

АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ В МЕДИЦИНСКОЙ СТАТИСТИКЕ

I. Абсолютные величины

При проведении медико-социального исследования статистическая сводка материала завершается получением ряда таблиц, числа в которых отражают абсолютные размеры изучаемого явления и составляющих его частей (**абсолютные величины** – количественное выражение любого явления).

Переход к анализу статистического материала включает его обработку, т.е. вычисление на основе этих абсолютных величин системы обобщающих показателей.

Однако неправильно было бы рассматривать абсолютные величины как промежуточный этап и лишать их самостоятельного значения и ценности.

Абсолютные величины имеют в медицинской статистике определенное значение.

В большинстве случаев они интересны сами по себе, характеризуя, например численность населения стран и населенных пунктов, число врачей, больничных коек, количество рождений и смертей, редкие случаи некоторых заболеваний (малярия, полиомиелит, столбняк, сыпной тиф) или наоборот массовое распространение других (туберкулез, ВИЧ-инфекция).

Таким образом абсолютные цифры применяются:

- для характеристики абсолютных размеров явления в целом (показывают массовость явления);
- для характеристики редко встречающихся явлений (показывают единичность явлений).

Сравнивая размеры двух явлений (рождаемость, смертность, заболеваемость и т.п.), либо изучая изменения этих явлений во времени, необходимо абсолютные числа, выражающие эти размеры преобразовать в **производные величины** – относительные и средние.

II. Относительные величины

Относительные величины представляют собой соотношение двух абсолютных величин в виде кратных величин или в виде процентов, получающихся при сравнении двух значений одного признака или

сравнении части с целым. Эталонами для сравнения служат базисные данные, принимаемые за 100, 1000, 10 000, 100 000, т. е. за единицу с нулями.

По своему содержанию относительные величины (показатели, коэффициенты) делятся на 4 основные группы (вида):

- экстенсивные;
- интенсивные;
- наглядности;
- соотношения.

К особым видам относительных статистических показателей относятся:

- относительные показатели, характеризующие динамику процесса, изменение во времени (темпы роста);
- относительные показатели отношения фактически наблюдаемых величин признака к его нормативным, плановым, оптимальным или максимально возможным величинам;
- относительные показатели, характеризующие взаимосвязи между разными признаками объекта, объектом и средой и т. п. К ним принадлежат коэффициенты регрессии, эластичности, детерминации, корреляции, а так же аналитические индексы.

Экстенсивные показатели или показатели структуры (распределения) характеризуют распределение явлений (структуру объекта) внутри одной совокупности (целого) на составляющие его части по их удельному весу или отношение части к целому.

Обычно экстенсивные показатели выражаются в процентах (%) или, реже в промилле к итогу.

Примеры экстенсивного показателя: лейкоцитарная формула (доля отдельных форменных элементов во всей массе лейкоцитов), структура населения по возрасту, полу, структура заболеваний и причин смерти и т.д.

Вычисление производится по формуле:

$$\text{Э. П.} = \frac{\text{часть явления}}{\text{целое явление}} \times 100$$

Например, для определения структуры инфекционной заболеваемости все зарегистрированные заболевания (120 случаев) принимаются за 100%, из них:

- вирусный гепатит – 25 случаев (x_1 %),
- скарлатина – 10 случаев (x_2 %),
- дизентерия – 45 случаев (x_3 %),
- корь – 25 случаев (x_4 %),
- эпидемический паротит – 15 случаев (x_5 %),

Отсюда:

$$\text{Удельный вес данного заболевания} = \frac{\text{число зарегистрированных заболеваний данной болезнью}}{\text{число всех зарегистрированных заболеваний}} \times 100$$

то есть, в данном случае:

$$= \frac{\text{число зарегистрированных случаев гепатита}}{\text{число всех зарегистрированных инфекционных заболеваний}} \times 100$$

$$= \frac{25 \times 100}{120} = 20,8 \%$$

Так же определяется удельный вес других заболеваний в структуре инфекционной заболеваемости. В результате структура инфекционной заболеваемости будет представлена в следующем виде:

№ п/п	Нозологические формы инфекционной заболеваемости	Абсолютные показатели	Экстенсивные показатели (удельный вес) %
1	Вирусный гепатит	25	20,8
2	Скарлатина	10	8,3
3	Дизентерия	50	41,7
4	Корь	20	16,7
5	Эпидемический паротит	15	12,5
	Всего:	120	100

Экстенсивные показатели (коэффициенты) нужны для определения структуры статистической совокупности и сравнительной оценки соотношения составляющих ее частей и применяются:

- при анализе структуры заболеваемости;
- при анализе причин смерти;
- при анализе структуры коечного фонда;
- при анализе структуры финансовых расходов на здравоохранение и т.д.

Характерной чертой экстенсивных коэффициентов является их взаимосвязанность. Так, например, при изучении структуры заболеваемости удельный вес какого-нибудь заболевания может возрасти:

- при подлинном его росте;

– при одном и том же уровне, если число других заболеваний в этот период снизилось;

– при снижении уровня данного заболевания, если уменьшение числа других заболеваний происходило более быстрым темпом.

На основании экстенсивных коэффициентов (показателей), которые отражают только структуру той или иной совокупности, нельзя судить о частоте изучаемого явления и динамике его во времени. Для этой цели необходимо знать численность среды, в которой происходят явления, и вычислить интенсивные коэффициенты.

Интенсивные показатели (показатели частоты, распространенности) характеризуют частоту явления в той среде, где мы это явление наблюдаем.

Эти показатели указывают о том, как часто встречается данное явление в той или иной среде.

Вычисление интенсивного показателя (коэффициента) производится по формуле:

$$\text{И. П.} = \frac{\text{явление} \times \text{основание показателя}}{\text{среда}}$$

Для вычисления показателя интенсивности недостаточно знать лишь величину интересующего нас явления, необходимо знать еще величину той среды, в которой данное явление наблюдается.

В медико-социальных исследованиях за " **среду** ", как правило, принимается численность населения в целом или отдельных его групп (по возрасту, полу, профессии, месту жительства и т.д.) При этом численность населения берут на середину изучаемого периода времени (чаще всего года), либо, что правильнее, рассчитывают среднюю численность населения (полусумму численности населения на начало и конец периода наблюдения).

Явление представляет собой как бы продукт "среды". Например: больные (среда) и умершие из их числа (явление); женщина детородного возраста (среда) и родившиеся у них дети (явление) и т. д.

Основание – единица с нулями (100, 1000, 10000, 100000 и т. д.). При перемене основания величина коэффициента изменяется в соответствующее число раз.

Вычисляется коэффициент интенсивности: на 100, на 1000, на 10000, на 100 000 (чаще на 1000 – в промилле (‰)).

Например, в цехе А. из 500 работающих заболело 35 человек, а в цехе Б. из 300 работающих – 45 человек.

$$\text{И.П. (заболеваемость)} \\ \text{в цехе А.} = \frac{\text{«явление»} \times 1000}{\text{«среда»}} = \frac{35 \times 1000}{500} = 70\text{‰}$$

$$\text{И.П. (заболеваемость) в цехе Б.} = \frac{\text{«явление»} \times 1000}{\text{«среда»}} = \frac{45 \times 1000}{300} = 150\%$$

Таким образом, уровень (частота) заболеваемости работающих в цехах А. и Б. будет представлен в следующем виде:

Подразделение работающих	Количество работающих	Абсолютные показатели заболеваемости	Интенсивные показатели заболеваемости (‰)
Цех А.	500	35	70,0
Цех Б.	300	45	150,0
В целом	800	80	100,0

В медицинской статистике при вычислении размеров (уровней) рождаемости, смертности, естественного прироста населения, общей заболеваемости за основание обычно принимается 1000 человек населения; вычисление размеров смертности или заболеваемости в отношении какой-либо отдельной болезни или причины смерти чаще производится на 10 000 или 100 000 (например, уровень распространённости туберкулёза, онкозаболеваний, психических расстройств и др.); вычисление показателей заболеваемости с временной утратой трудоспособности производятся на 100 работающих, показатель летальности на 100 болевших (или выбывших из стационара).

Интенсивные коэффициенты могут быть **общими и специальными**.

Общие коэффициенты характеризуют частоту явления в целом, например, общие коэффициенты рождаемости, смертности, заболеваемости, вычисленные по отношению к всему населению территории (города, района и т.д.)

Особенностью специальных коэффициентов является уточнение группировки. Например, при вычислении специальных коэффициентов рождаемости (коэффициентов плодовитости) за среду принимаются не всё население, а только женщины детородного (фертильного) возраста (15 – 49 лет).

Интенсивные показатели применяются:

- для определения уровня, частоты, распространённости того или иного явления (при анализе рождаемости, заболеваемости, смертности и др.);
- для сравнения ряда различных совокупностей по степени частоты того или иного явления (например, для сравнения уровней рождаемости в разных странах, разных районах, уровней смертности – в разных возрастных группах и т.д.).

– для выявления в динамике изменений степени частоты явления в наблюдаемой совокупности (например, изменения в заболеваемости населения города за 5 лет и т.д.).

Показатели интенсивности обобщают вторичные признаки объектов. Например, интенсивность труда врачей-терапевтов стационара (число выписанных больных терапевтического профиля на одного врача).

Показатели интенсивности бывают **прямыми и обратными**.

Например, отношение затрат труда врача-стоматолога на производство медицинской услуги к объему услуг дает показатель трудоемкости стоматологических медицинских услуг. Эта величина является обратной прямому показателю производительности труда врача-стоматолога, вычисляемому как отношение оказываемых услуг в натуральном или стоимостном выражении к затратам труда на их производство.

К этой же группе показателей относятся и такие, как соотношение между ростом и весом человека, характеризующие пропорциональность его тела.

Показатели соотношения характеризуют численное соотношение двух не связанных между собой совокупностей и указывает на частоту данного явления в неоднородной среде.

По методике вычисления коэффициенты соотношения сходны с интенсивными коэффициентами, но различны по существу.

$$\text{П.С.} = \frac{\text{«явление»} \times 100 (1000, 10\ 000, 100\ 000)}{\text{«среда», не являющаяся основанием для возникновения данного явления.}}$$

Примером коэффициента соотношения может служить обеспеченность населения больничными койками, врачами и т.д.

Например, в лечебных учреждениях г. Благовещенска с численностью населения 220 тыс. человек развернуто 1500 коек.

В этом случае показатель обеспеченности населения города больничными койками (показатель соотношения) составит:

$$\begin{aligned} \text{Обеспеченность населения койками} &= \frac{\text{число коек}}{\text{среднее число населения}} \times 10000 \\ \text{(показатель соотношения)} & \\ &= \frac{1500 \times 10000}{220000} = 68,2 \text{ на } 10\ 000 \text{ населения} \end{aligned}$$

Отличие данного показателя от коэффициента интенсивности будет заключаться в том, что интересующее нас «явление» (больничные койки) не представляют собой продукт той «среды» на которую производится

расчёт (население), т.е. эти показатели определяют отношение между разными совокупностями.

Показатели соотношения наиболее часто применяются при характеристике обеспеченности населения медицинскими кадрами, различными видами медицинской помощи, расходов здравоохранения на душу населения.

Показатели наглядности характеризуют динамику изучаемого явления при условии, что один из исходных уровней принимается за 100 (в %) или за единицу (в кратности).

Показатели наглядности применяются с целью более наглядного и доступного сравнения рядов абсолютных, относительных, или средних величин, для определения изменений, происходящих с тем или иным явлением во времени или сравнения друг с другом каких либо величин на различных территориях.

Они не имеют какого-либо качественного содержания, а представляют собой технический прием преобразования цифровых показателей. При вычислении коэффициентов наглядности одна из сравниваемых величин принимается за 100, а остальные величины с помощью обычной пропорции пересчитываются в коэффициенты по отношению к этому числу.

Чаще всего за 100 принимается первая исходная величина ряда, однако, это не обязательно (за 100 может быть взята величина из середины, или конца ряда или его средняя величина).

Вычисляется : – в процентах (%) если изменения не значительны,
– в кратности (во сколько раз произошло увеличение или уменьшение того или иного явления).

Например, уровень заболеваемости туберкулёзом населения России составил в 1991 году 34,0 на 100 000 населения, а в 1999 году – 85,2. Принимая уровень заболеваемости населения туберкулёзом в 1991 году за 100, определяем показатель наглядности заболеваемости туберкулёзом в 1999 году:

$$\frac{34,0 - 100}{85,2 - x} \quad x = \frac{85,2 \times 100}{34,0} = 250,5\%$$

Таким образом, делаем вывод о том, что уровень заболеваемости населения России туберкулёзом в 1999 году возрос на 250,5% или в 2,5 раза.

Наиболее частые ошибки, встречаемые при пользовании относительными величинами:

– Иногда делаются выводы об изменении частоты явлений, основанные на коэффициентах структуры, тогда как правильное заключение позволяют сделать только интенсивные коэффициенты.

– Не следует производить складывание или вычитание статистических коэффициентов, вычисленных из совокупностей неодинаковой численности, так как это может привести к грубым ошибкам.

– Необходимо правильно выбрать основание для коэффициента (например, коэффициент смертности мужчин следует вычислять по отношению к числу мужского, а не к числу всего населения; коэффициент летальности при гриппозной пневмонии вычисляется по отношению к больным, имевшим это осложнение, а не ко всем случаям гриппа вообще и т.д.).

– Следует учитывать фактор времени и не сравнивать коэффициенты, вычисленные за различный период времени, что так же может привести к ошибочным заключениям.

При построении относительных показателей целесообразно руководствоваться следующими принципами:

1. Сравнимые показатели должны быть чем-то связаны в реальной жизни независимо от нашего желания. При этом большое внимание необходимо уделять тому, чтобы сравниваемые показатели соответствовали смыслу.

2. Сравнимые исходные показатели могут различаться только одним атрибутом или видом признака (при одинаковом объекте, периоде времени, плановом или профилактическом характере показателей), или времени (при том же признаке, объекте и т.п.), или только фактическим, плановым или нормативным характером показателей (тот же объект, признак, время) и т. д. Нельзя сопоставлять показатели, различные по двум и более атрибутам.

3. Необходимо знать возможные границы существования относительного показателя. Например, относительные показатели вариации теряют смысл и не могут применяться в тех случаях, когда их знаменатели – средние значения признаков – близки к нулю, потому что при стремлении знаменателя к нулю относительный показатель стремится к абсурдному значению.

Аналогично если исходные показатели в текущем и базисном периодах имеют разные знаки, то теряет смысл и не может применяться относительная величина динамики – темп роста. Если ЛПУ имело в 1997 году убыток по хозрасчётной деятельности 100 млн. руб., а в 1998 году получило доход 300 млн. руб., не верно ни то, что «финансовый результат вырос втрое» (если отбросить знаки), ни то, что он «вырос в минус 3 раза», если делить +300 на – 100.

Относительные показатели, измеряющие степень приближения некоторого признака к предельному значению, должны строиться так, чтобы в пределе увеличения они стремились к единице, а в другом пределе

своего уменьшения – к нулю. Так строятся коэффициенты, измеряющие тесноту связи признаков, степень эффективности использования ресурсов.

Литература:

1. Лисицин Ю.П., Улумбекова Г.Э. Общественное здоровье и здравоохранение: учебник. - 3-е изд. перед.и доп. – М.: ГЭОТАР-Медиа, 2011.- 544. – 3 экз.
2. Медик В.А. Общественное здоровье и здравоохранение: учебник. – 2-е изд. исп. и перер. – М.: ГЭОТАР-Медиа, 2012. – 608с. – 30 экз.
3. Полунина Н.В. Общественное здоровье и здравоохранение: учебник. - М.: ООО «МИА», 2010. – 544 с. - 20 экз.
4. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: учебное пособие для практических занятий/Под ред. В.З.Кучеренко. – М.: ГЭОТАР-Медиа, 2011.

Оглавление

Цель занятия:	3
Студент должен уметь	3
Студент должен знать:	3
Информационный материал.....	4
Основные обозначения вариационного ряда	4
Виды вариационных рядов.....	4
Средняя величина.....	4
Применение средних величин	4
Методика расчета простой средней арифметической.....	5
Методика расчета взвешенной средней арифметической (табл. 1).....	5
Методика расчета среднеквадратического отклонения (см. табл. 1)	5
Применение среднеквадратического отклонения.....	5
Коэффициент вариации (C_v)	6
Применение коэффициента вариации.....	6
РАСЧЕТ ПОКАЗАТЕЛЕЙ ВАРИАЦИОННОГО РЯДА, ИСПОЛЬЗУЯ МАСТЕР ФУНКЦИЙ (f_x) MS EXCEL	8
КОНТРОЛЬНЫЕ ВОПРОСЫ	18
ТЕСТОВЫЕ ЗАДАНИЯ:	18
СИТУАЦИОННЫЕ ЗАДАЧИ:	20

Хронология занятия

1. Формулировка и обоснования цели занятия (10 мин.);
 2. Изложение основных вопросов темы (60 мин.);
 3. Перерыв (20 мин.)
 4. Самостоятельная работа студентов с методическим материалом - (30 мин.)
 5. Разбор типовых задач по изучаемой теме (20 мин.)
 6. Самостоятельное решение задач (40 мин.)
 7. Тестовый контроль на ПК (15 мин.)
- Общая продолжительность занятия – 195 минут.

Цель занятия:

ознакомить студентов методам вариационной статистики для оценки и анализа статистической совокупности при изучении общественного здоровья и деятельности медицинских учреждений.

Студент должен уметь

- выявлять основную закономерность изучаемого признака путем вычисления средних величин;
- обосновывать методику применения критериев разнообразия вариационного ряда;
- давать характеристику разнообразия вариационного ряда;
- делать выводы о типичности обобщающей характеристики признака в изучаемой совокупности, используя критерии разнообразия вариационного ряда;
- рассчитывать средние величины и критерии вариационного ряда, используя мастер функций MS Excel.

Студент должен знать:

- основные понятия темы (вариационный ряд, средние величины, среднеквадратическое отклонение, коэффициент вариации, правило трех сигм, нормальное распределение Гаусса);
- методику расчета средних величин и критериев разнообразия вариационного ряда;
- методику анализа средних величин: значение среднеквадратического отклонения и коэффициента разнообразия для оценки вариабельности изучаемого признака и типичности средней величины;
- нормальное распределение вариационного ряда и его значение для оценки статистических показателей;
- область применения характеристик вариационного ряда.
-

Место проведения

Дисплейный класс кафедры общественного здоровья и организации здравоохранения с курсом медицинской информатики

Оснащение занятия

Мультимедийный проектор

Наглядный материал в виде мультимедийных презентаций

Персональные компьютеры

Информационный материал

При изучении общественного здоровья (например, показателей физического развития), анализе деятельности учреждений здравоохранения за год (длительность пребывания больных на койке и др.), оценке работы медицинского персонала (нагрузка врача на приеме и др.) часто возникает необходимость получить представление о размерах изучаемого признака в совокупности для выявления его основной закономерности.

Оценить размер признака в совокупности, изменяющегося по своей величине, позволяет лишь его обобщающая характеристика, называемая средней величиной.

Для более детального анализа изучаемой совокупности по какому-либо признаку помимо средней величины необходимо также вычислить критерии разнообразия признака, которые позволяют оценить, насколько типична для данной совокупности ее обобщающая характеристика.

1. **Определение вариационного ряда.** Вариационный ряд - это числовые значения признака, представленные в ранговом порядке с соответствующими этим значениям частотами.

Основные обозначения вариационного ряда

V — варианта, отдельное числовое выражение изучаемого признака;
p — частота ("вес") варианты, число ее повторений в вариационном ряду;

n — общее число наблюдений (т.е. сумма всех частот, $n = \sum p$);

V_{max} и **V_{min}** — крайние варианты, ограничивающие вариационный ряд (лимиты ряда);

A — амплитуда ряда (т.е. разность между максимальной и минимальной вариантами, $A = V_{\max} - V_{\min}$)

Виды вариационных рядов

- а) простой** — это ряд, в котором каждая варианта встречается по одному разу ($p=1$); **б) взвешенный** — ряд, в котором отдельные варианты встречаются неоднократно (с разной частотой).
2. **Назначение вариационного ряда** Вариационный ряд необходим для определения средней величины (**M**) и критериев разнообразия признака, подлежащего изучению (**σ**, **C_v**).

Средняя величина

- это обобщающая характеристика размера изучаемого признака. Она позволяет одним числом количественно охарактеризовать качественно однородную совокупность.

Применение средних величин

- для оценки состояния здоровья — например, параметров физического развития (средний рост, средняя масса тела, среднее значение жизненной емкости легких и др.), соматических показателей (средний уровень сахара в крови, средняя величина пульса, средняя СОЭ и др.);

- для оценки организации работы лечебно-профилактических и санитарно-противоэпидемических учреждений, а также деятельности отдельных врачей и других медицинских работников (средняя длительность пребывания больного на койке, среднее число посещений на 1 ч приема в поликлинике и др.);
 - для оценки состояния окружающей среды.
3. В статистике принято выделять следующие **виды средних величин**: мода (M_o), медиана (M_e) и средняя арифметическая (M). Мода – величина варьирующего признака, наиболее часто встречающаяся в совокупности. В вариационном ряду это варианта, имеющая наибольшую частоту встречаемости. Обычно мода является величиной довольно близкой к средней арифметической, совпадает с ней при полной симметрии распределения. Медиана – варианта, делящая вариационный ряд на две равные половины. При нечетном числе наблюдений медианой является варианта, имеющая в вариационном ряду порядковый номер $(n + 1) : 2$. Средняя арифметическая величина (M) – в отличие от моды и медианы опирается на все произведенные наблюдения, поэтому является важной характеристикой для всего распределения.

Методика расчета простой средней арифметической

1. Суммировать варианты: $V_1 + V_2 + V_3 + \dots + V_n = \Sigma V$;
2. Сумму вариант разделить на общее число наблюдений: $M = \Sigma V / n$

Методика расчета взвешенной средней арифметической (табл. 1)

3. Получить произведение каждой варианты на ее частоту — V_p
4. Найти сумму произведений вариант на частоты: $V_{1p_1} + V_{2p_2} + V_{3p_3} + \dots + V_{np_n} = \Sigma V_p$
5. Полученную сумму разделить на общее число наблюдений: $M = \Sigma V_p / n$

Методика расчета среднеквадратического отклонения (см. табл. 1)

6. Найти отклонение (разность) каждой варианты от среднеарифметической величины ряда ($d = V - M$);
7. Возвести каждое из этих отклонений в квадрат (d^2);
8. Получить произведение квадрата каждого отклонения на частоту (d^2p);
9. Найти сумму этих отклонений: $d^2_1p_1 + d^2_2p_2 + d^2_3p_3 + \dots + d^2_np_n = \Sigma d^2p$;
10. Полученную сумму разделить на общее число наблюдений (при $n < 30$ в знаменателе $n-1$): $\Sigma d^2p / n$
11. Извлечь квадратный корень: $\sigma = \sqrt{\Sigma d^2p / n}$
при $n < 30$ $\sigma = \sqrt{\Sigma d^2p / n-1}$

Применение среднеквадратического отклонения

- для суждения о колеблемости вариационных рядов и сравнительной оценки типичности (представительности) средних

арифметических величин. Это необходимо в дифференциальной диагностике при определении устойчивости признаков;

- для реконструкции вариационного ряда, т.е. восстановления его частотной характеристики на основе правила "трех сигм". В интервале $M \pm 3\sigma$ находится 99,7% всех вариантов ряда, в интервале $M \pm 2\sigma$ — 95,5% и в интервале $M \pm 1\sigma$ — 68,3% вариант ряда — нормальное распределение (распределение Гаусса), при этом M — находится в максимуме (Рис. 1);

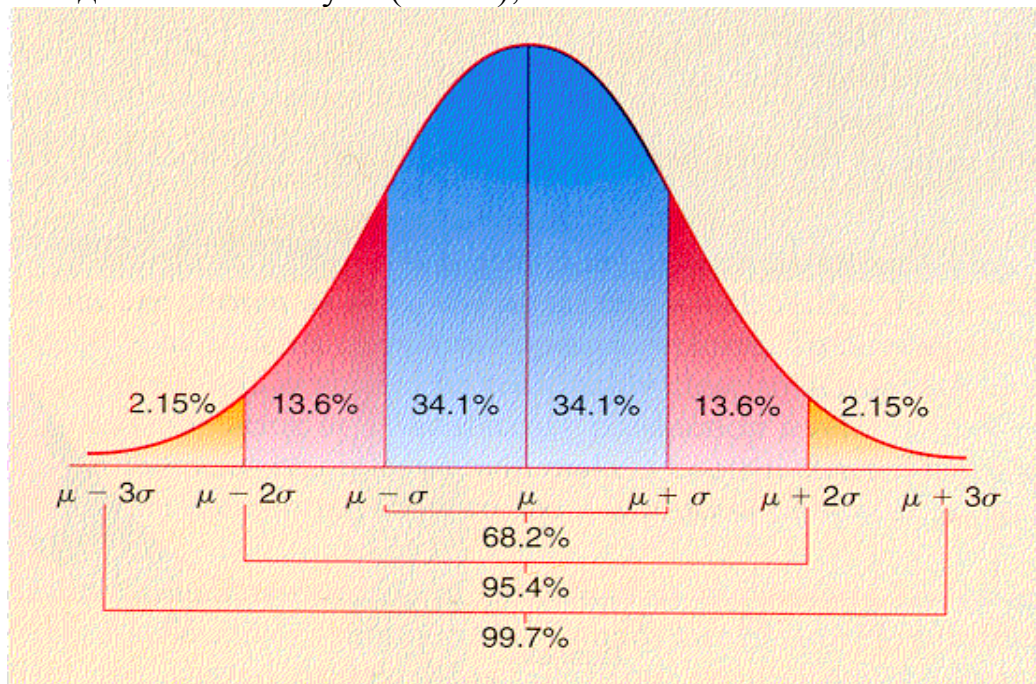


Рис. 1 График плотности нормального распределения

- для выявления "выскакивающих" вариант (при сопоставлении реального и реконструированного вариационных рядов);
- для определения параметров нормы и патологии с помощью сигмальных оценок;
- для расчета коэффициента вариации;
- для расчета средней ошибки средней арифметической величины.

Коэффициент вариации (C_v)

- это процентное отношение среднеквадратического отклонения к среднеарифметической величине: $C_v = \sigma / M \times 100\%$. Коэффициент вариации — это относительная мера колеблемости вариационного ряда.

Применение коэффициента вариации

- для оценки разнообразия каждого конкретного вариационного ряда и, соответственно, суждения о типичности отдельной средней (т.е. ее способности быть полноценной обобщающей характеристикой данного ряда). При $C_v < 10\%$ разнообразие ряда считается слабым, при C_v от 10 до 20% — средним, а при $C_v > 20\%$ — сильным. Сильное разнообразие ряда свидетельствует о малой представительности (типичности) соответствующей средней величины и, следовательно, о нецелесообразности ее использования в практических целях;

- для сравнительной оценки разнообразия (колеблемости) разноименных вариационных рядов и выявления более и менее стабильных признаков, что имеет значение в дифференциальной диагностике.

Таблица 1. Результаты измерения массы тела 7-летних мальчиков города Н. в 2007 г.

Масса тела (в кг) V	Середина интервала (центральная варианта) V1	Число мальчиков P	V _p	d = V - M	d ²	d ² p
15-18,9	17	16	272	-7	49	784
19-22,9	21	27	567	-3	9	243
23-26,9	25	32	800	+1	1	32
27-30,9	29	16	464	+5	25	400
31-34,9	33	9	297	+9	81	729

Задача - эталон

		n = 100	$\sum V_p =$	2400		$\sum d^2p =$	2188
--	--	---------	--------------	------	--	---------------	------

Условие задачи. В городе Н. в 2007 г. проведено измерение массы тела 7-летних мальчиков (данные представлены в табл. 1). По данным аналогичного исследования, выполненного в городе Н. в 1997 г., средняя масса тела 7-летних мальчиков составила 23,8 кг, $\sigma \pm 3,6$ кг.

Задание.

1. Вычислить среднюю арифметическую величину (M) и критерии разнообразия вариационного ряда (σ , C_v).
2. Оценить полученные результаты, сравнить их с данными предыдущего исследования, сделать соответствующие выводы.

Решение

В сгруппированном вариационном ряду центральная варианта рассчитывается как полусумма начальных вариантов соседних интервалов:

$$M = \sum V_p / n = 2400 / 100 = 24,0 \text{ (кг);}$$

$$\sigma = \sqrt{\sum d^2p / n} = \sqrt{2188 / 100} = \pm 4,68 \text{ (кг);}$$

$$C_v = \sigma / M \times 100\% = (4,68 / 24,0 \times 100) = 19,5\%.$$

Выводы:

1. Средняя масса тела 7-летних мальчиков в городе Н. в 2007 г. составляет 24,0 кг,
2. $\sigma = \pm 4,68$ (кг).

3. Величина коэффициента вариации, равная 19,5% свидетельствует о среднем разнообразии признака (приближающемся к сильному)

Таким образом, можно считать, что полученная средняя величина массы тела является достаточно представительной (типичной). По сравнению с 1997 г., в 2007 г. отмечается более значительная вариабельность массы тела у мальчиков 7 лет ($\pm 4,68$ кг против 3,6 кг).

Аналогичный вывод вытекает и из сопоставления коэффициентов вариации (C_v в 1997 году равен $(3,6 / 23,8 \times 100 = 15,1\%)$).

РАСЧЕТ ПОКАЗАТЕЛЕЙ ВАРИАЦИОННОГО РЯДА, ИСПОЛЬЗУЯ МАСТЕР ФУНКЦИЙ (f_x) MS EXCEL

Пример 1. Рассмотрим частоту сердечных сокращений у 20 студентов и рассчитаем средние величины.

Введем в ячейку A1 – A20 значения частоты сердечных сокращений (ЧСС) студентов: 48 56 54 57 47 50 59 60 67 68 70 69 74 75 53 58 86 51 88 60 (Рис.1).

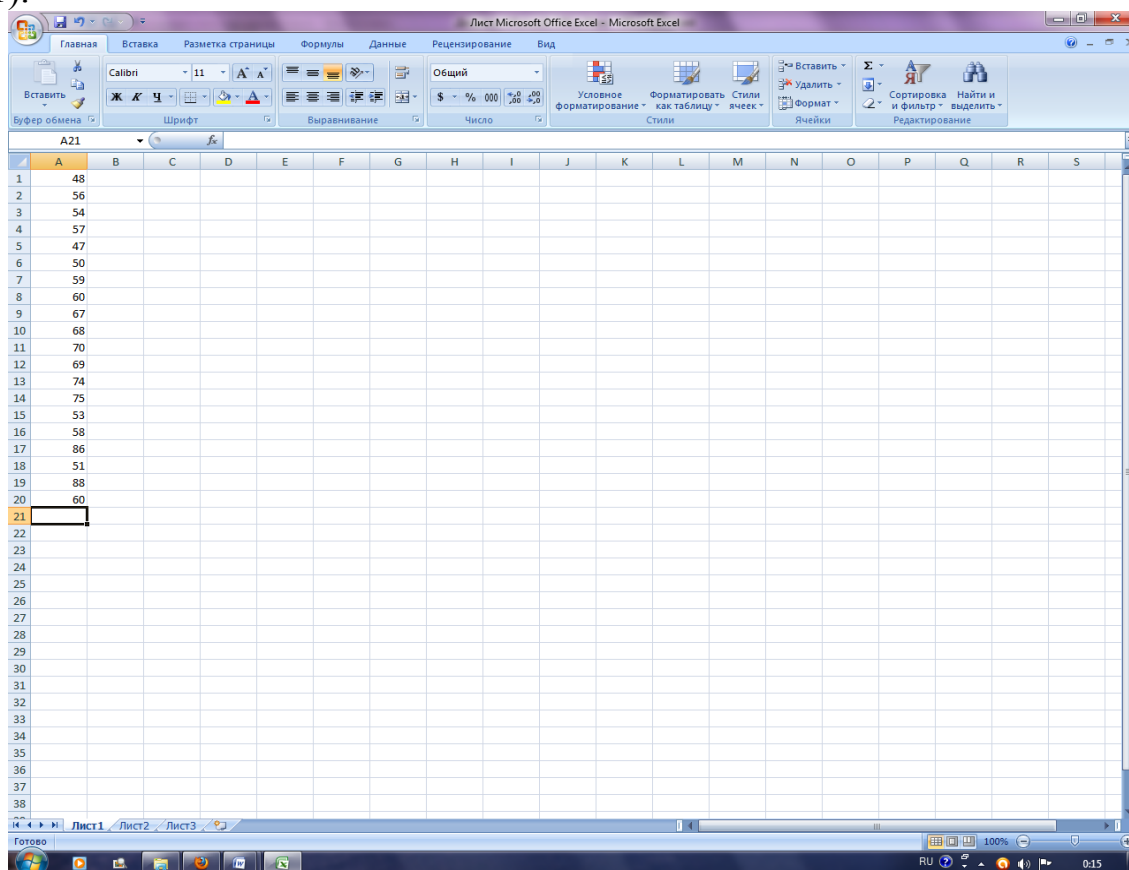


Рис. 1 Ввод данных

Для расчета Моды, открываем **Мастер функций**. Для этого определяем курсором ячейку B22, щелкаем мышью на кнопке f_x в строке состояния и открывается окно **Мастер функций**. В категории выбираем **Статистические**, в перечне функций находим функцию **Мода** и нажимаем **ОК** (Рис. 2, 3).

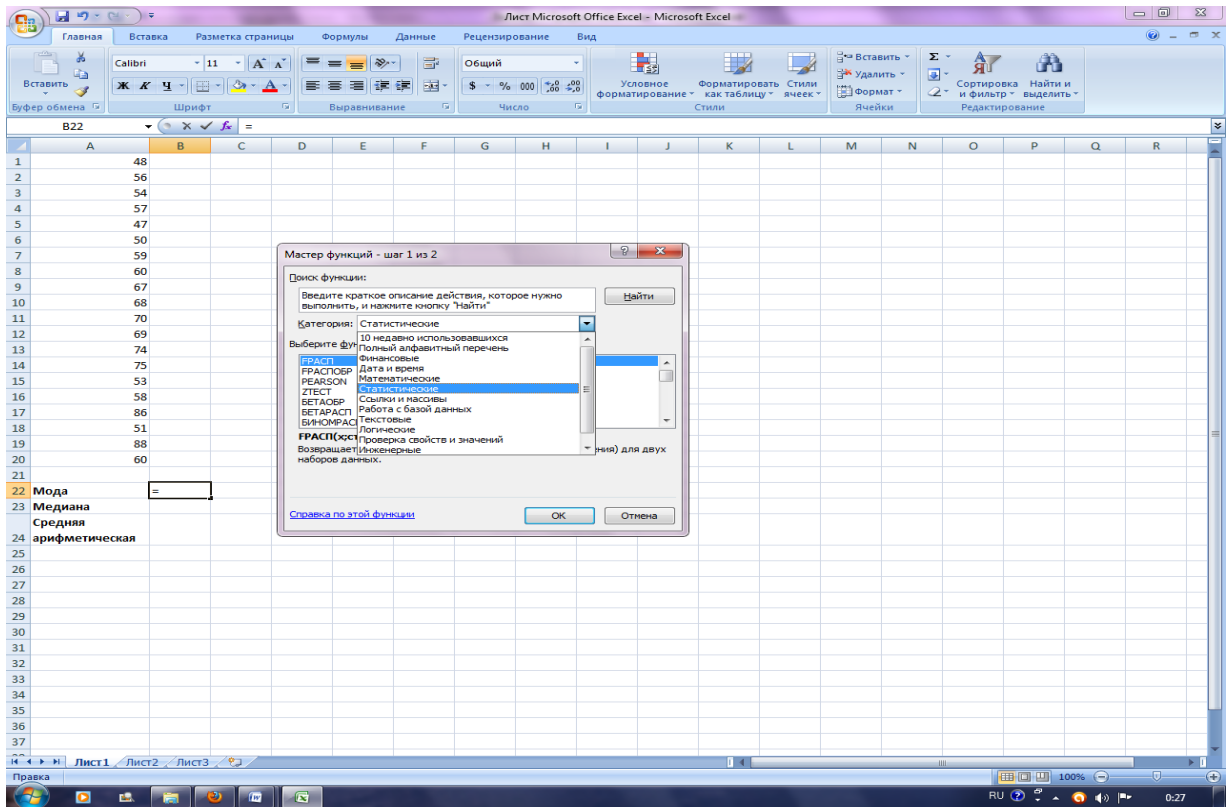


Рис. 2 Выбор категории

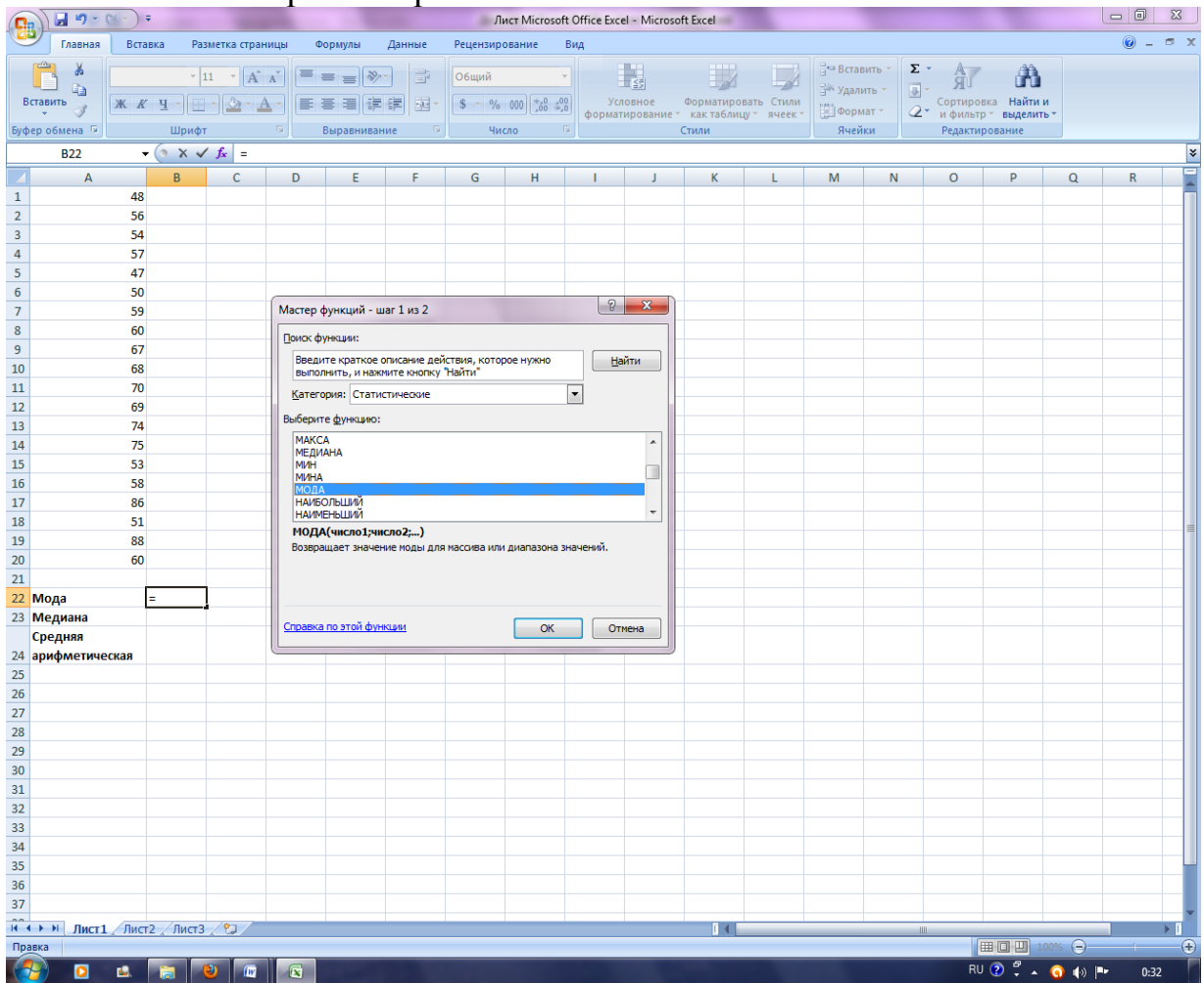


Рис. 3 Выбор функции Мода

В появившемся диалоговом окне **Аргументы функции** определяем массив A1:A20 и нажимаем **ОК** (Рис. 4).

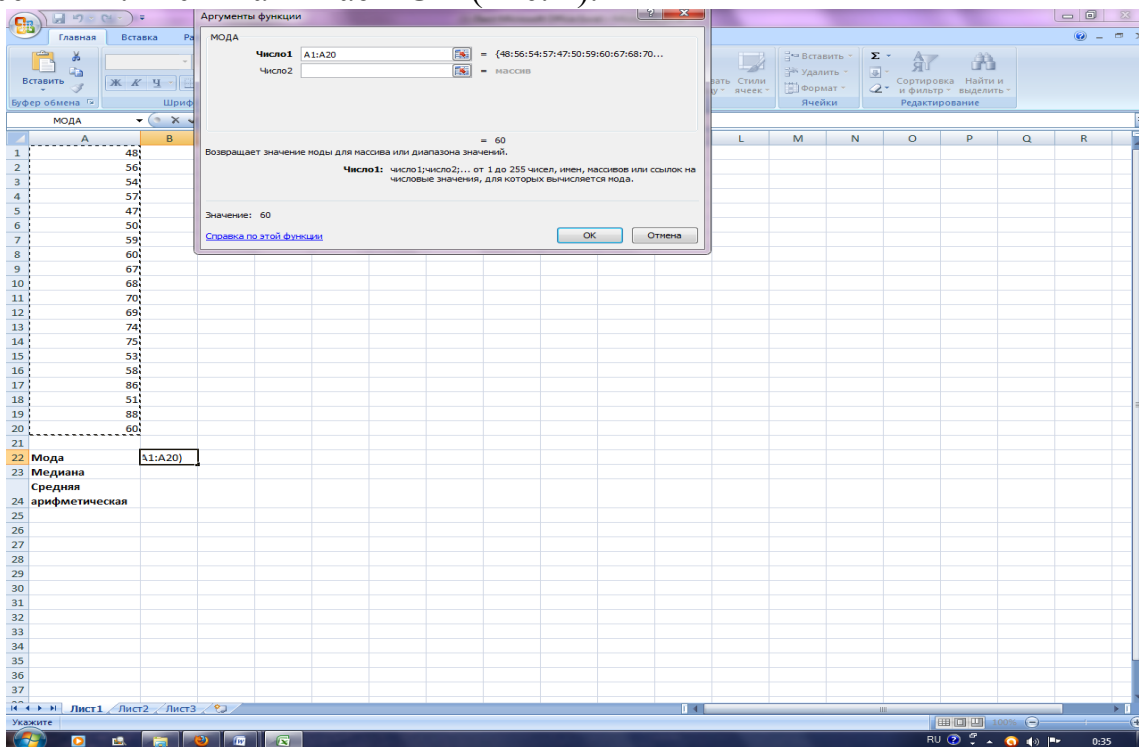


Рис. 4 Выбор массива

В нашем примере Мода равна 60.

Аналогичным способом вычисляем Медиану и среднюю арифметическую (Рис. 5 - 8).

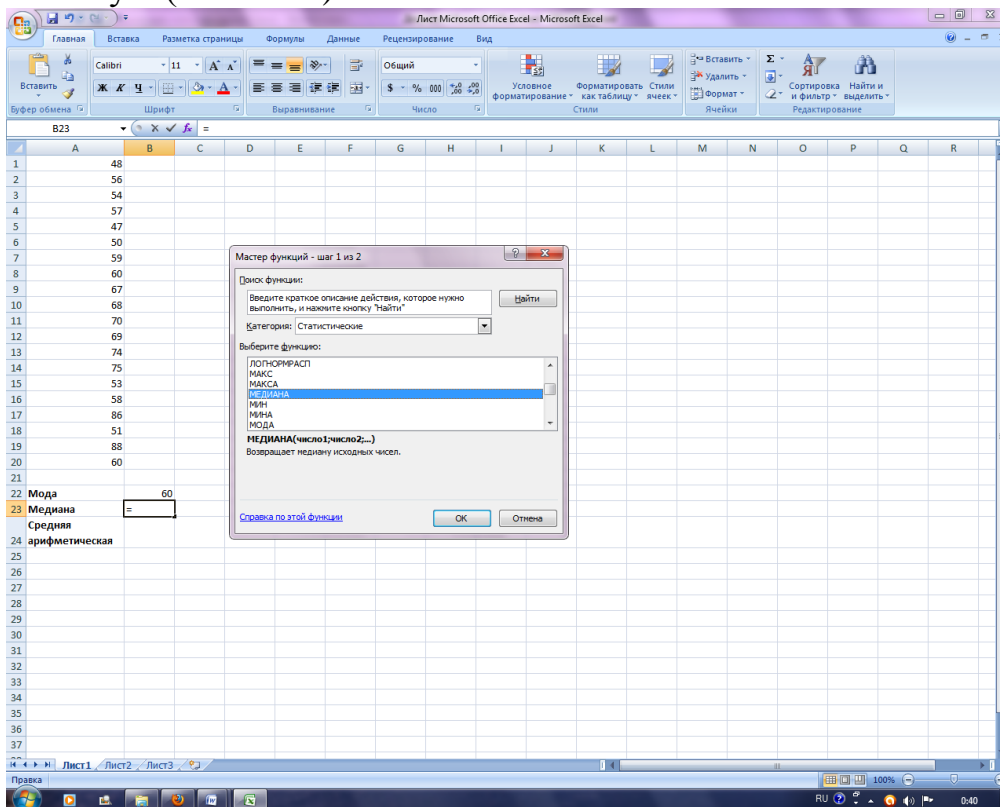


Рис. 5 Выбор функции Медиана

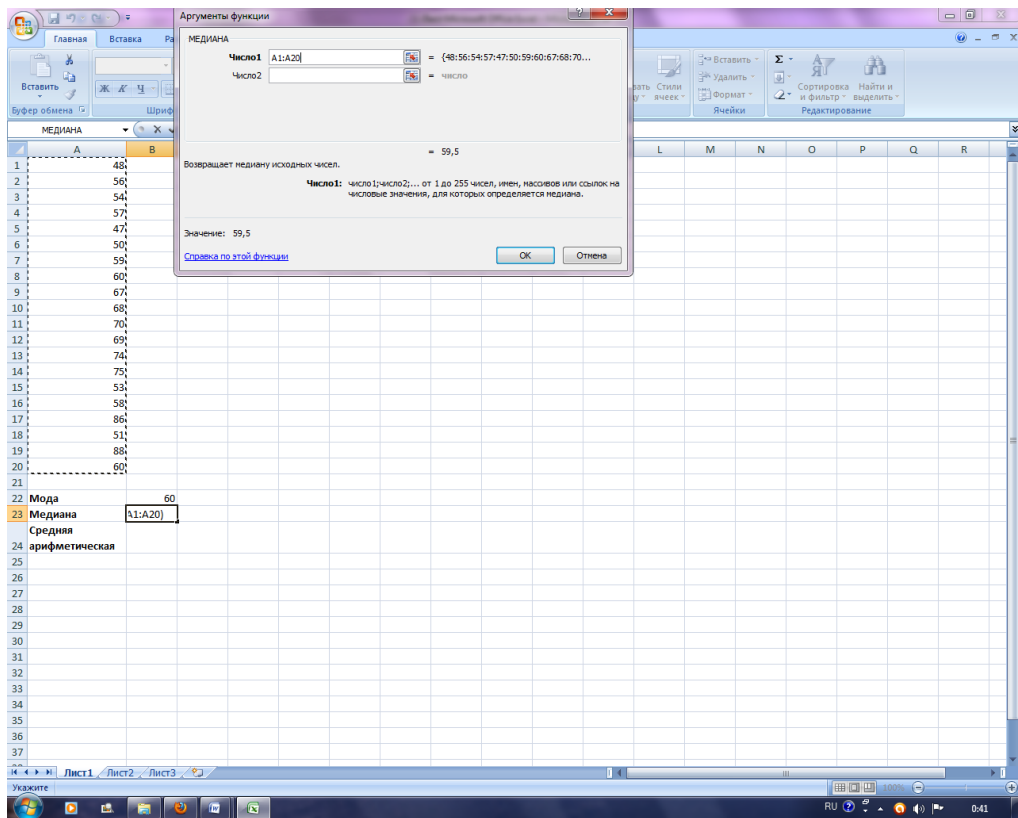


Рис. 6 Выбор массива

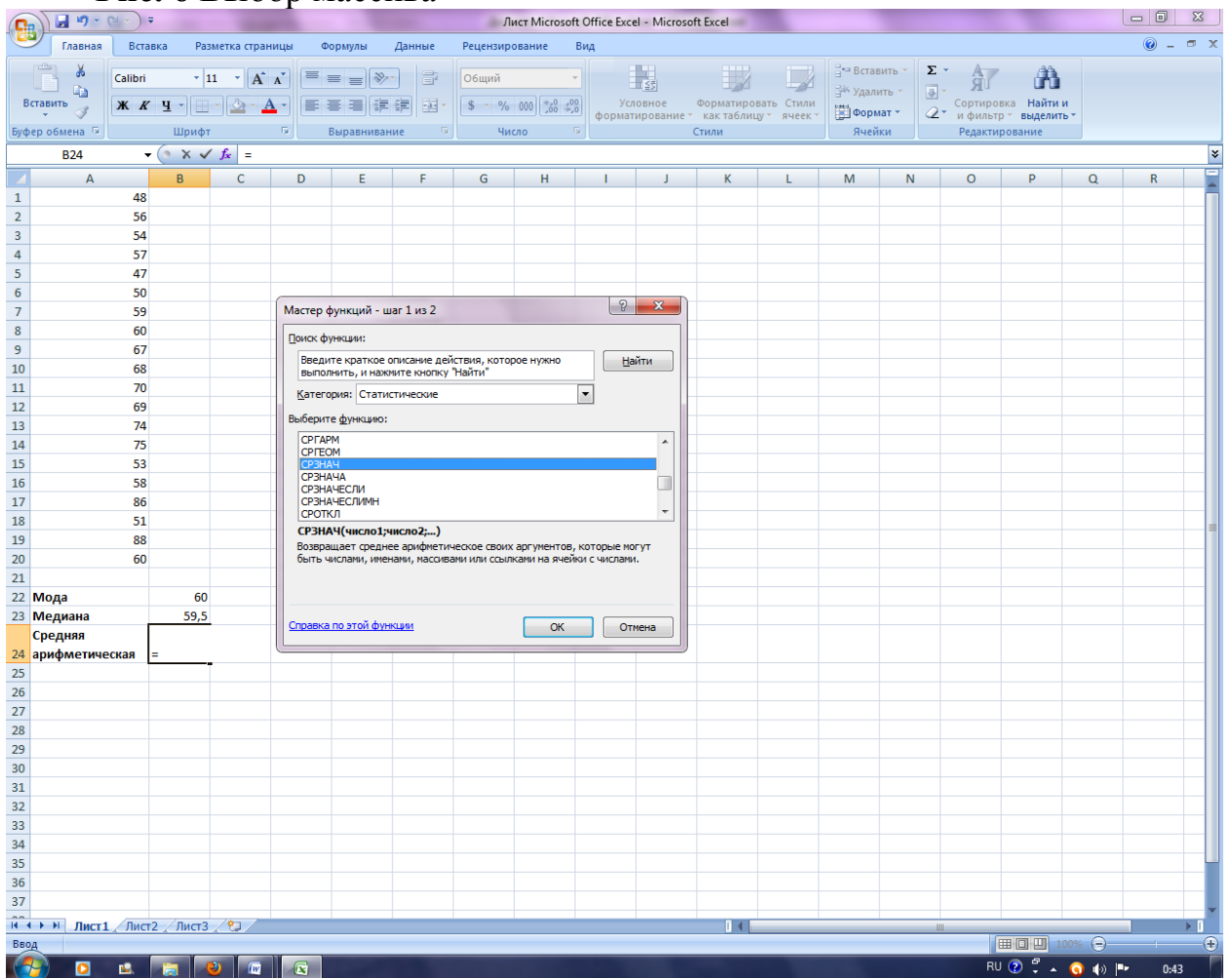


Рис. 7 Выбор функции Средзнач

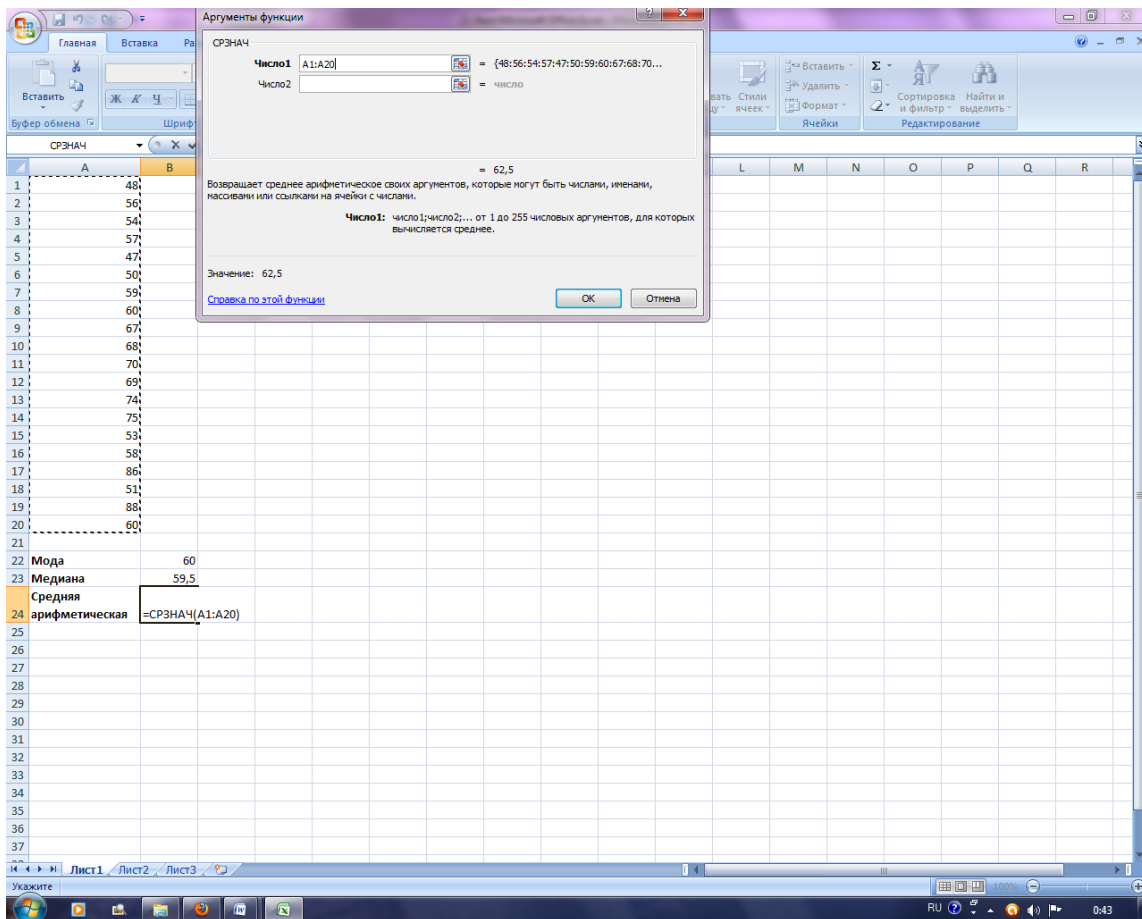


Рис. 8 Выбор массива

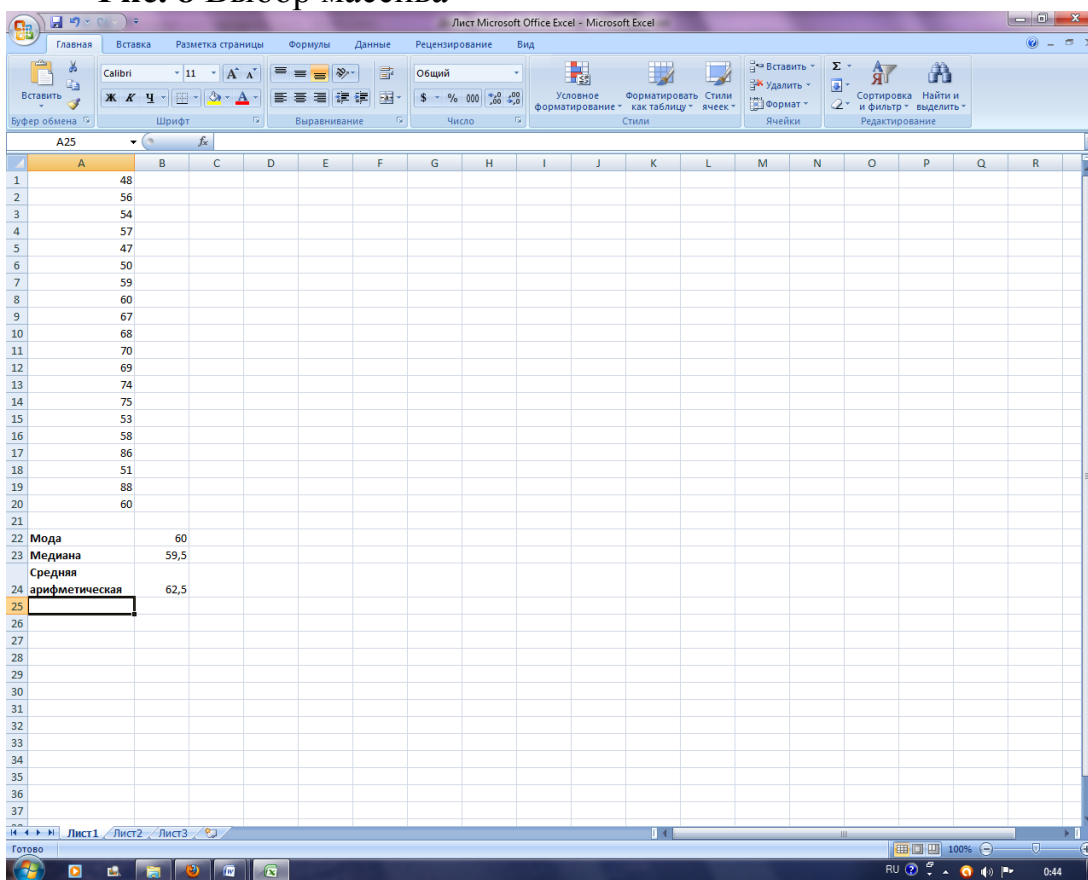


Рис. 9 Результат анализа

Итого, Медиана равна 59,5, а средняя арифметическая – 62,5 (Рис. 9).

Пример 2. Определить средний рост юношей 17 лет, среднеквадратичное отклонение, коэффициент вариации.

Рост, см (V)	Число юношей (p)
154-156	45
157-159	60
160-162	83
163-165	83
166-168	53
169-171	6
всего	330

Введем данные в таблицу Excel (Рис. 10).

Рост, см	Среднее значение интервала	Число юношей	Произведение среднего значения интервала на число наблюдений	Отклонение (разность) каждой варианты от среднearифметической величины ряда	Квадрат отклонения	Произведение квадрата отклонения на число наблюдений
V	V1	p	V1*p	d = V1 - M	d ²	d ² *p
154-156		45				
157-159		60				
160-162		83				
163-165		83				
166-168		53				
169-171		6				
всего		330				

Рис. 10 Ввод данных для расчета
 В ячейки В3:В8 внесем показатели центров интервалов роста (V1)
 (Рис. 11).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Рост, см	Среднее значение интервала	Число юношей	Произведение среднего значения интервала на число наблюдений	Отклонение (разность) каждой варианты от среднеарифметической величины ряда	Квадрат отклонения	Произведение квадрата отклонения на число наблюдений							
2	V	V1	p	V1*p	d = V1 - M	d ²	d ² *p							
3	154-156	155	45											
4	157-159	158	60											
5	160-162	161	83											
6	163-165	164	83											
7	166-168	167	53											
8	169-171	170	6											
9	всего		330											

Рис. 11 Ввод среднего значения интервала

В ячейку D3 вводим формулу $= B3 * C3$ (Рис. 12), копируем на остальные ячейки D4:D8 и полученные данные суммируем в ячейке D9. А в ячейку D10 введем формулу расчета средней арифметической $= D9 / C9$, после чего и получаем искомую величину взвешенной средней арифметической (Рис. 13).

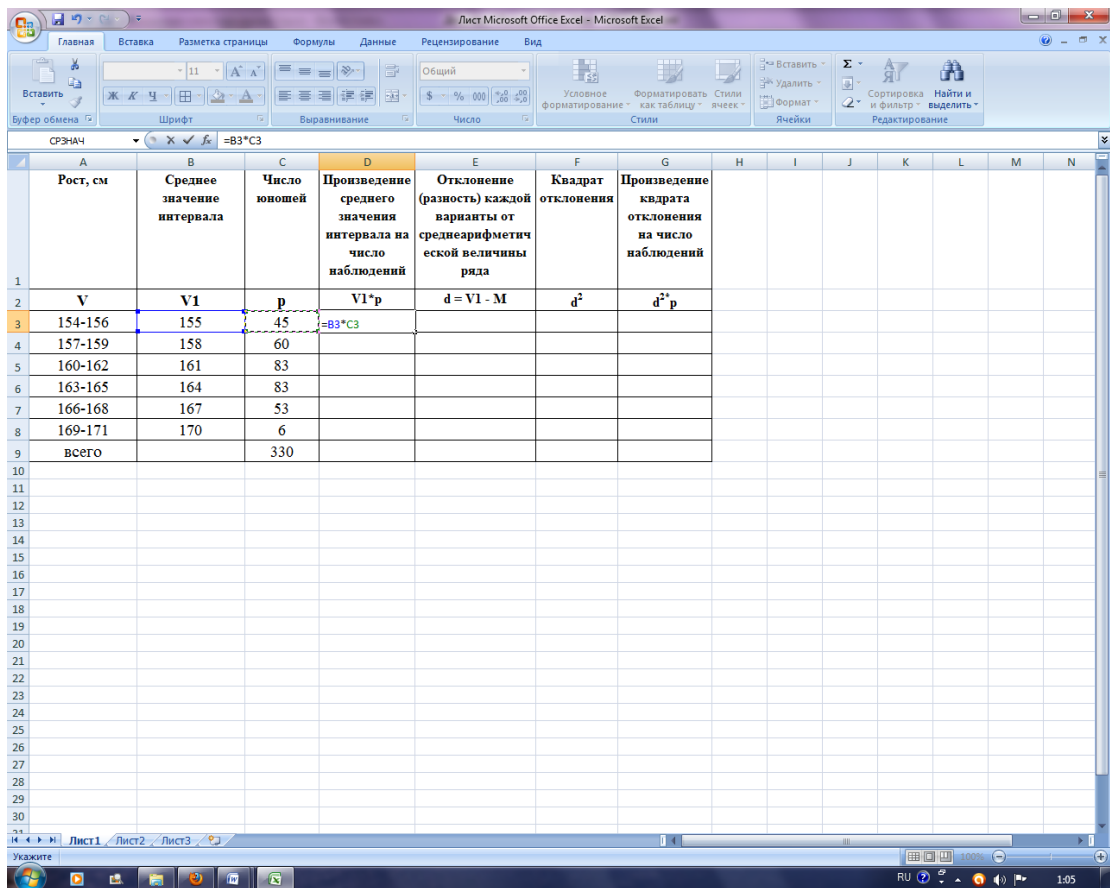


Рис. 12 Ввод формулы

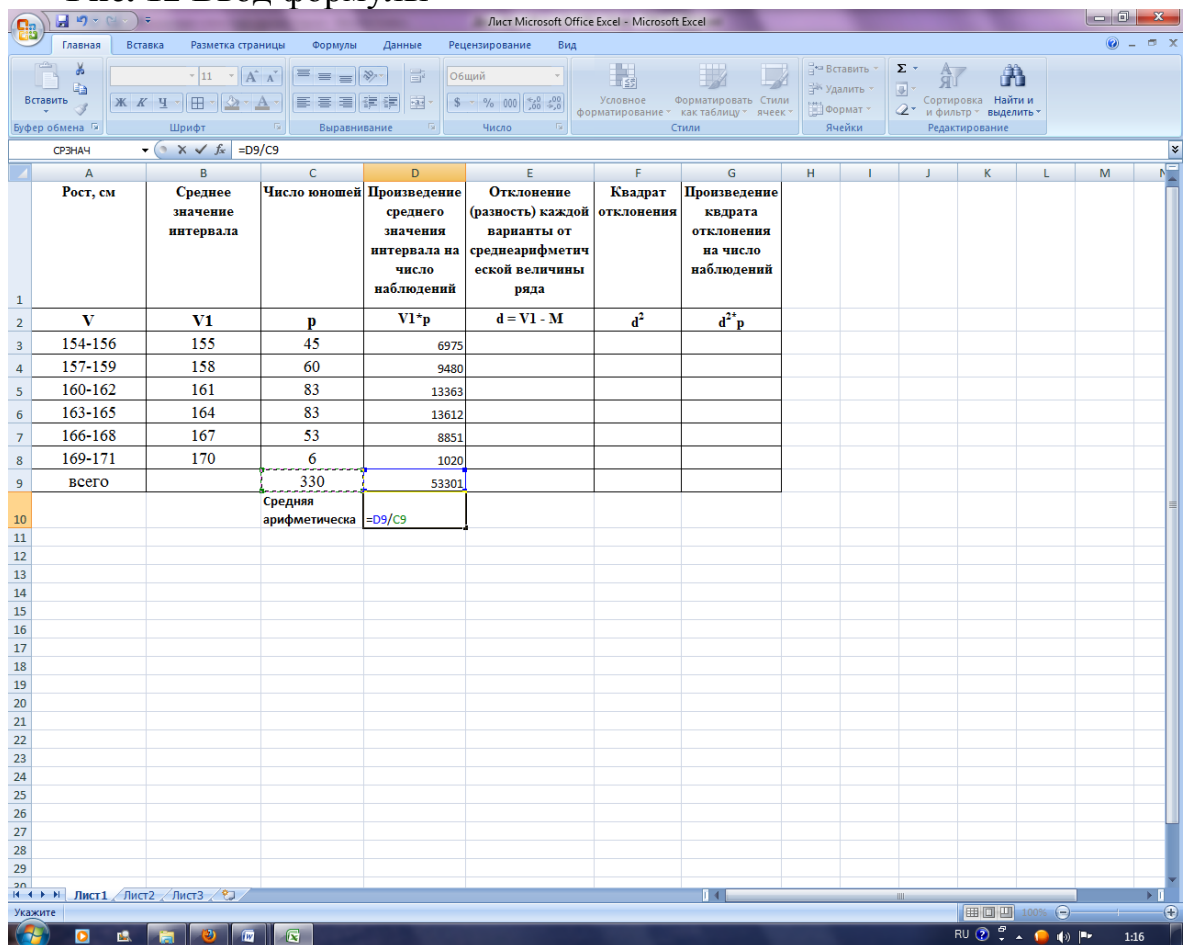


Рис. 13 Расчет средней арифметической величины

Для вычисления отклонения (разности) каждой варианты от средней арифметической величины вводим формулу в ячейку E3= $V3-D10$ (Рис. 14) и аналогично в остальные ячейки E4:E8. Находим квадрат отклонения и умножаем полученное число на число наблюдений.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Рост, см	Среднее значение интервала	Число юношей	Произведение среднего значения интервала на число наблюдений	Отклонение (разность) каждой варианты от среднеарифметической величины ряда	Квадрат отклонения	Произведение квадрата отклонения на число наблюдений						
1													
2	V	V1	p	V1*p	d = V1 - M	d²	d²*p						
3	154-156	155	45	6975	=B3-D10								
4	157-159	158	60	9480									
5	160-162	161	83	13363									
6	163-165	164	83	13612									
7	166-168	167	53	8851									
8	169-171	170	6	1020									
9	всего		330	53301									
10			Средняя арифметическая	161,5181818									

Рис. 14 Ввод формулы

В ячейке G9 суммируем полученные произведения квадрата отклонения на число наблюдений, делим на общее число наблюдений и находим дисперсию ($D11 = G9/C9$). В ячейке D12 вводим формулу расчета среднеквадратического отклонения – извлечение квадратного корня из величины дисперсии (Рис 15). Для получения квадратного корня щелкаем мышью на кнопке f_x в строке состояния, в открывшемся окне Мастера функций выбираем категорию Математические, находим функцию Корень. Нажимаем кнопку ОК и в открывшемся диалоговом окне указываем ячейку G9, из которой следует извлечь квадратный корень.

	A	B	C	D	E	F	G
	Рост, см	Среднее значение интервала	Число юношей	Произведение среднего значения интервала на число наблюдений	Отклонение (разность) каждой варианты от среднеарифметической величины ряда	Квадрат отклонения	Произведение квадрата отклонения на число наблюдений
1							
2	V	V1	p	V1*p	d = V1 - M	d ²	d ² *p
3	154-156	155	45	6975	-6,518181818	42,48669421	1911,90124
4	157-159	158	60	9480	-3,518181818	12,37760331	742,6561983
5	160-162	161	83	13363	-0,518181818	0,268512397	22,28652893
6	163-165	164	83	13612	2,481818182	6,159421488	511,2319835
7	166-168	167	53	8851	5,481818182	30,05033058	1592,667521
8	169-171	170	6	1020	8,481818182	71,94123967	431,647438
9	всего		330	53301			5212,390909
10			Средняя арифметическая	161,5181818			
11			Дисперсия	15,79512397			
12			Среднеквадратическое отклонение	=КОРЕНЬ(D11)			
13			Коэффициент вариации				

Рис. 15 Ввод формулы
 Для получения коэффициента вариации в ячейке D13 вводим формулу $=D12/D10*100$ (Рис. 16).

	A	B	C	D	E	F	G
	Рост, см	Среднее значение интервала	Число юношей	Произведение среднего значения интервала на число наблюдений	Отклонение (разность) каждой варианты от среднеарифметической величины ряда	Квадрат отклонения	Произведение квадрата отклонения на число наблюдений
1							
2	V	V1	p	V1*p	d = V1 - M	d ²	d ² *p
3	154-156	155	45	6975	-6,518181818	42,48669421	1911,90124
4	157-159	158	60	9480	-3,518181818	12,37760331	742,6561983
5	160-162	161	83	13363	-0,518181818	0,268512397	22,28652893
6	163-165	164	83	13612	2,481818182	6,159421488	511,2319835
7	166-168	167	53	8851	5,481818182	30,05033058	1592,667521
8	169-171	170	6	1020	8,481818182	71,94123967	431,647438
9	всего		330	53301			5212,390909
10			Средняя арифметическая	161,5181818			
11			Дисперсия	15,79512397			
12			Среднеквадратическое отклонение	3,974307986			
13			Коэффициент вариации	=D12/D10*100			

Рис. 16 Ввод формулы

Выводы:

4. Средний рост юношей составляет 1461,52 см;
5. $\sigma = \pm 3,97$ (см).
6. Величина коэффициента вариации, равная 2,46% свидетельствует о слабом разнообразии признака.

Таким образом, можно считать, что полученная средняя величина массы тела является достаточно представительной (типичной).

КОНТРОЛЬНЫЕ ВОПРОСЫ

- Что такое вариационный ряд?
- Для чего используются средние величины?
- По каким критериям можно оценить разнообразие признака?
- В каких случаях применяют среднее квадратическое отклонение?
- Каково назначение коэффициента вариации?
- Как оценить величину коэффициента вариации?
- Правило нормального распределения Гаусса.

ТЕСТОВЫЕ ЗАДАНИЯ:

Выберите один или несколько правильных ответов:

1. Вариационный ряд — это:

- а) числовые значения изучаемого признака статистической совокупности, расположенные в ранговом порядке;
- б) числовые значения изучаемого признака, расположенные в ранговом порядке с соответствующими этим значениям частотами;
- в) числовые значения изучаемого признака с соответствующими этим значениям частотами.

2. Средняя величина — это:

- а) варианта с повторяющимся числовым значением;
- б) варианта, имеющая наибольший «вес» (частоту) в вариационном ряду;
- в) обобщающая числовая характеристика размера изучаемого признака.

3. Впишите недостающий вид вариационного ряда по частоте представленных в нем вариант:

- а) простой;
- б) _____

4. Средние величины применяются для оценки:

- а) состояния здоровья населения;
- б) организации работы и деятельности лечебно-профилактических учреждений в целом, отдельных его подразделений и врачей;
- в) организации работы и деятельности всей системы здравоохранения;
- г) состояния окружающей среды.

5. В каком вариационном ряду используются следующие методы расчета средней арифметической величины (подберите соответствующие ответы):

Вариационный ряд:	Методы расчета:
1) простой вариационный ряд 2) взвешенный вариационный ряд	а) $M = (\text{сумма } V_p)/n$ б) $M = \text{сумма } V/n$

6. Укажите соответствующий алгоритм расчета для простых и взвешенных средних арифметических величин:

Средняя величина:	Алгоритм расчета:
1) простая средняя арифметическая величина 2) взвешенная средняя арифметическая величина	а) перемножить каждую варианту на соответствующую ей частоту (V_p); б) получить сумму произведений вариант на частоты (сумма V_p); г) суммировать числовые значения вариант (сумма V); д) полученную сумму V_p разделить на число наблюдений (n); г) полученную сумму V разделить на число наблюдений (n).

7. Характеристиками разнообразия вариационного ряда являются: а) лимиты ряда (V_{\max} и V_{\min});
б) амплитуда ряда (A);
в) среднеквадратическое отклонение (сигма);
г) отклонение (разность) каждой варианты от среднеарифметической величины вариационного ряда ($d = V - M$);
д) коэффициент вариации (C_v).

8. Каково значение сигмы для анализа вариационного ряда (укажите правильные ответы):

а) характеризует внутреннее разнообразие (колеблемость) вариационного ряда;
б) применяется для сравнительной оценки типичности средних арифметических величин;
в) позволяет оценить достоверность средней величины;
г) позволяет восстановить (реконструировать) вариационный ряд по частоте на основе правила «трех сигм»;
д) применяется для выявления «выскакивающих» вариант;
е) применяется для расчета коэффициента вариации (C_v);
ж) применяется для вычисления ошибки репрезентативности средней арифметической (m_M).

9. "Нормальное" распределение вариационного ряда означает:

- а) распределение вариационного ряда по частоте на основе правила «трех сигм»;
- б) что в пределах $M \pm 1$ сигма находится 68,3% вариант ряда;
- в) что в пределах $M \pm 2$ сигмы находятся 95,5% всех вариант;
- г) что в пределах $M \pm 3$ сигмы находятся 99,7% всех вариант.

СИТУАЦИОННЫЕ ЗАДАЧИ:

Задача 1 Результаты исследования здоровья студентов 2 групп по характеристике частоты сердечных сокращений (ЧСС) показали одинаковую среднюю величину (85 уд/мин). Критерий разнообразия ЧСС в одной группе — 2 удара в минуту, в другой — 4 удара в минуту.

1. Определите, для какой группы средняя величина пульса при одинаковой средней частоте сердечных сокращений (M) и одинаковом числе студентов типичнее, т.е. лучше отражает состояние здоровья студентов.
2. Какой критерий разнообразия был использован для определения разнообразия признака?

Задача 2 При изучении физического развития школьников 7-го класса было установлено значительное разнообразие по росту (от 151 до 170 см). Средняя величина роста этих мальчиков равна 160 см, $\sigma = \pm 3$ см.

1. Находятся ли крайние значения роста детей в пределах нормального распределения признака?
2. Какую методику (значение сигмы) Вы при этом использовали?

Задача 3 При медицинском осмотре студентов военно-медицинской академии изучены различные показатели крови, в т.ч. количество лейкоцитов колебалось в пределах 6000-9500. Среднее значение числа лейкоцитов равно 7500, $\sigma = \pm 0,5$ тыс. лейкоцитов.

1. Какая величина в данном случае является «выскакивающей вариантой»?
2. Какая методика позволила определить ее?

Задача 4 В первые часы после инфаркта миокарда у больных изменяется целый ряд параметров, в том числе уровень артериального давления количество лейкоцитов и ферментов крови.

1. Какой критерий необходимо применить для оценки разнообразия признаков?
2. Обоснуйте его применение.

Задача 5 При проведении всеобщей диспансеризации детского населения в городе Н. были получены результаты изучения физического развития детей (по массе тела). При этом получили следующие данные: средняя масса тела новорожденных детей составила 2,9 кг, $\sigma = \pm 0,3$ кг; средняя масса тела детей 1-го года жизни — 12 кг, $\sigma = \pm 1,0$ кг.

1. Достаточно ли представленной в условии задачи информации для вывода о степени разнообразия (устойчивости) признака?
2. В какой группе более разнообразна масса тела?

Основная литература:

1. Кобринский Б.А., Зарубина Т.В. Медицинская информатика: Учебник. М: изд. "Академия", 2009.
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: Учебное пособие для практических занятий / Под ред. В.З.Кучеренко. – М.:ГЭОТАР-МЕД, 2004. – 192 с.

Дополнительная литература:

1. Зайцев В.М., Лифляндский В.Г. Маринкин В.И. Прикладная медицинская статистика. – Спб: ООО «Издательства ФОЛИАНТ», 2003. – 432 с.
2. Карась С.И. Информационные основы принятия решений в медицине: Учебное пособие. – Томск: Печатная мануфактура, 2003.- 145с.
3. Чернов В.И., Родионов О.В., Есауленко И.Э. и др. Медицинская информатика: Учебное пособие.- Воронеж, 2004. – 282с.: ил.
5. Гельман В.Я. Медицинская информатика: практикум. – СПб: Питер, 2001. -480с. – (Серия "Национальная медицинская библиотека").
6. Богданов А.К., Проценко В.Д. Практические применения современных методов анализа изображений в медицине: Учебное пособие. – М.: РУДН, 2008. – 119с.: ил.
7. Санников А.Г., Егоров Д.Б., Скудных А.С., Рухлова С.А. Практикум по медицинской информатике: автоматизированное рабочее место врача и системы поддержки принятия врачебного решения. – Тюмень: П.П.Ш., 2009. – 116с.

Оглавление

Цель занятия	3
Студент должен уметь	3
Студент должен знать:.....	3
Информационный материал.....	4
Функциональная связь	4
Корреляционная связь	4
Практическое значение установления корреляционной связи.	4
Величина, характеризующая направление и силу связи между признаками.	4
Способы представления корреляционной связи.....	5
Направление корреляционной связи.....	5
Сила корреляционной связи.....	5
Методы определения коэффициента корреляции и формулы	5
Методические требования к использованию коэффициента корреляции	5
Рекомендации по применению метода ранговой корреляции (метод Спирмена)	5
Рекомендации к применению метода квадратов (метод Пирсона).....	5
Методика и порядок вычисления коэффициента корреляции	5
1) Метод квадратов	5
2) Ранговый метод.....	6
Схема оценки корреляционной связи по коэффициенту корреляции.....	6
Вычисление ошибки коэффициента корреляции	6
Оценка достоверности коэффициента корреляции, полученного методом ранговой корреляции и методом квадратов	7
Расчет коэффициента корреляции средствами MS Excel.....	11
Контрольные вопросы	13
Тестовые задания	14
Задачи	15
Приложение	17

Хронология занятия

1. Формулировка и обоснования цели занятия (10 мин.);
2. Изложение основных вопросов темы (60 мин.);
3. Перерыв (20 мин.)
4. Самостоятельная работа студентов с методическим материалом - (30 мин.)
5. Разбор типовых задач по изучаемой теме (20 мин.)
6. Самостоятельное решение задач (40 мин.)
7. Тестовый контроль на ПК (15 мин.)
8. Общая продолжительность занятия – 195 минут.

Цель занятия

на основе применения методов корреляции уметь выявлять влияние факторного признака на результативный при анализе общественного здоровья и деятельности медицинских учреждений.

Студент должен уметь

- устанавливать корреляционную зависимость методом квадратов и ранговой корреляции;
- рассчитывать коэффициент корреляции в MS Excel;
- оценивать силу, направление и достоверность полученного коэффициента корреляции и делать соответствующие выводы.

Студент должен знать:

- виды проявления количественных связей;
- понятие функциональной и корреляционной зависимости;
- практическое значение установления корреляционной связи;
- характеристики коэффициента корреляции;
- методику и порядок определения коэффициента корреляции.

Место проведения: аудитория кафедры общественного здоровья и организации здравоохранения с курсом медицинской информатики, дисплейный класс.

Оснащение занятия

Мультимедийный проектор

Ноутбук

Наглядный материал в виде мультимедийной презентации

Персональный компьютер

Информационный материал

При изучении общественного здоровья и здравоохранения в научных и практических целях исследователю часто приходится проводить статистический анализ связей между факторными и результативными признаками статистической совокупности (причинно-следственная связь) или определение зависимости параллельных изменений нескольких признаков этой совокупности от какой либо третьей величины (от общей их причины). Необходимо уметь изучать особенности этой связи, определять ее размеры и направление, а также оценивать ее достоверность. Для этого используются методы корреляции.

1. **Виды проявления количественных связей между признаками**
 - функциональная связь
 - корреляционная связь
2. **Определения функциональной и корреляционной связи**

Функциональная связь

такой вид соотношения между двумя признаками, когда каждому значению одного из них соответствует строго определенное значение другого (площадь круга зависит от радиуса круга и т.д.). Функциональная связь характерна для физико-математических процессов.

Корреляционная связь

такая связь, при которой каждому определенному значению одного признака соответствует несколько значений другого взаимосвязанного с ним признака (связь между ростом и массой тела человека; связь между температурой тела и частотой пульса и др.). Корреляционная связь характерна для медико-биологических процессов.

Практическое значение установления корреляционной связи.

Выявление причинно-следственной между факторными и результативными признаками (при оценке физического развития, для определения связи между условиями труда, быта и состоянием здоровья, при определении зависимости частоты случаев болезни от возраста, стажа, наличия производственных вредностей и др.)

Зависимость параллельных изменений нескольких признаков от какой-то третьей величины. Например, под воздействием высокой температуры в цехе происходят изменения кровяного давления, вязкости крови, частоты пульса и др.

Величина, характеризующая направление и силу связи между признаками.

Коэффициент корреляции, который одним числом дает представление о направлении и силе связи между признаками (явлениями), пределы его колебаний от 0 до ± 1

Способы представления корреляционной связи

- график (диаграмма рассеяния)
- коэффициент корреляции

Направление корреляционной связи

- прямая
- обратная

Сила корреляционной связи

- сильная: $\pm 0,7$ до ± 1
- средняя: $\pm 0,3$ до $\pm 0,699$
- слабая: 0 до $\pm 0,299$

Методы определения коэффициента корреляции и формулы

- метод квадратов (метод Пирсона)
- ранговый метод (метод Спирмена)

Методические требования к использованию коэффициента корреляции

- измерение связи возможно только в качественно однородных совокупностях (например, измерение связи между ростом и весом в совокупностях, однородных по полу и возрасту)
- расчет может производиться с использованием абсолютных или производных величин
- для вычисления коэффициента корреляции используются не сгруппированные вариационные ряды (это требование применяется только при вычислении коэффициента корреляции по методу квадратов)
- число наблюдений менее 30

Рекомендации по применению метода ранговой корреляции (метод Спирмена)

- когда нет необходимости в точном установлении силы связи, а достаточно ориентировочных данных
- когда признаки представлены не только количественными, но и атрибутивными значениями
- когда ряды распределения признаков имеют открытые варианты (например, стаж работы до 1 года и др.)

Рекомендации к применению метода квадратов (метод Пирсона)

- когда требуется точное установление силы связи между признаками
- когда признаки имеют только количественное выражение

Методика и порядок вычисления коэффициента корреляции

1) Метод квадратов

- построить вариационные ряды для каждого из сопоставляемых признаков, обозначив первый и второй ряд чисел соответственно x и y ;
- определить для каждого вариационного ряда средние значения (M_1 и M_2);
- найти отклонения (d_x и d_y) каждого числового значения от среднего значения своего вариационного ряда;
- полученные отклонения перемножить ($d_x \times d_y$)

- каждое отклонение возвести в квадрат и суммировать по каждому ряду ($\sum d_x^2$ и d_y^2)
- подставить полученные значения в формулу расчета коэффициента корреляции:

$$r_{xy} = \frac{\sum(d_x \times d_y)}{\sqrt{(\sum d_x^2 \times \sum d_y^2)}}$$

при наличии вычислительной техники расчет производится по формуле:

$$r_{xy} = \frac{n \sum xy - \sum x \times \sum y}{\sqrt{[(n \sum x^2 - \sum x^2 / n)(n \sum y^2 - \sum y^2 / n)]}}$$

2) Ранговый метод

- составить два ряда из парных сопоставляемых признаков, обозначив первый и второй ряд соответственно x и y. При этом представить первый ряд признака в убывающем или возрастающем порядке, а числовые значения второго ряда расположить напротив тех значений первого ряда, которым они соответствуют

- величину признака в каждом из сравниваемых рядов заменить порядковым номером (рангом). Рангами, или номерами, обозначают места показателей (значения) первого и второго рядов. При этом числовым значениям второго признака ранги должны присваиваться в том же порядке, какой был принят при раздаче их величинам первого признака. При одинаковых величинах признака в ряду ранги следует определять как среднее число из суммы порядковых номеров этих величин

- определить разность рангов между x и y (d): $d = x - y$
- возвести полученную разность рангов в квадрат (d^2)
- получить сумму квадратов разности ($\sum d^2$) и подставить полученные значения в формулу:

$$r_{xy} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Схема оценки корреляционной связи по коэффициенту корреляции

Сила связи	Направление связи	
	прямая (+)	обратная (-)
Сильная	от + 1 до +0,7	от - 1 до - 0,7
Средняя	от + 0,699 до + 0,3	от - 0,699 до - 0,3
Слабая	от + 0,299 до 0	от - 0,299 до 0

Вычисление ошибки коэффициента корреляции

- ошибка коэффициента корреляции, вычисленного методом квадратов (Пирсона):

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}$$

ошибка коэффициента корреляции, вычисленного ранговым методом (Спирмена):

$$m_{p_{xy}} = \sqrt{\frac{1 - p_{xy}^2}{n - 2}}$$

Оценка достоверности коэффициента корреляции, полученного методом ранговой корреляции и методом квадратов

Способ 1 Достоверность определяется по формуле:

$$t = \frac{r_{xy}}{m_{r_{xy}}} \quad \text{или} \quad t = \frac{p_{xy}}{m_{p_{xy}}}$$

Критерий t оценивается по таблице значений t с учетом числа степеней свободы ($n - 2$), где n — число парных вариантов. Критерий t должен быть равен или больше табличного, соответствующего вероятности $p \geq 99\%$.

Способ 2 Достоверность оценивается по специальной таблице стандартных коэффициентов корреляции. При этом достоверным считается такой коэффициент корреляции, когда при определенном числе степеней свободы ($n - 2$), он равен или более табличного, соответствующего степени безошибочного прогноза $p \geq 95\%$.

Задача - эталон

на применение метода квадратов

Задание: вычислить коэффициент корреляции, определить направление и силу связи между количеством кальция в воде и жесткостью воды, если известны следующие данные (табл. 1). Оценить достоверность связи. Сделать вывод.

Таблица 1

Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)
4	28
8	56
11	77
27	191
34	241
37	262

Обоснование выбора метода. Для решения задачи выбран метод квадратов (Пирсона), т.к. каждый из признаков (жесткость воды и количество кальция) имеет числовое выражение; нет открытых вариантов.

Решение.

Последовательность расчетов изложена в тексте, результаты представлены в таблице. Построив ряды из парных сопоставляемых признаков, обозначить их через x (жесткость воды в градусах) и через y (количество кальция в воде в мг/л).

Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)	d_x	d_y	$d_x \times d_y$	d_x^2	d_y^2
4	28	-16	-114	1824	256	12996
8	56	-12	-86	1032	144	7396
11	77	-9	-66	594	81	4356
27	191	+7	+48	336	49	2304
34	241	+14	+98	1372	196	9604
37	262	+16	+120	1920	256	14400
$M_x = \Sigma x / n$	$M_y = \Sigma y / n$			$\Sigma d_x \times d_y = 7078$	$\Sigma d_x^2 = 982$	$\Sigma d_y^2 = 51056$
$M_x = 120/6 = 20$	$M_y = 852/6 = 142$					

1. Определить средние величины M_x ряду вариант "x" и M_y в ряду вариант "y" по формулам:

$$M_x = \Sigma x / n \text{ (графа 1) и}$$

$$M_y = \Sigma y / n \text{ (графа 2)}$$

2. Найти отклонение (d_x и d_y) каждой варианты от величины вычисленной средней в ряду "x" и в ряду "y"

$$d_x = x - M_x \text{ (графа 3) и } d_y = y - M_y \text{ (графа 4).}$$

3. Найти произведение отклонений $d_x \times d_y$ и суммировать их: $\Sigma d_x \times d_y$ (графа 5)

4. Каждое отклонение d_x и d_y возвести в квадрат и суммировать их значения по ряду "x" и по ряду "y": $\Sigma d_x^2 = 982$ (графа 6) и $\Sigma d_y^2 = 51056$ (графа 7).

5. Определить произведение $\Sigma d_x^2 \times \Sigma d_y^2$ и из этого произведения извлечь квадратный корень

$$\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)} = \sqrt{(982 \times 51056)}$$

6. Полученные величины $\Sigma (d_x \times d_y)$ и $\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}$ подставляем в формулу расчета коэффициента корреляции:

$$r_{xy} = \frac{\Sigma (d_x \times d_y)}{\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}} = \frac{7078}{\sqrt{(982 \times 51056)}} = \frac{7078}{\sqrt{50136992}} = \frac{7078}{7080,7} = +0,99$$

7. Определить достоверность коэффициента корреляции: 1-й способ. Найти ошибку коэффициента корреляции ($m_{r_{xy}}$) и критерий t по формулам:

$$m_{r_{xy}} = \pm \sqrt{\frac{1 - 0,99^2}{6 - 2}} = \pm \sqrt{\frac{0,02}{4}} = \sqrt{0,005} = \pm 0,07$$

$$t = \frac{0,99}{0,07} = 14,1$$

Критерий $t = 14,1$, что соответствует вероятности безошибочного прогноза $p > 99,9\%$.

2-й способ. Достоверность коэффициента корреляции оценивается по таблице "Стандартные коэффициенты корреляции" (см. приложение 1). При числе степеней свободы $(n - 2) = 6 - 2 = 4$, наш расчетный коэффициент корреляции $r_{xy} = + 0,99$ больше табличного ($r_{\text{табл}} = + 0,917$ при $p = 99\%$).

Вывод. Чем больше кальция в воде, тем она более жесткая (связь **прямая, сильная и достоверная**: $r_{xy} = + 0,99$, $p > 99,9\%$).

Задача - эталон

на применение рангового метода

Задание: методом рангов установить направление и силу связи между стажем работы в годах и частотой травм, если получены следующие данные:

Стаж работы в годах	Число травм на 100 работающих
до 1 года	24
1-2	16
3-4	12
5-6	12
7 и более	6

Обоснование выбора метода: для решения задачи может быть выбран только метод ранговой корреляции, т.к. первый ряд признака "стаж работы в годах" имеет открытые варианты (стаж работы до 1 года и 7 и более лет), что не позволяет использовать для установления связи между сопоставляемыми признаками более точный метод — метод квадратов.

Решение. Последовательность расчетов изложена в тексте, результаты представлены в табл. 2.

Таблица 2

Стаж работы в годах	Число травм	Порядковые номера (ранги)		Разность рангов	Квадрат разности рангов
		X	Y		
До 1 года	24	1	5	-4	16
1-2	16	2	4	-2	4
3-4	12	3	2,5	+0,5	0,25
5-6	12	4	2,5	+1,5	2,25
7 и более	6	5	1	+4	16
					$\Sigma d^2 = 38,5$

1. Каждый из рядов парных признаков обозначить через "x" и через "y" (графы 1—2).

2. Величину каждого из признаков заменить ранговым (порядковым) номером. Порядок раздачи рангов в ряду "x" следующий: минимальному значению признака (стаж до 1 года) присвоен порядковый номер "1", последующим вариантам этого же ряда признака соответственно в порядке увеличения 2-й, 3-й, 4-й и 5-й порядковые номера — ранги (см. графу 3).

Аналогичный порядок соблюдается при раздаче рангов второму признаку "y" (графа 4).

В тех случаях, когда встречаются несколько одинаковых по величине вариант (например, в задаче-эталоне это 12 и 12 травм на 100 работающих при стаже 3—4 года и 5—6 лет, порядковый номер обозначить средним числом из суммы их порядковых номеров. Эти данные о числе травм (12 травм) при ранжировании должны занимать 2 и 3 места, таким образом среднее число из них равно $(2 + 3)/2 = 2,5$.

Таким образом, числу травм "12" и "12" (признаку) следует раздать ранговые номера одинаковые — "2,5" (графа 4).

3. Определить разность рангов $d = (x - y)$ — (графа 5)

4. Разность рангов возвести в квадрат (d^2) и получить сумму квадратов разности рангов Σd^2 (графа 6).

5. Произвести расчет коэффициента ранговой корреляции по формуле:

$$r_{xy} = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)},$$

где n — число сопоставляемых пар вариант в ряду "x" и в ряду "y"

$$\rho_{xy} = 1 - \frac{6 \times 38,5}{5(5^2 - 1)} = 1 - \frac{325}{5(25 - 1)} = 1 - \frac{325}{120} = 1 - 2,708\bar{3} = -0,708\bar{3}$$

6. Определить достоверность коэффициента ранговой корреляции.

1-й способ. Определить ошибку ($m\rho_{xy}$) коэффициента ранговой корреляции и оценить достоверность его с помощью критерия t:

$$m\rho_{xy} = \sqrt{\frac{1 - 0,92^2}{n - 2}} = \sqrt{\frac{0,08}{3}} = \sqrt{0,026} = 0,16$$

$$t = \frac{0,92}{0,16} = 5,75$$

Полученный критерий $t = 5,75$ соответствует вероятности безошибочного прогноза (p) больше 95 %:

$$\rho_{xy} = -0,92; m\rho_{xy} = \pm 0,16; t = 5,75; p > 95\%$$

2-й способ. По таблице "Стандартных коэффициентов корреляции": при числе степеней свободы $(n - 2) = 5 - 2 = 3$ наш расчетный коэффициент корреляции $\rho_{xy} = -0,92$ больше табличного 0,878 и меньше 0,933, что соответствует вероятности безошибочного прогноза больше 95% и меньше 98%. Это позволяет считать полученный коэффициент ранговой корреляции достоверным.

Вывод. С вероятностью безошибочного прогноза (p) больше 95% установлено, что чем больше стаж работы, тем меньше частота травм (связь обратная, сильная, достоверная корреляционная: $\rho_{xy} = -0,92, p > 95\%$).

Расчет коэффициента корреляции средствами MS Excel

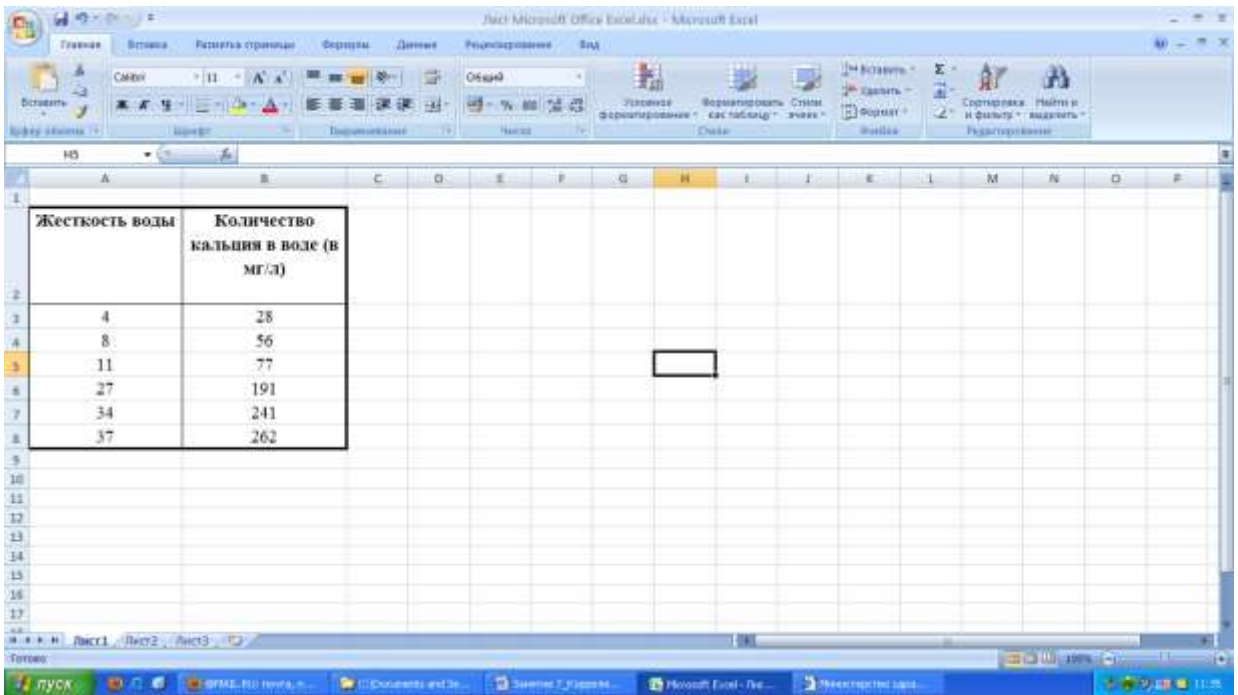
Задание: вычислить коэффициент корреляции, определить направление и силу связи между количеством кальция в воде и жесткостью воды, если известны следующие данные (табл. 1). Оценить достоверность связи. Сделать вывод.

Таблица 1

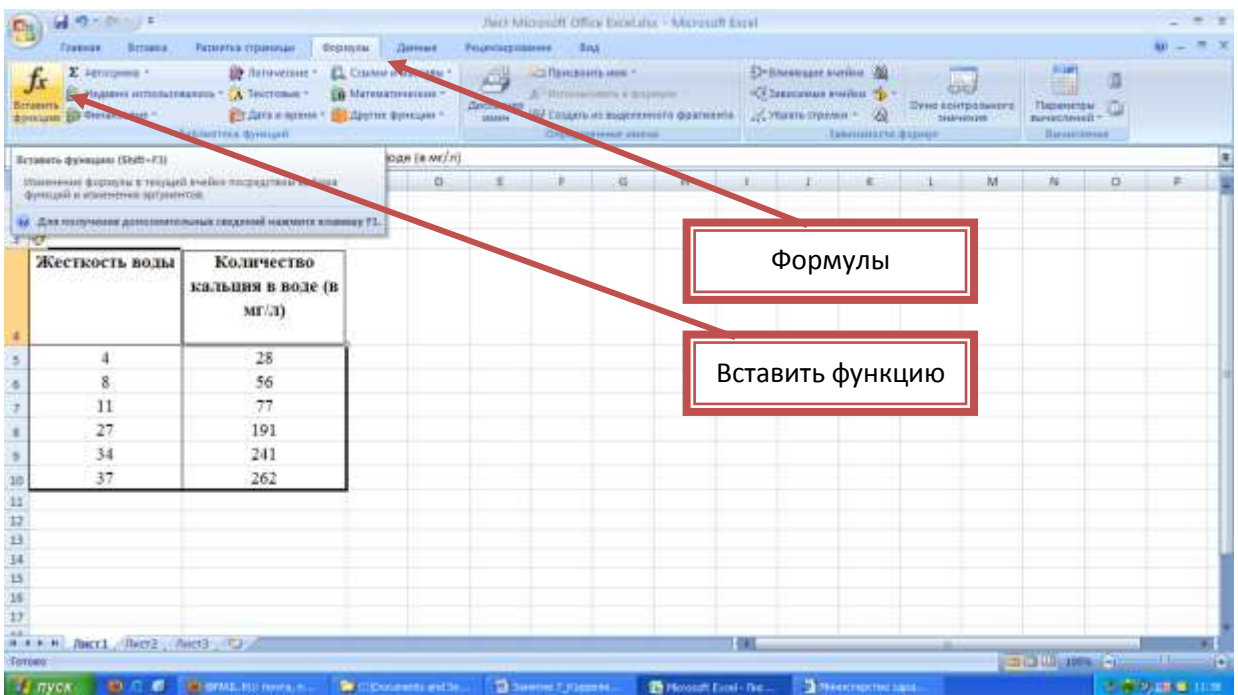
Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)
4	28
8	56
11	77
27	191
34	241
37	262

Решение:

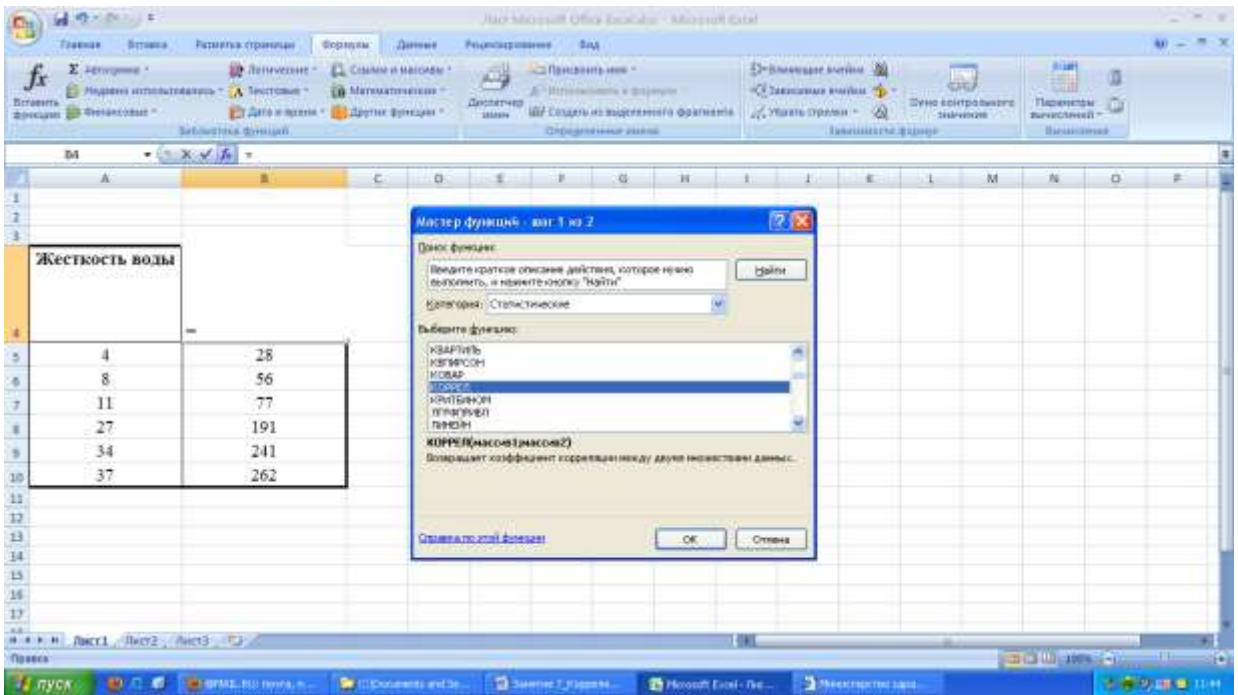
1 этап: ввод информации:



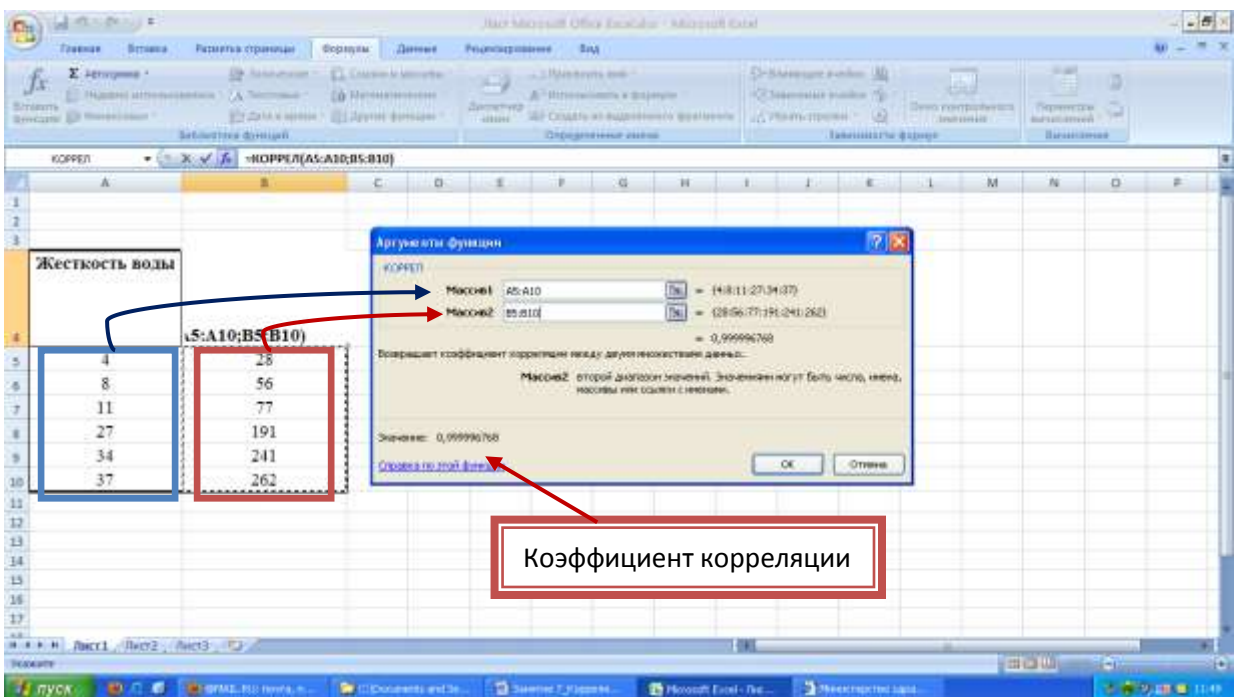
2 этап – выбор вкладки «Формулы» и «Вставка функции f_x »:



3 этап – из категории статистических функций, выбираем «Коррел» (коэффициент корреляции между двумя множествами данных):



4 этап – внесение массивов данных и расчет коэффициента корреляции:



5 этап: Вывод. Чем больше кальция в воде, тем она более жесткая (связь прямая, сильная и достоверная: $r_{xy} = + 0,99$).

Контрольные вопросы

1. Дайте определение функциональной и корреляционной связи.

2. Приведите примеры прямой и обратной корреляционной связи.
3. Укажите размеры коэффициентов корреляции при слабой, средней и сильной связи между признаками.
4. В каких случаях применяется ранговый метод вычисления коэффициента корреляции?
5. В каких случаях применяется метод квадратов?
6. Каковы основные этапы вычисления коэффициента корреляции методом квадратов?
7. Каковы основные этапы вычисления коэффициента корреляции ранговым методом?
8. Укажите способы определения достоверности коэффициента корреляции.

Тестовые задания

При проведении корреляционного анализа необходимо учитывать следующие параметры:

- Направление связи между признаками, её силу и ошибку репрезентативности;
- Направление связи между признаками, её силу, ошибку репрезентативности и величину коэффициента вариации;
- Направление связи между признаками, её силу, ошибку репрезентативности и величину критерия достоверности.

Между какими из ниже перечисленных признаков может устанавливаться корреляционная связь:

- Ростом и массой тела у детей;
- Содержанием кислорода в клетках крови и уровнем осмотического давления;
- Уровнем систолического и диастолического давления;
- Частотой случаев хронических заболеваний и возрастом.

Коэффициент корреляции между уровнем шума и снижением слуха с учетом стажа у рабочих механосборочного цеха равен

$r_{xy} + 0,91$. Установленная связь:

- Обратная и слабая;
- Обратная и сильная;
- Прямая и слабая;
- Прямая и сильная.

Получить представление о силе и направлении связи между признаками можно с помощью:

- таблиц, в которых записаны размеры признаков;
- графического изображения зависимости;
- коэффициента корреляции.

Направление корреляционной зависимости может быть представлено с помощью:

- таблиц, в которых записаны размеры признаков;
- графического изображения зависимости;
- коэффициента корреляции.

Задачи

Задача 1. В связи с ростом ревматизма в районе А. врач провел обследование семей жителей своего участка с целью выявления носителей стрептококковой инфекции в каждой семье. Специалист Роспотребнадзора оценил санитарно-гигиеническую характеристику жилищных условий этих семей (см. табл.).

1. Определите, какой метод позволит установить корреляцию между факторным признаком и результативным?
2. Обоснуйте свой вывод.

Жилищные условия	Носительство стрептококковой инфекции (на 100 обследованных)
Очень плохие	12
Плохие	8
Удовлетворительные	6
Хорошие	6
Наиболее благоприятные	2

Задача 2. В городе Н. было проведено изучение зависимости заболеваемости инфарктом миокарда по месяцам года в зависимости от средней температуры воздуха:

Месяцы года	Заболеваемость инфарктом миокарда по месяцам (на 10 тыс. жителей)	Среднемесячная температура воздуха
Январь	1,6	-7,1
Февраль	1,23	-7,7
Март	1,14	-5,8
Апрель	1,13	-4,1
Май	1,12	+13
Июнь	1,02	+14,9
Июль	0,91	+18,8
Август	0,82	+15,6
Сентябрь	1,06	+9,0
Октябрь	1,22	+6,0
Ноябрь	1,33	-1,0
Декабрь	1,4	-7,7

Какой из методов корреляции следует применять для установления связи? Обоснуйте свой вывод.

Задача 3. Между стажем работы ткачих и частотой понижения слуха у них установлена прямая корреляционная связь ($r_{xy}=+0.8$)/ Ошибка коэффициента корреляции $\pm 0,1$.

Оцените коэффициент корреляции.

Какая дополнительная информация необходима для оценки достоверности этой связи?

Задача 4. В научном исследовании между частотой материнской смертности и частотой внебольничного аборта установлена корреляционная зависимость.

Какой метод корреляции более предпочтителен для установления связи в данной ситуации?

Назовите факторные и результативные признаки.

Задача 5. В трех районах города N. Проводилось изучение заболеваемости кариесом детей в зависимости от содержания фтора в питьевой воде. При этом была установлена связь ($r_{xy}=-0.85$).

Оцените силу и направление связи.

Можно ли утверждать, что при едином централизованном водоснабжении эта закономерность характерна для заболеваемости кариесом детей всего города?

Является ли условие задачи достаточным для такого утверждения?

Приложение

**Стандартные коэффициенты корреляции, которые считаются
достоверными (по Л.С. Каминскому)**

Число степеней свободы — 2	Уровень вероятности р (%)		
	95%	98%	99%
1	0,997	0,999	0,999
2	0,950	0,980	0,990
3	0,878	0,934	0,959
4	0,811	0,882	0,917
5	0,754	0,833	0,874
6	0,707	0,789	0,834
7	0,666	0,750	0,798
8	0,632	0,716	0,765
9	0,602	0,685	0,735
10	0,576	0,658	0,708
11	0,553	0,634	0,684
12	0,532	0,612	0,661
13	0,514	0,592	0,641
14	0,497	0,574	0,623
15	0,482	0,558	0,606
16	0,468	0,542	0,590
17	0,456	0,528	0,575
18	0,444	0,516	0,561
19	0,433	0,503	0,549
20	0,423	0,492	0,537
25	0,381	0,445	0,487
30	0,349	0,409	0,449

Основная литература:

1. Кобринский Б.А., Зарубина Т.В. Медицинская информатика: Учебник. М: изд. "Академия", 2009.
2. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: Учебное пособие для практических занятий / Под ред. В.З.Кучеренко. – М.:ГЭОТАР-МЕД, 2004. – 192 с.

Дополнительная литература:

1. Зайцев В.М., Лифляндский В.Г. Маринкин В.И. Прикладная медицинская статистика. – Спб: ООО «Издательства ФОЛИАНТ», 2003. – 432 с.
2. Карась С.И. Информационные основы принятия решений в медицине: Учебное пособие. – Томск: Печатная мануфактура, 2003.- 145с.
3. Чернов В.И., Родионов О.В., Есауленко И.Э. и др. Медицинская информатика: Учебное пособие.- Воронеж, 2004. – 282с.: ил.
5. Гельман В.Я. Медицинская информатика: практикум. – СПб: Питер, 2001. -480с. – (Серия "Национальная медицинская библиотека").
6. Богданов А.К., Проценко В.Д. Практические применения современных методов анализа изображений в медицине: Учебное пособие. – М.: РУДН, 2008. – 119с.: ил.
7. Санников А.Г., Егоров Д.Б., Скудных А.С., Рухлова С.А. Практикум по медицинской информатике: автоматизированное рабочее место врача и системы поддержки принятия врачебного решения. – Тюмень: П.П.Ш., 2009. – 116с.