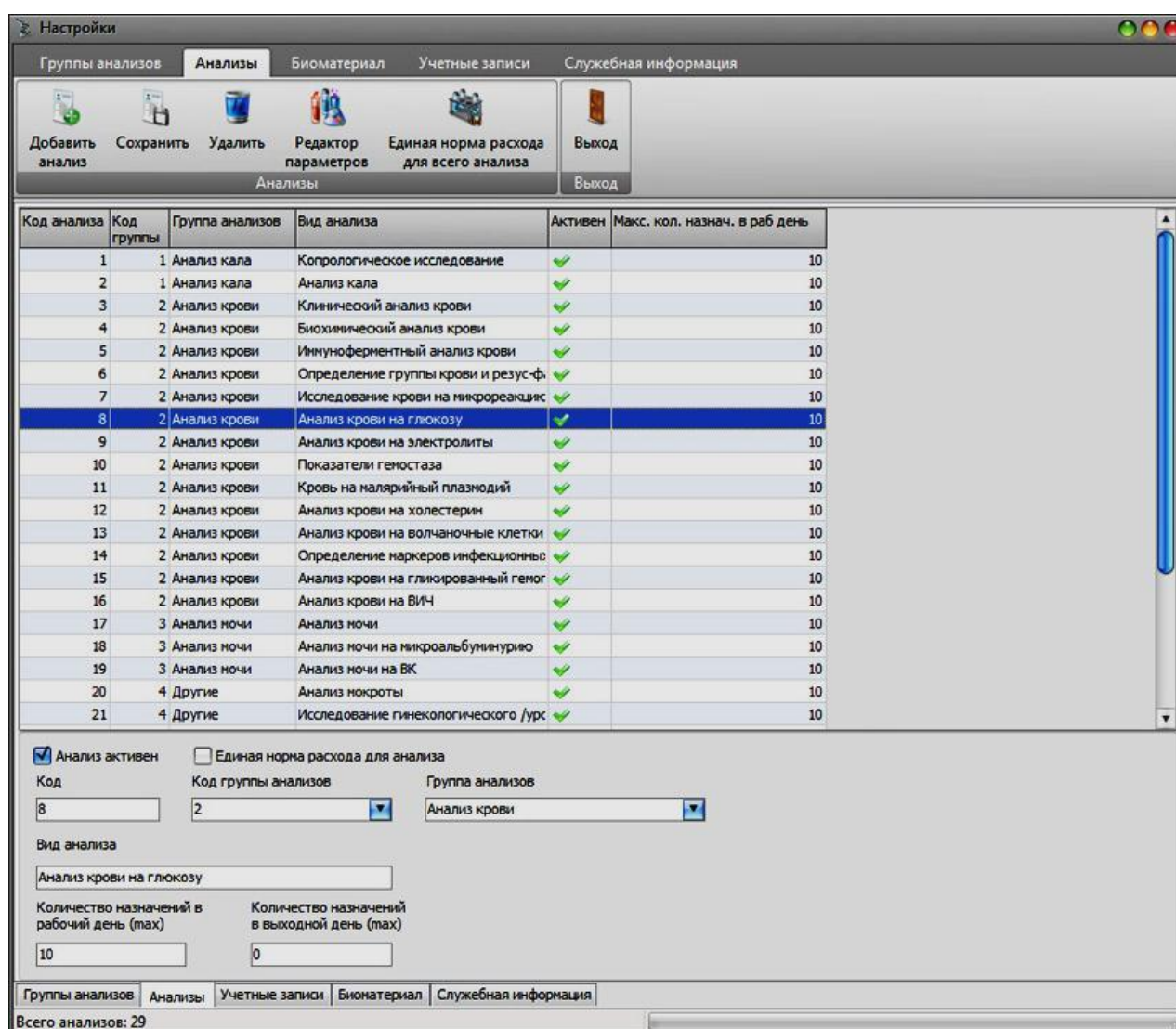
Справочная система

Встроенная справочная подсистема позволяет оперативно получать справочную информацию из любого документа или программы системы.

Возможности настройки лабораторной системы

В состав лабораторной информационной системы включен удобный модуль настроек, которые позволяет выполнять следующие действия:

- Управлять списком доступных лабораторных исследований;
- Разрешать/Запрещать заказ любого параметра;
- Ограничение суммы назначений определенного вида и т.д.;
- Управление списком биоматериалов
- Управление нормами расхода
- Управление пользователями.



Модуль настроек лабораторной информационной системы

3. Цель деятельности студентов на занятии:

Студент должен знать:

1. Что такое информационные медицинские системы диагностических служб и для чего они предназначены.
2. Что такое Радиологическая информационная система (РИС) и ее основные задачи.
3. Что такое лабораторная информационная система ее основные функции и назначение.

Студент должен уметь:

1. Иметь навыки работы с медицинскими информационными системами.
2. Знать их функциональные возможности

4. Содержание обучения:

1. Комплексная медицинская информационная система

2. Радиологическая информационная система

- а. Функциональные возможности
- б. Описание решения

3. Лабораторная информационная система

- а. Выполнение лабораторных назначений
- б. Работа с бланком лабораторного заказа
- в. Работа с листом заказов
- г. Внесение результатов исследований
- д. Получение результатов лабораторных исследований
- е. Печать результатов
- ж. Построение графиков динамического измерения
- з. Справочная система

5. Перечень вопросов для проверки уровня знаний:

1. Какие основные документы предусмотрены в Комплексной медицинской информационной системе (КМИС)?
2. Какую информацию содержит электронная медицинская карта пациента?
3. Что такое Лабораторная информационная система и в чем цель ее внедрения?

6. Перечень вопросов для проверки конечного уровня знаний:

1. Как осуществляется работа пользователей с лабораторной подсистемой.
2. Как осуществляется работа с бланком лабораторного заказа.

7. Используя лабораторную информационную систему осуществите заказ лабораторного исследования.

8. Самостоятельная работа студентов:

Изучите основные существующие информационные медицинские системы диагностических служб (отделений функциональной диагностики и лабораторных исследований).

9. Хронокарта учебного занятия:

1. Организационный момент – 5 мин.
2. Текущий контроль знаний – 30 мин.
3. Разбор темы – 20 мин.
4. Практическая работа – 30 мин.
5. Подведение итогов занятия – 10 мин.

10. Перечень учебной литературы к занятию:

1. Кишкун А.А., Гузовский А.Л. Лабораторные информационные системы и экономические аспекты деятельности лаборатории
2. Лекция по данной теме.

