

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 07.09.2023 09:15:24
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
« 21 » 09 2023 г.



СИСТЕМНЫЙ АНАЛИЗ

**Методические указания по выполнению практических работ
для студентов направления подготовки
12.03.04 «Биотехнические системы и технологии»**

Курск 2023

УДК 303.732.4

Составитель: Н.М. Агарков

Рецензент

Доктор медицинских наук, профессор *Л.В. Шульга*

Системный анализ : методические указания по выполнению практических работ для студентов направления подготовки 12.03.04 «Биотехнические системы и технологии» / Юго-Зап. гос. ун-т; сост.: Н.М. Агарков. – Курск, 2023. – 68 с.

Содержатся сведения, необходимые для выполнения практических занятий по системному анализу.

Методические указания по структуре, содержанию и стилю изложения материала соответствуют методическим и научным требованиям, предъявляемым к учебным и методическим пособиям.

Предназначены для студентов направления подготовки 12.03.04 очной формы обучения.

Текст печатается в авторской редакции

Подписано в печать . Формат 60x84 1/16.
Усл.печ. л. __. Уч.-изд. л. __. Тираж 30 экз. Заказ ~~650~~ . Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Практическая работа № 1.

«Описание строения и функционирования биотехнических систем»

Цель работы: получить практические навыки в выделении компонент системы, описании свойств и структуры системы, ее взаимодействия со средой, функционирования системы во времени и управления системой.

Самостоятельная работа: изучение основных понятий теории систем и системного анализа, закономерностей строения и функционирования систем.

Порядок выполнения

1. Выбор задания.

Выберите систему, на примере которой будет выполняться вся работа.

Примеры систем: - технические устройства (автомобили, компьютеры, телевизоры, фотоаппараты, мобильные устройства и т.д.); - организации или предприятия (библиотеки, театры, заводы, магазины, гостиницы, рестораны и т.д.); - биологические системы (человек, животные, растения и т.д.); - информационные системы (системы поддержки принятия решений системы автоматизированного проектирования, бухгалтерские и финансовые системы и т.д.); - целевые системы (система выборов, система водоснабжения, система безопасности, транспортная система и т.д.).

2. Построение иерархии состава.

Выделите основные подсистемы исследуемой системы (не менее 5 подсистем).

В рамках каждой из них выделите более мелкие подсистемы и элементы. Представьте компоненты системы в виде иерархической структуры. Пример иерархии состава для радиоприемника представлен на рис. 1.



Рис. 1. Иерархия состава радиоприемника

3. Описание сущностных свойств системы.

Опишите сущностное свойство системы и его внешнее проявление (явление).

Определите, является ли данное свойство эмерджентным. Ответ обоснуйте.

Например, для радиоприемника сущностное свойство - способность воспроизводить звук, закодированный в виде радиоволн и посланный радиостанцией. Явление - звучание приемника, передающего радиопередачу. Данное свойство является

эмерджентным, т.к. ни один из компонентов радиоприемника по отдельности не обладает им: антенна способна только улавливать радиоволны, преобразователь - преобразовывать радиоволны в звуковые и т.п.

4. Описание структуры системы и ее взаимодействия с окружением.

Выделите объекты окружающей среды. Составьте схему взаимодействия компонентов системы, а также схему взаимодействия со средой (это может быть одна общая схема). Если система слишком большая и сложная, можете составить схему для некоторой подсистемы. Опишите внутренние и внешние связи.

Пример схемы взаимодействия компонент радиоприемника друг с другом и с окружающей средой представлен на рис. 2.

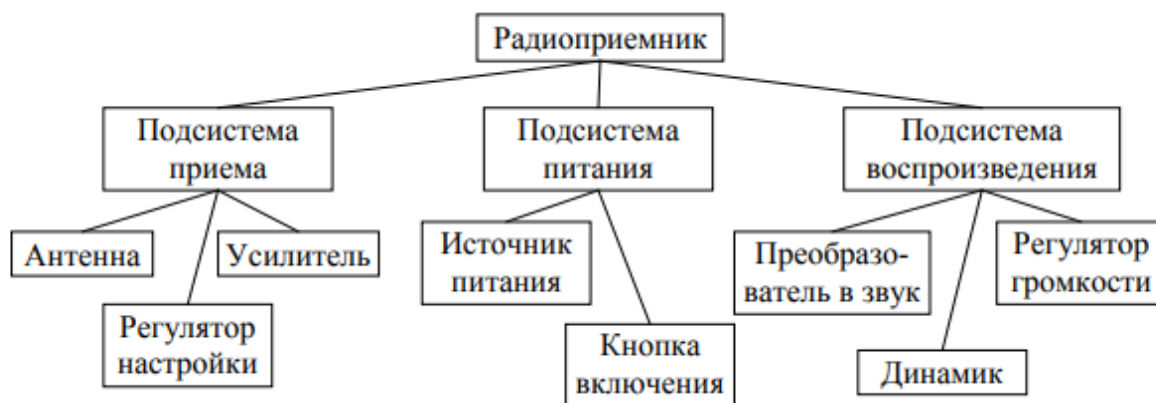


Рис. 2. Структура радиоприемника

5. Описание функционирования системы в пространстве состояний.

Выделите характеристики (параметры) системы. Параметры могут быть сгруппированы по типам: физические характеристики (размер, местоположение, цвет, материал), технические характеристики, экономические показатели и т.д.

Например, для радиоприемника могут быть выделены следующие параметры:

- физические (размер, цвет, материал, дизайн корпуса);
- технические (напряжение питания, диапазон радиоволн);
- параметры производителя (компания-производитель, страна, марка, дата производства, гарантийный срок) и т.д.

Из множества параметров выделите те, которые характеризуют поведение (функционирование) системы, т.е. которые изменяются во времени. Опишите различные состояния системы, указав конкретные значения параметров. Пример описания состояний радиоприемника приведен в таблице 1.

Таблица 1

Параметр	I состояние	II состояние	III состояние
Принимаемый диапазон радиоволн	101 FM	150 FM	150 FM
Уровень громкости	70 дб	70 дб	75 дб
Наличие помех	слабые	нет	нет

Укажите события, вызывающие переход из состояния в состояние. Например, для состояний, приведенных в таблице 1: настройка регулятора на другой диапазон радиоволн (переход из I в II), настройка уровня звука (переход из II в III).

6. Описание управления системой.

Определите основную цель системы. Если система является неживым объектом, цель, как правило, определяется пользователем. Например, для радиоприемника цель, задаваемая владельцем радиоприемника. — получить качественное (необходимой громкости и без помех) воспроизводство выбранной радиопередачи.

Определите, кто (что) и как управляет системой, с помощью каких управляющих воздействий осуществляется управление, используется ли в процессе управления обратная связь и если используется, то каким образом.

Например, управление радиоприемником осуществляет пользователь (внешнее управление) посредством изменения положений ручек и переключателей на корпусе радиоприемника. Управляющие воздействия: поворот ручки настройки диапазона и ручки тюнинга, поворот регулятора громкости. Обратная связь — определение на слух, наличия помех и уровня громкости звука, наблюдение за положением индикатора принимаемого диапазона радиоволн.

7. Составление отчета.

В отчет должны войти все схемы и описания, указанные в каждом из 6 пунктов.

Контрольные вопросы:

1. Что такое биотехническая система?
2. На какие типы можно разделить биотехнические системы?
3. Назовите роли технического компонента в процессе управления БТС.
4. Назовите свойства БТС.
5. Назовите основные этапы синтеза БТС.
6. Как происходит информационное согласование.
7. Что необходимо для установления связи между режимами функционирования воспринимающих систем и состоянием организма оператора?
8. Какие существуют основные принципы сопряжения технических и биологических элементов?
9. Какие выделяют медицинские БТС?

Практическая работа № 2.

Построение статических и динамических моделей.

Данная практическая работа выполняется методом творческих заданий. На примере, разбираемом на практическом занятии, студенты учатся анализировать экспериментальные данные, находить математические зависимости. В качестве творческих заданий предлагаются таблицы экспериментальных данных, полученных студентами заранее во время занятий научной работой. При решении творческих заданий, студенты могут выполнять роль экспертов, помогая другим студентам в группе найти правильное решение. Экспертами выбираются студенты, быстро построившие модель процесса.

Регрессионный анализ позволяет оценить степень связи между переменными, предлагая механизм вычисления предполагаемого значения переменной из нескольких уже известных значений. Используя регрессионный анализ, можно продлить линию тренда в диаграмме за пределы реальных данных для предсказания будущих значений.

Описание последовательности действий при моделировании:

Полученные в результате эксперимента данные зависимости между величинами x и y можно представить в виде таблицы 1:

Таблица 1-Экспериментальные данные

x	x_1	x_2	x_3	...	x_n
y	y_1	y_2	y_3	...	y_n

Необходимо найти эмпирическую формулу $y = f(x)$, связывающую между собой соответствующие значения переменных так, чтобы значения этой функции при $x = x_i$ возможно мало отличались бы от y_i , полученных из опыта.

Цели работы:

- 1) построить математическую модель в виде эмпирической формулы;
- 2) сделать оценку параметров модели;
- 3) проверить модель на адекватность.

Выбор общего вида эмпирической формулы может быть произведен на основе теоретических представлений о характере изучаемой зависимости. В других случаях приходится подбирать формулу, сравнивая кривую, построенную по данным

наблюдений с типичными графиками формул. Такими графиками могут служить линии тренда, которые можно добавить на диаграмму Microsoft Excel.

Линия тренда – это графическое представление направления изменения ряда данных. Линии тренда используются для анализа ошибок предсказания.

Точность аппроксимации. Линия тренда в наибольшей степени приближается к представленной на диаграмме зависимости, если значение R-квадрат равно или близко к 1. При аппроксимации данных с помощью линии тренда значение R-квадрат рассчитывается автоматически. Полученный результат можно вывести на диаграмме.

При этом можно использовать следующие функциональные зависимости:

Линейная: $Y = a + bx$, где a – координата пересечения оси абсцисс и b – угол наклона константы;

Логарифмическая: $Y = c \ln x + b$, где c и b – константы, \ln – функция натурального логарифма.

Экспоненциальная: $Y = ce^{bx}$, где c и b – константы, e – основание натурального логарифма.

Степенная: $Y = cx^b$, где c и b – константы;

Полиномиальная: $Y = b + c_1x + c_2x^2 + c_3x^3 + \dots + c_bx^b$, где b и $c_1 \dots c_b$ – константы.

Величина достоверности аппроксимации – R. Число от 0 до 1, которое отражает близость значений линии тренда к фактическим данным. Линия тренда наиболее соответствует действительности, когда значение R в квадрате близко к 1. Оно также называется квадратом смешанной корреляции.

Классический подход к оцениванию параметров линейной регрессии $Y = a + bx$ основан на методе наименьших квадратов, который позволяет получить такие оценки параметров a и b , при которых сумма квадратов отклонений фактических значений Y результативного признака от расчетных (теоретических) $f(x)$ будет минимальна:

$$OШ = \sum (f(x) - Y)^2 \Rightarrow \min,$$

$$OШ_{\text{лин}} = (f_{\text{лин}}(x_i) - y_i)^2; OШ_{\text{экс}} = (f_{\text{экс}}(x_i) - y_i)^2; OШ_{\text{лог}} = (f_{\text{лог}}(x_i) - y_i)^2.$$

т.е. из всего множества линий регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной.

Средняя квадратическая ошибка вычисляется по формуле:

$$\delta = \sqrt{\frac{\sum(F(x) - Y)^2}{n}}$$

Методика выполнения работы

1. Оформить исходные данные в виде сводной таблицы Microsoft Excel.
2. С помощью Мастера диаграмм М. Excel построить график зависимости всего диапазона данных сводной таблицы.
3. Построить линию тренда.
4. Для полученных математических моделей сделать оценку параметров:
 - а) провести вычисление средней квадратичной ошибки δ ;
 - б) сравнить δ с величиной достоверности аппроксимации – R.
5. Проверить модель на адекватность. Функция, которой соответствует минимальное значение δ и максимальное значение R, является математической моделью, наиболее близко описывающей исходные данные.

Пример 1: Проведено исследование зависимости функционально-технологических свойств (показателя активной кислотности рН и щелочности) водно-спиртовых смесей от вариации объемных долей спирта V1 и воды V2 (V1+ V2 = 100 мл). В результате эксперимента были получены следующие зависимости (таблица 2).

Таблица 2 - Зависимость щелочности и показателя активной кислотности рН от объемной доли спирта

Объемная доля спирта V1, мл	20	25	30	35	40	45	50	55	60	65	70	75	80
Объемная доля воды V2, мл	80	75	70	65	60	55	50	45	40	35	30	25	20
рН Y1	7,35	7,35	7,52	7,77	7,84	7,86	7,92	7,98	8,03	8,25	8,29	8,4	8,6
Щелочность Y2	3	2,7	2,6	2,4	2,1	1,9	1,8	1,6	1	1,3	1,2	1	0,9

Необходимо построить различные виды зависимостей рН и щелочности спирта от объемной доли спирта в водно-спиртовой смеси и выбрать уравнение линии тренда наиболее соответствующее действительности для предсказания будущих значений.

Решение задачи

Осуществим выбор прогнозной модели, позволяющей наиболее точно указать зависимость уровня рН водно-спиртовой смеси от объемной доли спирта. Для этого построим зависимость величины Y от V_1 (рис.1).

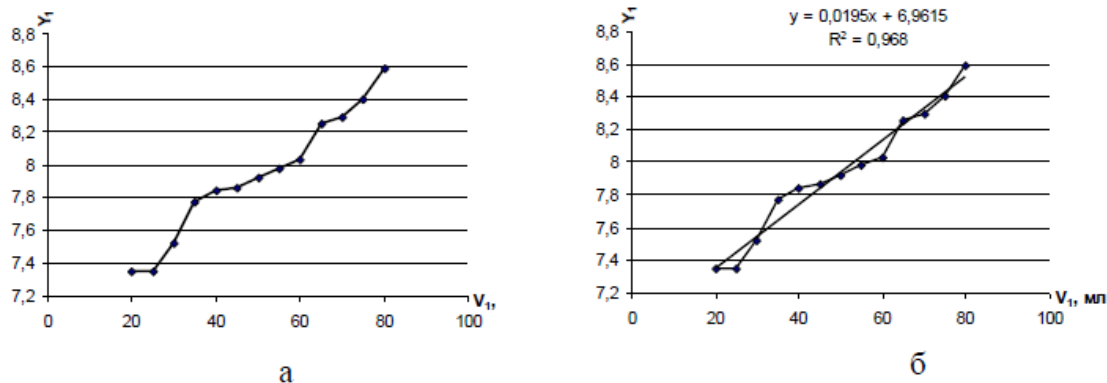


Рисунок 1 – График зависимости рН от объемной доли спирта (X – объемная доля спирта, Y – уровень рН): а – без линии тренда, б – с линией линейного тренда

Добавим к построенному графику линию тренда, которая позволяет однозначно определить характер наблюдаемой динамики (рис. 2, 3).

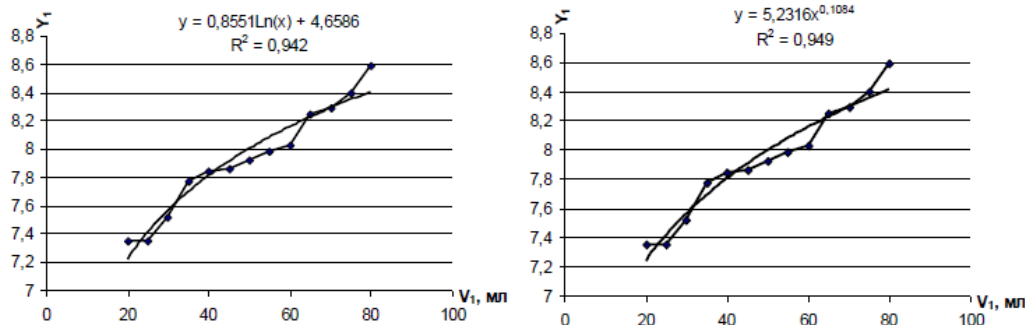


Рисунок 2 – Логарифмический тренд

Рисунок 3 – Степенной тренд

Итак, по значению коэффициента детерминации R^2 (квадрата корреляции) наиболее значимой оказывается линейная линия тренда (для данной модели R^2 принимает наибольшее значение). Получаем математические модели:

$$f_{\text{лин}}(x) = 0,0195x + 6,9615;$$

$$f_{\text{эксп}}(xi) = 5,2316x^{0,1084};$$

$$f_{\text{лог}}(xi) = 0,8551\ln(x) + 4,6586.$$

2. Для полученных моделей оценим параметры: а) проведем вычисление средней квадратичной ошибки δ ;

$$\text{ОШ}_{\text{лин}} = (f_{\text{лин}}(xi) - yi)^2; \text{ОШ}_{\text{эксп}} = (f_{\text{эксп}}(xi) - yi)^2; \text{ОШ}_{\text{лог}} = (f_{\text{лог}}(xi) - yi)^2.$$

3. Сравним значения δ полученных формул и величины достоверности аппроксимации – R. Приходим к выводу, что наилучшим образом исходные данные описывает линейная регрессионная модель.

Для того чтобы «улучшить» построенную эмпирическую зависимость построим полиномиальную зависимость 6-й степени. Зависимость

$$y = 2E-09x^6 - 7E-07x^5 + 8E-05x^4 - 0,0053x^3 + 0,1844x^2 - 3,1966x + 28,928$$

имеет большой коэффициент детерминации $R^2 = 0,9958$ (т.е. доля вариации величины рН уже на 99,58% будет объясняться вариацией объемной долей спирта в водно-спиртовой смеси).

Регрессионно-факторный анализ технологических моделей.

Данная практическая работа выполняется методом творческих заданий. Предлагается использовать таблицы экспериментальных данных, полученных студентами заранее во время занятий научной работой. При решении творческих заданий, студенты могут выполнять роль экспертов, помогая другим студентам в группе найти правильное решение. Экспертами выбираются студенты, быстро построившие модель процесса.

Регрессионный анализ позволяет оценить степень связи между переменными, предлагая механизм вычисления предполагаемого значения $y = 0,8551 \ln(x) + 4,6586$ переменной из нескольких уже известных значений.

В состав M. Excel входит набор средств анализа данных (пакет анализа), предназначенный для решения сложных статистических и инженерных задач. Для анализа данных следует указать входные данные и выбрать параметры; анализ будет выполнен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещен в выходной диапазон. Другие средства позволяют представить результаты анализа в графическом виде. Средства, включенные в пакет анализа данных, доступны через команду Анализ данных меню Сервис. Если этой команды нет в меню, необходимо загрузить надстройку Пакет анализа.

Пример: В лабораторных условиях было исследовано влияние ферментного препарата глюкозооксидазы (x_2) в сочетании с аскорбиновой кислотой (x_1) на качество хлеба (табл.). Необходимо с помощью ЭВМ рассчитать, какой фактор (x_1 или x_2) оказывает большее влияние на пористость хлеба; построить эмпирическую линейную модель зависимости пористости хлеба y от фактора x_1 или x_2 , оказывающего на него большее влияние; выявить, как изменится пористость хлеба, если величину глюкозооксидазы увеличить на 30% от среднего значения выборки.

Таблица 3- Влияние ферментного препарата глюкозооксидазы в сочетании с аскорбиновой кислотой на качество хлеба

Количество аскорбиновой кислоты x_1 , %	Количество глюкозооксидазы x_2 , %	Пористость y , %
0,003	0,0146	87
0,003	0,0853	85
0,017	0,0146	87
0,017	0,0854	85
0,000	0,0500	85

1. Создадим шаблон-таблицу в M. Excel.

2. Определим, какой фактор (количество аскорбиновой кислоты x_1 или количество глюкозооксидазы x_2) оказывает большее влияние на пористость хлеба с помощью коэффициента регрессии. Для этого построим матрицу коэффициентов корреляции: Сервис → Анализ данных → Корреляция (рис. 4).

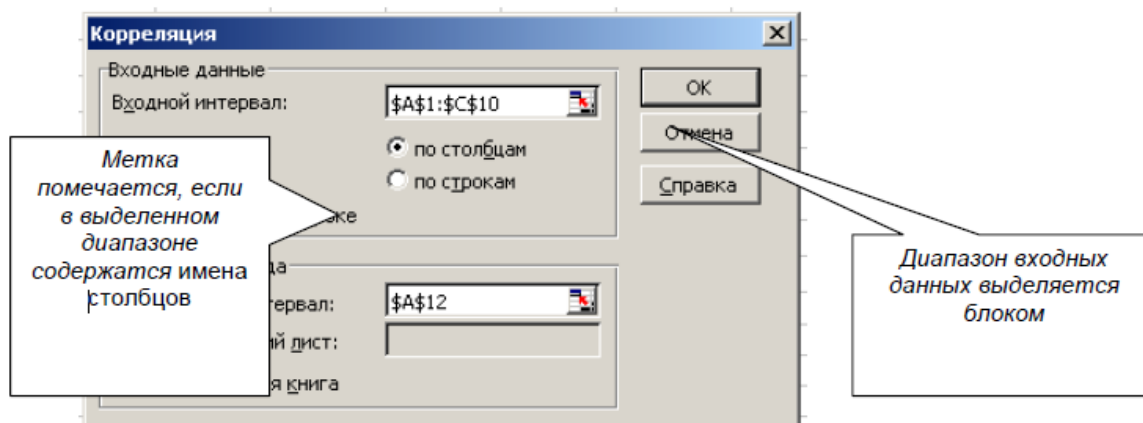


Рисунок 4 – Определение матрицы коэффициентов корреляции

Корреляционный анализ дает возможность установить, ассоциированы ли наборы данных по величине, то есть, большие значения из одного набора данных связаны с большими значениями другого набора (положительная корреляция), или, наоборот, малые значения одного набора связаны с большими значениями другого (отрицательная корреляция), или данные двух диапазонов никак не связаны (нулевая корреляция). Коэффициент корреляции, как ковариационный анализ, характеризует область, в которой два измерения изменяются вместе. Коэффициент масштабируется таким образом, что его значение не зависит от единиц, в которых выражены переменные двух измерений.

Таблица 4- Матрица коэффициентов корреляции

	Количество аскорбиновой	Количество глюкозооксидазы	Пористость y , %

	кислоты x1, %	x2, %	
Количество аскорбиновой кислоты x1, %	1		
Количество глюкозооксидазы x2, %	0,000351665	1	
Пористость y, %	0,304572452	-0,88317	1

Любое значение коэффициента корреляции должно находиться в диапазоне от -1 до $+1$ включительно. Чем ближе по модулю коэффициент корреляции r_{xy} , тем теснее связь между x и y . Если $|r_{x1y}| > |r_{x2y}|$, то фактор $x1$ оказывает большее влияние на y , чем $x2$. В результате имеем: $r_{x1y} = 0,304572452$ – связь (между $x1$ и y) слабая;

$r_{x2y} = -0,88317$ – связь (между $x2$ и y) умеренная обратная (т.е. с увеличением количества глюкооксидазы, пористость хлеба уменьшается).

Значит $|r_{x1y}| < |r_{x2y}|$, поэтому фактор $x1$ оказывает меньшее влияние на y , чем $x2$. Следовательно, будем строить эмпирическую зависимость *Метка помечается, если в выделенном диапазоне содержатся имена столбцов Диапазон входных данных выделяется блоком* пористости хлеба y от количества глюкооксидазы $x2$.

Построим эмпирическую модель зависимости пористости хлеба y от количества глюкооксидазы $x2$. Для этого используем встроенный пакет регрессионного анализа: Сервис → Анализ данных → Регрессия (рис. 5).

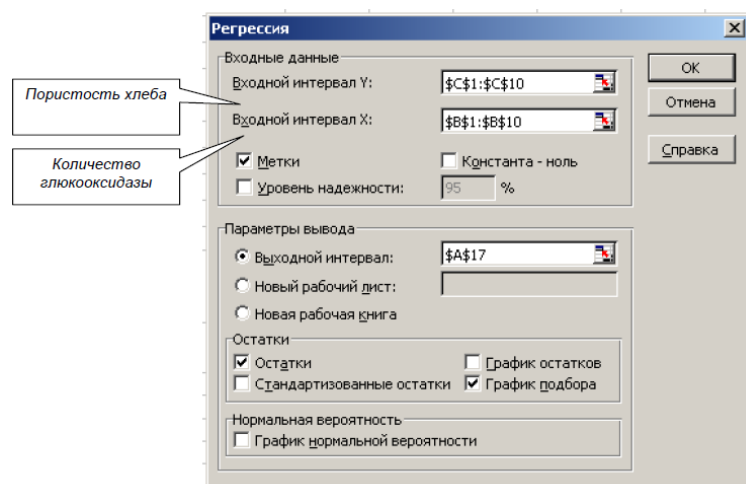


Рисунок 5 – Вид окна пакета «Регрессия»

Линейный регрессионный анализ заключается в подборе графика для набора наблюдений с помощью метода наименьших квадратов.

Получили зависимость: $y=87,3-29,1x^2$. Коэффициент детерминации $R^2=0,78$, т.е. доля вариации y объясняется лишь на 78% вариацией (есть смысл «улучшить» модель для вычисления значения).

Метод наименьших квадратов (МНК).

Пример.

Экспериментальные данные о значениях переменных x и y приведены в таблице.

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
x_i	0	1	2	4	5
y_i	2,1	2,4	2,6	2,8	3,0

В результате их выравнивания получена функция $g(x) = \sqrt[3]{x+1} + 1$

Используя метод наименьших квадратов, аппроксимировать эти данные линейной зависимостью $y=ax+b$ (найти параметры a и b). Выяснить, какая из двух линий лучше (в смысле метода наименьших квадратов) выравнивает экспериментальные данные. Сделать чертеж.

Суть метода наименьших квадратов (МНК).

Задача заключается в нахождении коэффициентов линейной зависимости, при

$$F(a,b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

которых функция двух переменных a и b принимает наименьшее значение. То есть, при данных a и b сумма квадратов отклонений экспериментальных данных от найденной прямой будет наименьшей. В этом вся суть метода наименьших квадратов.

Таким образом, решение примера сводится к нахождению экстремума функции двух переменных.

Вывод формул для нахождения коэффициентов.

Составляется и решается система из двух уравнений с двумя неизвестными.

$$F(a,b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

Находим частные производные функции по переменным a и b , приравняем эти производные к нулю.

$$\begin{cases} \frac{\partial F(a,b)}{\partial a} = 0 \\ \frac{\partial F(a,b)}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} -2 \sum_{i=1}^n (y_i - (ax_i + b))x_i = 0 \\ -2 \sum_{i=1}^n (y_i - (ax_i + b)) = 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases} \Leftrightarrow \begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases}$$

Решаем полученную систему уравнений любым методом (например *методом подстановки* или [методом Крамера](#)) и получаем формулы для нахождения

коэффициентов по методу наименьших квадратов (МНК).

$$\begin{cases} a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \\ b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \end{cases}$$

$$F(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

При данных a и b функция принимает наименьшее значение. Доказательство этого факта приведено [ниже по тексту в конце страницы](#).

Вот и весь метод наименьших квадратов. Формула для нахождения параметра a содержит суммы $\sum_{i=1}^n x_i$, $\sum_{i=1}^n y_i$, $\sum_{i=1}^n x_i y_i$, $\sum_{i=1}^n x_i^2$ и параметр n - количество экспериментальных данных. Значения этих сумм рекомендуем вычислять отдельно. Коэффициент b находится после вычисления a .

Пришло время вспомнить про исходный пример.

Решение.

В нашем примере $n=5$. Заполняем таблицу для удобства вычисления сумм, которые входят в формулы искомым коэффициентов.

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$\sum_{i=1}^5$
x_i	0	1	2	4	5	12
y_i	2,1	2,4	2,6	2,8	3	12,9
$x_i y_i$	0	2,4	5,2	11,2	15	33,8
x_i^2	0	1	4	16	25	46

Значения в четвертой строке таблицы получены умножением значений 2-ой строки на значения 3-ей строки для каждого номера i .

Значения в пятой строке таблицы получены возведением в квадрат значений 2-ой строки для каждого номера i .

Значения последнего столбца таблицы – это суммы значений по строкам.

Используем формулы метода наименьших квадратов для нахождения коэффициентов a и b . Подставляем в них соответствующие значения из последнего столбца таблицы:

$$\begin{cases} a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \\ b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \end{cases} \Rightarrow \begin{cases} a = \frac{5 \cdot 33,8 - 12 \cdot 12,9}{5 \cdot 46 - 12^2} \Rightarrow \begin{cases} a \approx 0,165 \\ b \approx 2,184 \end{cases} \\ b = \frac{12,9 - a \cdot 12}{5} \end{cases}$$

Следовательно, $y = 0.165x + 2.184$ - искомая аппроксимирующая прямая.

Осталось выяснить какая из линий $y = 0.165x + 2.184$ или $g(x) = \sqrt[3]{x+1} + 1$ лучше аппроксимирует исходные данные, то есть произвести оценку методом наименьших квадратов.

Оценка погрешности метода наименьших квадратов.

Для этого требуется вычислить суммы квадратов отклонений исходных данных от

этих линий $\sigma_1 = \sum_{i=1}^n (y_i - (ax_i + b))^2$ и $\sigma_2 = \sum_{i=1}^n (y_i - g(x_i))^2$, меньшее значение соответствует линии, которая лучше в смысле метода наименьших квадратов аппроксимирует исходные данные.

$$\sigma_1 = \sum_{i=1}^n (y_i - (ax_i + b))^2 =$$

$$\sum_{i=1}^5 (y_i - (0.165x_i + 2.184))^2 \approx 0.019$$

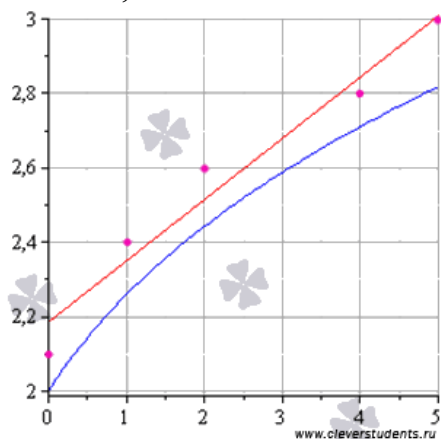
$$\sigma_2 = \sum_{i=1}^n (y_i - g(x_i))^2 =$$

$$= \sum_{i=1}^5 (y_i - (\sqrt[3]{x_i+1} + 1))^2 \approx 0.096$$

Так как $\sigma_1 < \sigma_2$, то прямая $y = 0.165x + 2.184$ лучше приближает исходные данные.

Графическая иллюстрация метода наименьших квадратов (мнк).

На графиках все прекрасно видно. Красная линия – это найденная прямая $y = 0.165x + 2.184$, синяя линия – это $g(x) = \sqrt[3]{x+1} + 1$, розовые точки – это исходные данные.



На практике при моделировании различных процессов - в частности, экономических, физических, технических, социальных - широко используются те или иные способы вычисления приближенных значений функций по известным их значениям в некоторых фиксированных точках.

Такого рода задачи приближения функций часто возникают:

- при построении приближенных формул для вычисления значений характерных величин исследуемого процесса по табличным данным, полученным в результате эксперимента;
- при численном интегрировании, дифференцировании, решении дифференциальных уравнений и т. д.;

- при необходимости вычисления значений функций в промежуточных точках рассматриваемого интервала;
- при определении значений характерных величин процесса за пределами рассматриваемого интервала, в частности при прогнозировании.

Если для моделирования некоторого процесса, заданного таблицей, построить функцию, приближенно описывающую данный процесс на основе метода наименьших квадратов, она будет называться аппроксимирующей функцией (регрессией), а сама задача построения аппроксимирующих функций - задачей аппроксимации.

В Excel для построения регрессий имеются две возможности.

1. Добавление выбранных регрессий (линий тренда - trendlines) в диаграмму, построенную на основе таблицы данных для исследуемой характеристики процесса (доступно лишь при наличии построенной диаграммы);

2. Использование встроенных статистических функций рабочего листа Excel, позволяющих получать регрессии (линии тренда) непосредственно на основе таблицы исходных данных.

Добавление линий тренда в диаграмму

Для таблицы данных, описывающих некоторый процесс и представленных диаграммой, в Excel имеется эффективный инструмент регрессионного анализа, позволяющий:

- строить на основе метода наименьших квадратов и добавлять в диаграмму пять типов регрессий, которые с той или иной степенью точности моделируют исследуемый процесс;
- добавлять к диаграмме уравнение построенной регрессии;
- определять степень соответствия выбранной регрессии отображаемым на диаграмме данным.

На основе данных диаграммы Excel позволяет получать линейный, полиномиальный, логарифмический, степенной, экспоненциальный типы регрессий, которые задаются уравнением:

$$y = y(x)$$

где x - независимая переменная, которая часто принимает значения последовательности натурального ряда чисел (1; 2; 3; ...) и производит, например, отсчет времени протекания исследуемого процесса (характеристики).

1. Линейная регрессия хороша при моделировании характеристик, значения которых увеличиваются или убывают с постоянной скоростью. Это наиболее простая в построении модель исследуемого процесса. Она строится в соответствии с уравнением:

$$y = mx + b$$

где m - тангенс угла наклона линейной регрессии к оси абсцисс; b - координата точки пересечения линейной регрессии с осью ординат.

2. Полиномиальная линия тренда полезна для описания характеристик, имеющих несколько ярко выраженных экстремумов (максимумов и минимумов). Выбор степени полинома определяется количеством экстремумов исследуемой характеристики. Так, полином второй степени может хорошо описать процесс, имеющий только один максимум или минимум; полином третьей степени - не более двух экстремумов; полином четвертой степени - не более трех экстремумов и т. д.

В этом случае линия тренда строится в соответствии с уравнением:

$$y = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5x^5 + c_6x^6$$

где коэффициенты $c_0, c_1, c_2, \dots, c_6$ - константы, значения которых определяются в ходе построения.

3. Логарифмическая линия тренда с успехом применяется при моделировании характеристик, значения которых вначале быстро меняются, а затем постепенно стабилизируются.

Строится в соответствии с уравнением:

$$y = c \ln(x) + b$$

где коэффициенты b, c - константы.

4. Степенная линия тренда дает хорошие результаты, если значения исследуемой зависимости характеризуются постоянным изменением скорости роста. Примером такой зависимости может служить график равноускоренного движения автомобиля. Если среди данных встречаются нулевые или отрицательные значения, использовать степенную линию тренда нельзя.

Строится в соответствии с уравнением:

$$y = c x^b$$

где коэффициенты b, c - константы.

5. Экспоненциальную линию тренда следует использовать в том случае, если скорость изменения данных непрерывно возрастает. Для данных, содержащих нулевые или отрицательные значения, этот вид приближения также неприменим.

Строится в соответствии с уравнением:

$$y = c e^{bx}$$

где коэффициенты b, c - константы.

При подборе линии тренда Excel автоматически рассчитывает значение величины R^2 , которая характеризует достоверность аппроксимации: чем ближе значение R^2 к единице, тем надежнее линия тренда аппроксимирует исследуемый процесс. При необходимости значение R^2 всегда можно отобразить на диаграмме.

Определяется по формуле:

$$R^2 = 1 - \frac{\Sigma_1}{\Sigma_2}$$

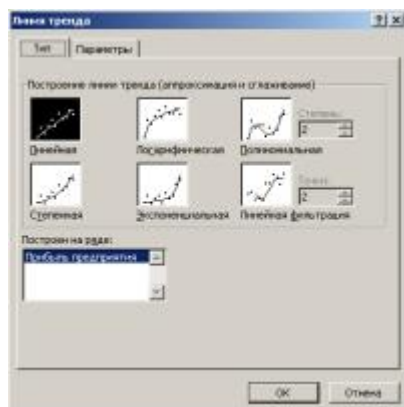
где $\Sigma_1 = \sum_j (y_j - Y_j)^2$;

$$\Sigma_2 = \sum_j Y_j^2 - \frac{1}{n} \cdot (\sum_j Y_j)^2$$

Для добавления линии тренда к ряду данных следует:

- активизировать построенную на основе ряда данных диаграмму, т. е. щелкнуть в пределах области диаграммы. В главном меню появится пункт Диаграмма;
- после щелчка на этом пункте на экране появится меню, в котором следует выбрать команду Добавить линию тренда.

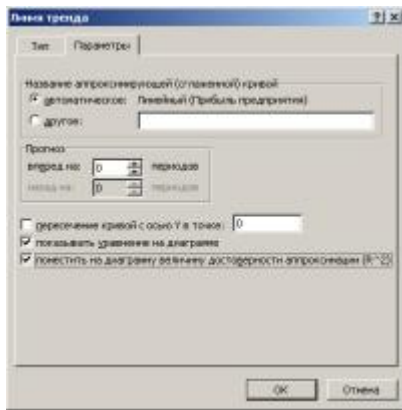
Эти же действия легко реализуются, если навести указатель мыши на график, соответствующий одному из рядов данных, и щелкнуть правой кнопкой мыши; в появившемся контекстном меню выбрать команду Добавить линию тренда. На экране появится диалоговое окно Линия тренда с раскрытой вкладкой Тип (рис. 1).



После этого необходимо:

Выбрать на вкладке Тип необходимый тип линии тренда (по умолчанию выбирается тип Линейный). Для типа Полиномиальная в поле Степень следует задать степень выбранного полинома.

1. В поле Построен на ряде перечислены все ряды данных рассматриваемой диаграммы. Для добавления линии тренда к конкретному ряду данных следует в поле Построен на ряде выбрать его имя.



При необходимости, перейдя на вкладку Параметры (рис. 2), можно для линии тренда задать следующие параметры:

- изменить название линии тренда в поле Название аппроксимирующей (сглаженной) кривой.
- задать количество периодов (вперед или назад) для прогноза в поле Прогноз;
- вывести в область диаграммы уравнение линии тренда, для чего следует включить флажок показать уравнение на диаграмме;
- вывести в область диаграммы значение достоверности аппроксимации R^2 , для чего следует включить флажок поместить на диаграмму величину достоверности аппроксимации (R^2);
- задать точку пересечения линии тренда с осью Y, для чего следует включить флажок пересечение кривой с осью Y в точке;
- щелкнуть на кнопке ОК, чтобы закрыть диалоговое окно.

Для того, чтобы начать редактирование уже построенной линии тренда, существует три способа:

1. воспользоваться командой Выделенная линия тренда из меню Формат, предварительно выбрав линию тренда;
2. выбрать команду Формат линии тренда из контекстного меню, которое вызывается щелчком правой кнопки мыши по линии тренда;
3. двойным щелчком по линии тренда.

На экране появится диалоговое окно Формат линии тренда (рис. 3), содержащее три вкладки: Вид, Тип, Параметры, причем содержимое последних двух полностью совпадает с аналогичными вкладками диалогового окна Линия тренда (рис.1-2). На вкладке Вид, можно задать тип линии, ее цвет и толщину.



Для удаления уже построенной линии тренда следует выбрать удаляемую линию тренда и нажать клавишу Delete.

Достоинствами рассмотренного инструмента регрессионного анализа являются:

- относительная легкость построения на диаграммах линии тренда без создания для нее таблицы данных;
- достаточно широкий перечень типов предложенных линий трендов, причем в этот перечень входят наиболее часто используемые типы регрессии;
- возможность прогнозирования поведения исследуемого процесса на произвольное (в пределах здравого смысла) количество шагов вперед, а также назад;
- возможность получения уравнения линии тренда в аналитическом виде;
- возможность, при необходимости, получения оценки достоверности проведенной аппроксимации.

К недостаткам можно отнести следующие моменты:

- построение линии тренда осуществляется лишь при наличии диаграммы, построенной на ряде данных;
- процесс формирования рядов данных для исследуемой характеристики на основе полученных для нее уравнений линий тренда несколько загроможден: искомые уравнения регрессий обновляются при каждом изменении значений исходного ряда данных, но только в пределах области диаграммы, в то время как ряд данных, сформированный на основе старого уравнения линии тренда, остается без изменения;
- в отчетах сводных диаграмм при изменении представления диаграммы или связанного отчета сводной таблицы имеющиеся линии тренда не сохраняются, то есть до проведения линий тренда или другого форматирования отчета сводных диаграмм следует убедиться, что макет отчета удовлетворяет необходимым требованиям.

Линиями тренда можно дополнить ряды данных, представленные на диаграммах типа график, гистограмма, плоские ненормированные диаграммы с областями, линейчатые, точечные, пузырьковые и биржевые.

Нельзя дополнить линиями тренда ряды данных на объемных, нормированных, лепестковых, круговых и кольцевых диаграммах.

Использование встроенных функций Excel

В Excel имеется также инструмент регрессионного анализа для построения линий тренда вне области диаграммы. Для этой цели можно использовать ряд статистических функций рабочего листа, однако все они позволяют строить лишь линейные или экспоненциальные регрессии.

В Excel имеется несколько функций для построения линейной регрессии, в частности:

- ТЕНДЕНЦИЯ;
- ЛИНЕЙН;
- НАКЛОН и ОТРЕЗОК.

А также несколько функций для построения экспоненциальной линии тренда, в частности:

- РОСТ;
- ЛГРФПРИБЛ.

Следует отметить, что приемы построения регрессий с помощью функций ТЕНДЕНЦИЯ и РОСТ практически совпадают. То же самое можно сказать и о паре функций ЛИНЕЙН и ЛГРФПРИБЛ. Для четырех этих функций при создании таблицы значений используются такие возможности Excel, как формулы массивов, что несколько загромождает процесс построения регрессий. Заметим также, что построение линейной регрессии, на наш взгляд, легче всего осуществить с помощью функций НАКЛОН и ОТРЕЗОК, где первая из них определяет угловой коэффициент линейной регрессии, а вторая - отрезок, отсекаемый регрессией на оси ординат.

Достоинствами инструмента встроенных функций для регрессионного анализа являются:

- достаточно простой однотипный процесс формирования рядов данных исследуемой характеристики для всех встроенных статистических функций, задающих линии тренда;
- стандартная методика построения линий тренда на основе сформированных рядов данных;
- возможность прогнозирования поведения исследуемого процесса на необходимое количество шагов вперед или назад.

А к недостаткам относится то, что в Excel нет встроенных функций для создания других (кроме линейного и экспоненциального) типов линий тренда. Это обстоятельство часто не позволяет подобрать достаточно точную модель исследуемого процесса, а также получить близкие к реальности прогнозы. Кроме того, при использовании функций ТЕНДЕНЦИЯ и РОСТ не известны уравнения линий тренда.

Следует отметить, что авторы не ставили целью статьи изложение курса регрессионного анализа с той или иной степенью полноты. Основная ее задача - на конкретных примерах показать возможности пакета Excel при решении задач аппроксимации; продемонстрировать, какими эффективными инструментами для построения регрессий и прогнозирования обладает Excel; проиллюстрировать, как относительно легко такие задачи могут быть решены даже пользователем, не владеющим глубокими знаниями регрессионного анализа.

Примеры решения конкретных задач

Рассмотрим решение конкретных задач с помощью перечисленных инструментов пакета Excel.

Задача 1

С таблицей данных о прибыли автотранспортного предприятия за 1995-2002 гг. необходимо выполнить следующие действия.

1. Построить диаграмму.
2. В диаграмму добавить линейную и полиномиальную (квадратичную и кубическую) линии тренда.
3. Вывести уравнения полученных линий тренда, а также величины достоверности аппроксимации R^2 для каждой из них.

4. Используя уравнения линий тренда, получить табличные данные по прибыли предприятия для каждой линии тренда за 1995-2004 г.г.

5. Составить прогноз по прибыли предприятия на 2003 и 2004 гг.

Решение задачи

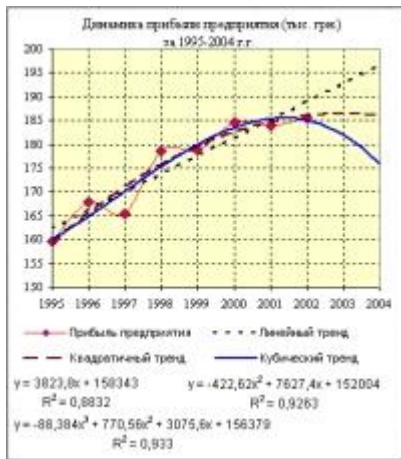
	A	B	C	D	E	F
1	Диагностика прибыли предприятия (грн.) за 1995-2004 г.г.					
2						
3	№ п/п	Годы	Прибыль, грн.			
4	1	1995	159 500,00			
5	2	1996	167 800,00			
6	3	1997	165 400,00			
7	4	1998	178 600,00			
8	5	1999	178 900,00			
9	6	2000	184 700,00			
10	7	2001	184 000,00			
11	8	2002	185 600,00			
12	9	2003	?			
13	10	2004	?			

1. В диапазон ячеек A4:C11 рабочего листа Excel вводим рабочую таблицу, представленную на рис. 4.

2. Выделив диапазон ячеек B4:C11, строим диаграмму.

3. Активизируем построенную диаграмму и по описанной выше методике после выбора типа линии тренда в диалоговом окне Линия тренда (см. рис. 1) поочередно добавляем в диаграмму линейную, квадратичную и кубическую линии тренда. В этом же диалоговом окне открываем вкладку Параметры (см. рис. 2), в поле Название аппроксимирующей (сглаженной) кривой вводим наименование добавляемого тренда, а в поле Прогноз вперед на: периодов задаем значение 2, так как планируется сделать прогноз по прибыли на два года вперед. Для вывода в области диаграммы уравнения регрессии и значения достоверности аппроксимации R² включаем флажки показывать уравнение на экране и поместить на диаграмму величину достоверности аппроксимации (R²). Для лучшего визуального восприятия изменяем тип, цвет и толщину построенных линий тренда, для чего воспользуемся вкладкой Вид диалогового окна Формат линии тренда (см. рис. 3). Полученная диаграмма с добавленными линиями тренда представлена на рис. 5.

4. Для получения табличных данных по прибыли предприятия для каждой линии тренда за 1995-2004 гг. воспользуемся уравнениями линий тренда, представленными на рис. 5. Для этого в ячейки диапазона D3:F3 вводим текстовую информацию о типе выбранной линии тренда: Линейный тренд, Квадратичный тренд, Кубический тренд. Далее вводим в ячейку D4 формулу линейной регрессии и, используя маркер заполнения, копируем эту формулу с относительными ссылками в диапазон ячеек D5:D13. Следует отметить, что каждой ячейке с формулой линейной регрессии из диапазона ячеек D4:D13 в качестве аргумента стоит соответствующая ячейка из диапазона A4:A13. Аналогично для квадратичной регрессии заполняется диапазон ячеек E4:E13, а для кубической регрессии - диапазон ячеек F4:F13. Таким образом, составлен прогноз по прибыли предприятия на 2003 и 2004 гг. с помощью трех трендов. Полученная таблица значений представлена на рис. 6.



№ п/п	Годы	Прибыль, грн.	Линейный тренд	Квадратичный тренд	Кубический тренд	
4	1	1995	159 500,00	162 166,80	159 209,78	160 136,78
5	2	1996	167 800,00	165 990,60	165 567,52	164 906,37
6	3	1997	165 400,00	169 814,40	171 081,82	170 154,47
7	4	1998	178 600,00	173 638,20	175 750,08	175 353,78
8	5	1999	178 900,00	177 462,00	179 573,90	179 973,00
9	6	2000	184 700,00	181 285,80	182 551,88	183 481,82
10	7	2001	184 000,00	185 109,60	184 684,62	185 349,93
11	8	2002	185 500,00	188 933,40	185 972,32	185 847,03
12	9	2003	?	192 757,20	186 414,76	182 042,82
13	10	2004	?	196 581,00	186 012,00	175 807,00

Задача 2

С таблицей данных о прибыли автотранспортного предприятия за 1995-2002 гг., приведенной в задаче 1, необходимо выполнить следующие действия.

1. Построить диаграмму.
2. В диаграмму добавить логарифмическую, степенную и экспоненциальную линии тренда.
3. Вывести уравнения полученных линий тренда, а также величины достоверности аппроксимации R^2 для каждой из них.
4. Используя уравнения линий тренда, получить табличные данные о прибыли предприятия для каждой линии тренда за 1995-2002 гг.
5. Составить прогноз о прибыли предприятия на 2003 и 2004 гг., используя эти линии тренда.

Решение задачи

Следуя методике, приведенной при решении задачи 1, получаем диаграмму с добавленными в нее логарифмической, степенной и экспоненциальной линиями тренда (рис. 7). Далее, используя полученные уравнения линий тренда, заполняем таблицу значений по прибыли предприятия, включая прогнозируемые значения на 2003 и 2004 гг. (рис. 8).



№ п/п	Годы	Прибыль, грн.	Логарифмический тренд	Степенный тренд	Экспоненциальный тренд	
4	1	1995	154 836,00	154 836,00	155 228,00	162 203,25
5	2	1996	167 800,00	165 418,66	165 139,91	165 893,04
6	3	1997	165 400,00	171 609,71	171 228,87	169 583,13
7	4	1998	178 600,00	176 002,33	175 684,73	173 356,30
8	5	1999	178 900,00	179 409,51	179 220,66	177 211,38
9	6	2000	184 700,00	182 193,38	182 162,50	181 153,23
10	7	2001	184 000,00	184 547,10	184 687,42	185 182,76
11	8	2002	185 500,00	186 585,99	186 902,88	189 301,93
12	9	2003	?	188 384,42	188 879,11	193 512,72
13	10	2004	?	189 993,17	190 664,60	197 817,17

На рис. 5 и рис. видно, что модели с логарифмическим трендом, соответствует наименьшее значение достоверности аппроксимации

$$R2 = 0,8659$$

Наибольшие же значения R2 соответствуют моделям с полиномиальным трендом: квадратичным ($R2 = 0,9263$) и кубическим ($R2 = 0,933$).

Задача 3

С таблицей данных о прибыли автотранспортного предприятия за 1995-2002 гг., приведенной в задаче 1, необходимо выполнить следующие действия.

1. Получить ряды данных для линейной и экспоненциальной линии тренда с использованием функций ТЕНДЕНЦИЯ и РОСТ.
 2. Используя функции ТЕНДЕНЦИЯ и РОСТ, составить прогноз о прибыли предприятия на 2003 и 2004 гг.
 3. Для исходных данных и полученных рядов данных построить диаграмму.
- Решение задачи

Воспользуемся рабочей таблицей задачи 1 (см. рис. 4). Начнем с функции ТЕНДЕНЦИЯ:

1. выделяем диапазон ячеек D4:D11, который следует заполнить значениями функции ТЕНДЕНЦИЯ, соответствующими известным данным о прибыли предприятия;

2. вызываем команду Функция из меню Вставка. В появившемся диалоговом окне Мастер функций выделяем функцию ТЕНДЕНЦИЯ из категории Статистические, после чего щелкаем по кнопке ОК. Эту же операцию можно осуществить нажатием кнопки (Вставка функции) стандартной панели инструментов.

3. В появившемся диалоговом окне Аргументы функции вводим в поле Известные_значения_у диапазон ячеек C4:C11; в поле Известные_значения_x - диапазон ячеек B4:B11;

4. чтобы вводимая формула стала формулой массива, используем комбинацию клавиш ++.

Введенная нами формула в строке формул будет иметь вид: =ТЕНДЕНЦИЯ(C4:C11;B4:B11)}.

В результате диапазон ячеек D4:D11 заполняется соответствующими значениями функции ТЕНДЕНЦИЯ (рис. 9).

	A	B	C	D	E
1	Динамика прибыли предприятия (грн.)				
2	за 1995-2004 гг.				
3	№	Годы	Прибыль, грн.	Функция ТЕНДЕНЦИЯ	Функция РОСТ
4	1	1995	159 500,00	162 166,67	162 286,26
5	2	1996	167 800,00	165 990,48	165 903,44
6	3	1997	165 400,00	169 814,29	169 599,15
7	4	1998	178 600,00	173 638,10	173 377,18
8	5	1999	178 900,00	177 461,90	177 239,37
9	6	2000	184 700,00	181 285,71	181 187,60
10	7	2001	184 000,00	185 109,52	185 223,79
11	8	2002	185 500,00	188 933,33	189 349,88
12	9	2003	?	192 757,14	193 567,88
13	10	2004	?	196 580,95	197 879,85

Для составления прогноза о прибыли предприятия на 2003 и 2004 гг. необходимо:

1. выделить диапазон ячеек D12:D13, куда будут заноситься значения, прогнозируемые функцией ТЕНДЕНЦИЯ.

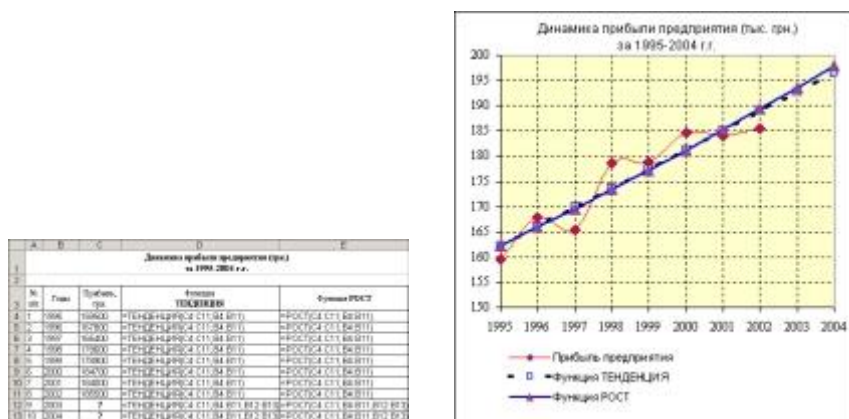
2. вызвать функцию ТЕНДЕНЦИЯ и в появившемся диалоговом окне Аргументы функции ввести в поле Известные_значения_y - диапазон ячеек C4:C11; в поле Известные_значения_x - диапазон ячеек B4:B11; а в поле Новые_значения_x - диапазон ячеек B12:B13.

3. превратить эту формулу в формулу массива, используя комбинацию клавиш Ctrl + Shift + Enter.

4. Введенная формула будет иметь вид: $=\{ТЕНДЕНЦИЯ(C4:C11;B4:B11;B12:B13)\}$, а диапазон ячеек D12:D13 заполнится прогнозируемыми значениями функции ТЕНДЕНЦИЯ (см. рис. 9).

Аналогично заполняется ряд данных с помощью функции РОСТ, которая используется при анализе нелинейных зависимостей и работает точно так же, как ее линейный аналог ТЕНДЕНЦИЯ.

На рис.10 представлена таблица в режиме показа формул.



Для исходных данных и полученных рядов данных построена диаграмма, изображенная на рис. 11.

Задача 4

С таблицей данных о поступлении в диспетчерскую службу автотранспортного предприятия заявок на услуги за период с 1 по 11 число текущего месяца необходимо выполнить следующие действия.

1. Получить ряды данных для линейной регрессии: используя функции НАКЛОН и ОТРЕЗОК; используя функцию ЛИНЕЙН.

2. Получить ряд данных для экспоненциальной регрессии с использованием функции ЛГРФПРИБЛ.

3. Используя вышеназванные функции, составить прогноз о поступлении заявок в диспетчерскую службу на период с 12 по 14 число текущего месяца.

4. Для исходных и полученных рядов данных построить диаграмму.

Решение задачи

Отметим, что, в отличие от функций ТЕНДЕНЦИЯ и РОСТ, ни одна из перечисленных выше функций (НАКЛОН, ОТРЕЗОК, ЛИНЕЙН, ЛГРФПРИБ) не

является регрессией. Эти функции играют лишь вспомогательную роль, определяя необходимые параметры регрессии.

Для линейной и экспоненциальной регрессий, построенных с помощью функций НАКЛОН, ОТРЕЗОК, ЛИНЕЙН, ЛГРФПРИБ, внешний вид их уравнений всегда известен, в отличие от линейной и экспоненциальной регрессий, соответствующих функциям ТЕНДЕНЦИЯ и РОСТ.

1. Построим линейную регрессию, имеющую уравнение:

$$y = mx+b$$

с помощью функций НАКЛОН и ОТРЕЗОК, причем угловой коэффициент регрессии m определяется функцией НАКЛОН, а свободный член b - функцией ОТРЕЗОК.

Для этого осуществляем следующие действия:

1. заносим исходную таблицу в диапазон ячеек A4:B14;
2. значение параметра m будет определяться в ячейке C19. Выбираем из категории Статистические функцию Наклон; заносим диапазон ячеек B4:B14 в поле известные_значения_y и диапазон ячеек A4:A14 в поле известные_значения_x. В ячейку C19 будет введена формула: =НАКЛОН(B4:B14;A4:A14);
3. по аналогичной методике определяется значение параметра b в ячейке D19. И ее содержимое будет иметь вид: =ОТРЕЗОК(B4:B14;A4:A14). Таким образом, необходимые для построения линейной регрессии значения параметров m и b будут сохраняться соответственно в ячейках C19, D19;
4. далее заносим в ячейку C4 формулу линейной регрессии в виде: =\$C*A4+\$D. В этой формуле ячейки C19 и D19 записаны с абсолютными ссылками (адрес ячейки не должен меняться при возможном копировании). Знак абсолютной ссылки \$ можно набить либо с клавиатуры, либо с помощью клавиши F4, предварительно установив курсор на адресе ячейки. Воспользовавшись маркером заполнения, копируем эту формулу в диапазон ячеек C4:C17. Получаем искомый ряд данных (рис. 12). В связи с тем, что количество заявок - целое число, следует установить на вкладке Число окна Формат ячеек числовой формат с числом десятичных знаков 0.

2. Теперь построим линейную регрессию, заданную уравнением:

$$y = mx+b$$

с помощью функции ЛИНЕЙН.

Для этого:

1. вводим в диапазон ячеек C20:D20 функцию ЛИНЕЙН как формулу массива: ={ЛИНЕЙН(B4:B14;A4:A14)}. В результате получаем в ячейке C20 значение параметра m , а в ячейке D20 - значение параметра b ;
2. вводим в ячейку D4 формулу: =\$C*A4+\$D;
3. копируем эту формулу с помощью маркера заполнения в диапазон ячеек D4:D17 и получаем искомый ряд данных.

3. Строим экспоненциальную регрессию, имеющую уравнение:

$$y = bmx$$

с помощью функции ЛГРФПРИБЛ оно выполняется аналогично:

1. в диапазон ячеек C21:D21 вводим функцию ЛГРФПРИБЛ как формулу массива: $=\{ \text{ЛГРФПРИБЛ} (B4:B14;A4:A14) \}$. При этом в ячейке C21 будет определено значение параметра m , а в ячейке D21 - значение параметра b ;
2. в ячейку E4 вводится формула: $=\$D*\$C^{\wedge}A4$;
3. с помощью маркера заполнения эта формула копируется в диапазон ячеек E4:E17, где и расположится ряд данных для экспоненциальной регрессии (см. рис. 12).

	A	B	C	D	E
1	Дневное поступление заявок на услуги в диспетчерскую				
2	автомобильного предприятия				
3	Дни	Количество заявок	Линейная регрессия (НАКЛОН +ОТРЕЗОК)	Линейная регрессия (ОБЪЕМЫ)	Экспоненциальная регрессия (ЛГРФПРИБЛ)
4	1	23	2	2	18
5	2	23	13	13	22
6	3	25	24	24	26
7	4	28	36	36	32
8	5	32	47	47	38
9	6	40	59	59	47
10	7	54	69	69	58
11	8	73	81	81	70
12	9	92	92	92	85
13	10	112	103	103	104
14	11	130	116	116	127
15	12	?	126	126	154
16	13	?	137	137	187
17	14	?	149	149	228
18					
19			11,309	-9,673	
20			11,309	-9,673	
21			1,217	14,532	

На рис. 13 приведена таблица, где видны используемые нами функции с необходимыми диапазонами ячеек, а также формулы.

Величина R^2 называется *коэффициентом детерминации*.

Задачей построения регрессионной зависимости является нахождение вектора коэффициентов m модели (1) при котором коэффициент R принимает максимальное значение.

Для оценки значимости R применяется F -критерий Фишера, вычисляемый по формуле

$$F = \frac{R^2}{(1-R^2)} \cdot \frac{n-k}{k-1},$$

где n - размер выборки (количество экспериментов);

k - число коэффициентов модели.

Если F превышает некоторое критическое значение для данных n и k и принятой доверительной вероятности, то величина R считается существенной. Таблицы критических значений F приводятся в справочниках по математической статистике.

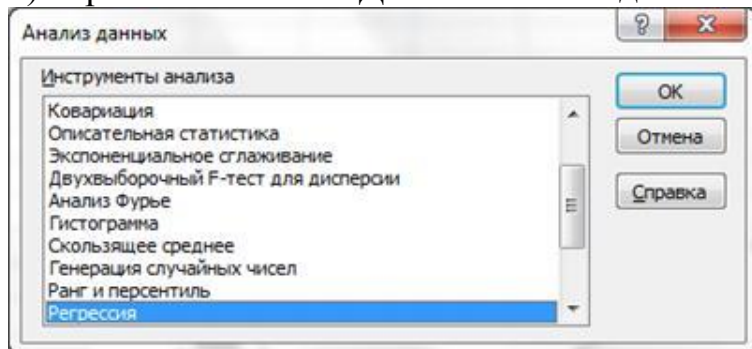
Таким образом, значимость R определяется не только его величиной, но и соотношением между количеством экспериментов и количеством коэффициентов (параметров) модели. Действительно, корреляционное отношение для $n=2$ для простой линейной модели равно 1 (через 2 точки на плоскости можно всегда провести единственную прямую). Однако если экспериментальные данные являются случайными величинами, доверять такому значению R следует с большой осторожностью. Обычно для получения значимого R и достоверной регрессии стремятся к тому, чтобы

количество экспериментов существенно превышало количество коэффициентов модели ($n > k$).

Для построения линейной регрессионной модели необходимо:

1) подготовить список из n строк и m столбцов, содержащий экспериментальные данные (столбец, содержащий выходную величину Y должен быть либо первым, либо последним в списке); для примера возьмем данные предыдущего задания, добавив столбец с названием "№ периода", пронумеруем номера периодов от 1 до 12. (это будут значения X)

2) обратиться к меню Данные/Анализ данных/Регрессия



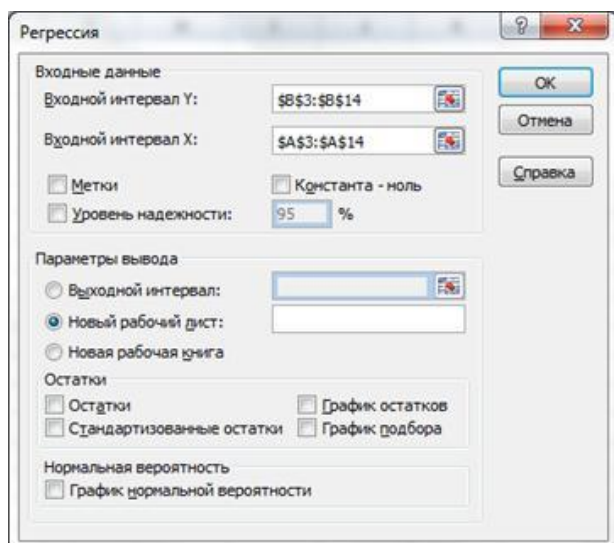
Если пункт "Анализ данных" в меню "Сервис" отсутствует, то следует обратиться к пункту "Надстройки" того же меню и установить флажок "Пакет анализа".

3) в диалоговом окне "Регрессия" задать:

входной интервал Y ;

входной интервал X ;

выходной интервал - верхняя левая ячейка интервала, в который будут помещаться результаты вычислений (рекомендуется разместить на новом рабочем листе);



4) нажать "Ok" и проанализировать результаты.

Пример использования множественной линейной регрессии

Предположим, что застройщик оценивает стоимость группы небольших офисных зданий в традиционном деловом районе.

Застройщик может использовать множественный регрессионный анализ для оценки цены офисного здания в заданном районе на основе следующих переменных.

y - оценочная цена здания под офис;

x_1 - общая площадь в квадратных метрах;

x_2 - количество офисов;

x_3 - количество входов (0,5 входа означает вход только для доставки корреспонденции);

x_4 - время эксплуатации здания в годах.

В этом примере предполагается, что существует линейная зависимость между каждой независимой переменной (x_1 , x_2 , x_3 и x_4) и зависимой переменной (y), то есть ценой здания под офис в данном районе. Исходные данные показаны на рисунке.

	A	B	C	D	E
1	x_1	x_2	x_3	x_4	y
2	2310	2	2	20	142000
3	2333	2	2	12	144000
4	2356	3	1,5	33	151000
5	2379	3	2	43	150000
6	2402	2	3	53	139000
7	2425	4	2	23	169000
8	2448	2	1,5	99	126000
9	2471	2	2	34	142900
10	2494	3	3	23	163000
11	2517	4	4	55	169000
12	2540	2	3	22	149000

Настройки для решения поставленной задачи показаны на рисунке окна "Регрессия". Результаты расчетов размещены на отдельном листе в трех таблицах



Регрессионная статистика

Множественный R	0,99837267
R-квадрат	0,99674799
Нормированный R-квадрат	0,99457999
Стандартная ошибка	970,578463
Наблюдения	11

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	4	1732393319	433098330	458,75367	1,37231E-07
Остаток	6	5652135,32	942022,55		
Итого	10	1738045455			

	Кэффиц иенты	Станд. ошибка	t-стату- стика	P-Значе- ние	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	52317,83	12237,36	4,28	0,0052328	22374,06	82261,60	22374,06	82261,60
Переменная X 1	27,64	5,43	5,09	0,0022410	14,36	40,93	14,36	40,93
Переменная X 2	12529,77	400,07	31,32	0,0000001	11550,84	13508,70	11550,84	13508,70
Переменная X 3	2553,21	530,67	4,81	0,0029663	1254,71	3851,71	1254,71	3851,71
Переменная X 4	-234,24	13,27	-17,65	0,0000021	-266,70	-201,77	-266,70	-201,77

В итоге мы получили следующую математическую модель:

$$y = 52318 + 27,64 \cdot x_1 + 12530 \cdot x_2 + 2553 \cdot x_3 - 234,24 \cdot x_4.$$

Теперь застройщик может определить оценочную стоимость здания под офис в том же районе. Если это здание имеет площадь 2500 квадратных метров, три офиса, два входа и время эксплуатации - 25 лет, можно оценить его стоимость, используя следующую формулу:

$$y = 27,64 \cdot 2500 + 12530 \cdot 3 + 2553 \cdot 2 - 234,24 \cdot 25 + 52318 = 158\,261 \text{ у.е.}$$

Контрольные вопросы:

1. Что описывает регрессионный анализ?
2. Как осуществляется регрессионно-факторный анализ
3. Дайте определение линии тренда.
4. Как построить математическую модель в виде эмпирической формулы?
5. Как оценить параметры линейной регрессии $Y = a + bx$?
6. Расскажите о методе Крамера
7. Какие возможности дает корреляционный анализ?

Практическое занятие №3

«Рассмотрение организма с позиций системного анализа»

Цель занятия

Целью занятия является изучение основ системного анализа применительно к живым биологическим организмам.

Подготовка к занятию

Самостоятельная подготовка студентов к практическому занятию *может* осуществляться в следующих направлениях:

- кибернетическая подсистема организма при рассмотрении его с позиций системного анализа;
- метаболическая подсистема организма при рассмотрении его с позиций системного анализа;
- виды биосистем;
- основные виды управления биотехническими системами.

Теоретическая часть

Говоря о биологических системах, часто применительно к ним употребляют термин “организм”. Но с позиций системного анализа организмом называют любую систему, обладающую собственными целями и способностью (ресурсами) для их достижения, т.е. целенаправленными действиями. С другой стороны, в настоящее время любой организм рассматривают как биохимическую машину с кибернетическим управлением, функционирование которой осуществляется за счет взаимодействия двух подсистем — метаболической и кибернетической.

Организм образуется из множества различных, качественно неоднородных элементов, каждый из которых играет в процессе жизнедеятельности строго определенную, зачастую незаменимую роль.

В общем виде взаимодействие основных элементов организма можно представить в виде схемы (рисунок 10) [1, 9]. В верхней части схемы расположена информационно-кибернетическая часть, в нижней — метаболическая подсистема (МП).

Верхняя часть представляет кибернетическую цепочку, состоящую из трех подсистем: рецепторная подсистема - управляющая подсистема - эффекторная подсистема. Именно эта часть системы

формирует поведение организма, осуществляя восприятие, хранение, переработку и использование информации. Информация из внешней среды поступает в рецепторную подсистему (рецепторы внешней среды), а далее сигналы от рецепторов доставляются в управляющую подсистему, функции которой выполняют рефлексy, а у высших животных - ЦНС. Управляющая подсистема является лидирующей и играет наиболее сложную роль в процессе управления всей системой.

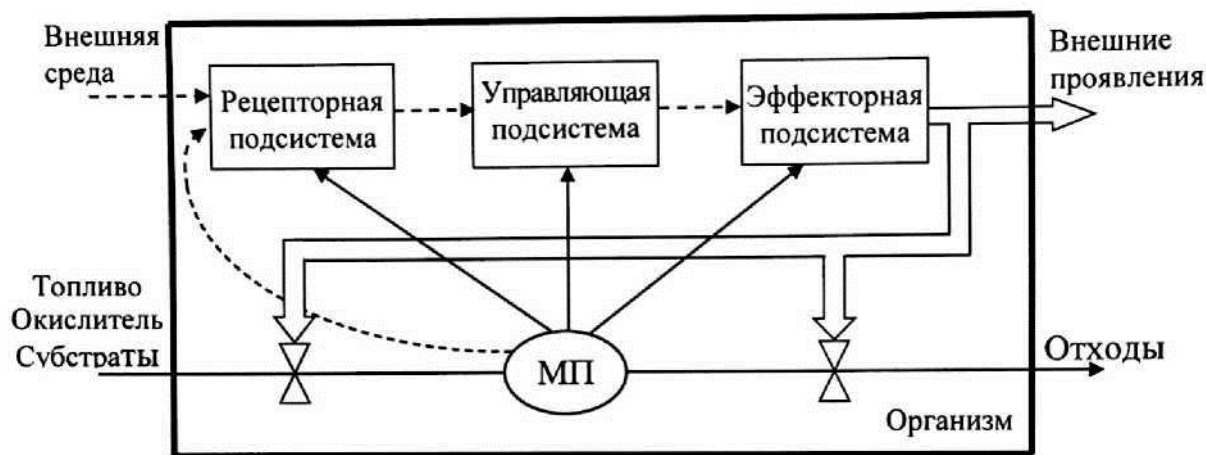


Рисунок 10 - Обобщенная структура живого организма

Доставляемые сведения позволяют поддерживать поведение организма в соответствии с условиями окружающей среды, но не менее важной является информация о состоянии самого организма, доставляемая из метаболической подсистемы через рецепторы внутренней среды.

Эта информация корректирует воздействия, вырабатываемые управляющей подсистемой в зависимости от потребностей организма. Информационные сигналы на рисунке обозначены пунктирными стрелками (рисунок 10), а кибернетические функции организма представлены в таблице 1.

Реализация поведенческих актов осуществляется через эффекторную подсистему. Как часть кибернетической системы у высших животных она включает органы движения, органы звуковой сигнализации (у человека - органы речи) и т. п. К ней же относят и все органы, оказывающие различного рода воздействия на внутреннюю сферу организма — железы внутренней секреции (поджелудочная, тимус), железы внешней секреции (например, потовые) и др. Действие эффекторов показано двойными стрелками (рисунок 10).

Таблица 1 - Кибернетические функции организма [1, 9]

Функция	Системы, реализующие функции
1. Восприятие информации	Сенсорные системы организма (зрение, слух, осязание и т.д.); рецепторы внутренней среды (хемотрецепторы, барорецепторы и т.п.)
2. Обработка информации, принятие решений, формирование программ поведения	Центральная нервная система (ЦНС)
3. Реализация принятых программ	Эффекторные системы организма (скелетно-мышечная, нервная, эндокринная, репродуктивная, органы речи или внешней сигнализации)

Кибернетическая система, представленная специальной системой органов, существует не у всех живых организмов. У низших животных из-за простоты управленческих функций также нет и постоянных элементов, которые выполняли бы сложные процессы управления. Подобные специальные элементы появляются у более сложных организмов, хотя при этом сохраняются и процессы управления без них. Усложнение кибернетической подсистемы организма идет параллельно с развитием специализированных органов, обеспечивающих все более дифференцированное снабжение метаболической подсистемы.

Нижняя часть схемы (рисунок 10) иллюстрирует *метаболическую подсистему* организма, которая представляет собой совокупность функционально и структурно связанных процессов преобразования химических веществ, протекающих в клетках

организма, и их транспортировки с целью обеспечения организма веществом и энергией для его жизнедеятельности, роста и размножения. В качестве входных величин для нее будут выступать вещества, доставляемые из окружающей среды. Дальнейшее перемещение и переработка доставленных продуктов и энергии в МП направлены на обеспечение функционирования кибернетической цепочки (и своего собственного).

Метаболическую подсистему организма иногда называют еще метаболическим “котлом”. В ней вещества, поступающие с пищей (субстраты), - белки, жиры, витамины, микроэлементы - преобразуются в более необходимые для организма, иногда сложные, а иногда и в более простые соединения.

Та часть процессов метаболизма, которая направлена на получение из простых веществ более сложных, называется *анаболизмом*. Компоненты, получаемые при анаболизме, идут на сборку элементов и ремонт клеточных структур. Процессы сборки можно относить к метаболической подсистеме, считая, что она осуществляет как синтез нужных веществ, так и формирование из них клеточных структур (мембран и других элементов), самих клеток и многоклеточных структур, необходимых для жизнедеятельности организма. Сам процесс сборки направлен не только на реставрацию отмирающих клеток или тканей. В молодом организме возникают новые органы, он растет и развивается. В организме, способном к деторождению, идет сборка половых клеток и синтезируются специальные вещества — секреты, необходимые для нормального оплодотворения. Эти процессы протекают непрерывно в течение довольно длительного времени в жизни живого организма.

Очень часто в организме возникает необходимость в реализации программ экстренной сборки. Почти постоянно в него вторгаются вирусы или бактерии, проникают чужеродные белки - антигены. Ответная реакция организма — сборка специфических белков-антител, строение которых каждый раз строго отвечает вторгшемуся антигену. При встрече с чужеродным белком антитела связывают его, но при разрушении чужеродных клеток их содержимое остается во внутренней среде организма. Также во внутренние межклеточные жидкости поступают непрерывно и все “осколки” собственных клеток организма, выходящих из строя.

Если это вещества, которые нельзя непосредственно использовать для синтеза нужных макромолекул или для сборки новых конструкций, подобные отходы разлагаются на более простые компоненты. Процессы разложения сложных веществ в метаболической подсистеме на более простые получили название *катаболизма*. Вещества, образующиеся при катаболизме, могут использоваться непосредственно или выступать в качестве строительных блоков, а ненужные вещества приобретают такие

свойства (например, растворимость), которые облегчают их окончательное выведение из организма.

Процессы анаболизма и катаболизма контролируются кибернетической подсистемой, поэтому сам факт протекания определенных процессов в МП и их скорость определяются именно управлением. Сплошными стрелками, направленными к МП и от нее, показано поступление в организм и выведение из него различных веществ. Эти процессы выполняются специальными физиологическими системами, но управляются они тоже кибернетической подсистемой. Вместе с тем в управлении процессами, происходящими в МП, определенную роль играют и механизмы “низшего уровня”, сосредоточенные в самой МП, - так называемые механизмы ауторегуляции (саморегуляции).

Функции организма как открытой метаболической системы, обменивающейся с окружением веществами и энергией (таблица 2), свойственны как отдельным клеткам, так и сложным организмам.

Таблица 2 - Метаболические функции организма [1,9]

Функция	Вещества	Системы организма
1	2	3
1. Доставка из окружающей среды вещества	<p>Субстраты: белки, липиды, углеводы, витамины, незаменимые аминокислоты</p> <p>Топливо: углеводы</p> <p>Окислитель: кислород</p> <p>Транспорт внутри организма</p>	<p>Физиологический комплекс: желудочно-кишечный тракт, система пищеварения</p> <p>Те же системы</p> <p>Дыхательная система</p> <p>Сердечно-сосудистая система</p>
2. Производство энергии	АТФ	Биохимическая система: митохондрии

3. Обмен веществ	Анаболизм: биополимеры Катаболизм: высокомолекулярные отходы, неклеточные элементы	Биохимическая система: аппарат Гольджи, эндоплазматическая сеть Клеточные мембраны
4. Выведение конечных продуктов из организма	Отходы: азот, креатинин, мочевина, вода, углекислота, билирубин Наследственный материал	Физиологический комплекс: почки, системы дыхания и кровообращения, печень и желчная система Репродуктивная система

Причем для управления поведением и состоянием организма, как объекта управления, используются различные управляющие воздействия.

К основным методам управления, используемым в живых системах, относят энергетический, вещественный и информационный [4].

Энергетическое управление предполагает воздействие на биологическую систему в целом или на ее подсистемы физических управляющих агентов, не изменяющих количества вещества биологического объекта. К этим агентам относятся физические поля: электрическое, магнитное, тепловое, акустическое, радиационное, электромагнитное. Воздействие может осуществляться как контактно через электроды, так и бесконтактным способом. Суть энергетического управления заключается в стимулировании функционирования отдельных подсистем организма или подавлении некоторых патологических процессов, протекающих в больном организме.

В отличие от энергетического управления, исключающего введение в организм каких-либо материальных управляющих агентов, *вещественное управление* использует самые различные фармакологические, гормональные, химические и другие агенты в твердом, жидком и газообразном состоянии для управления состоянием живого организма и его отдельными функциональными подсистемами. Таким образом, вещественное управление предусматривает непременно изменение количества вещества, содержащегося в организме.

Информационное управление - это управление состоянием человека с помощью воздействия специально сформированных потоков информации. Этот метод управления является наиболее эффективным, но и наименее количественно формализованным.

2.4 Контрольные тестовые задания

1 Выберите подсистемы, которые включает в себя обобщенная структура животного организма:

- а) метаболическая;
- б) информационная;
- в) контрольно-измерительная;
- г) кибернетическая

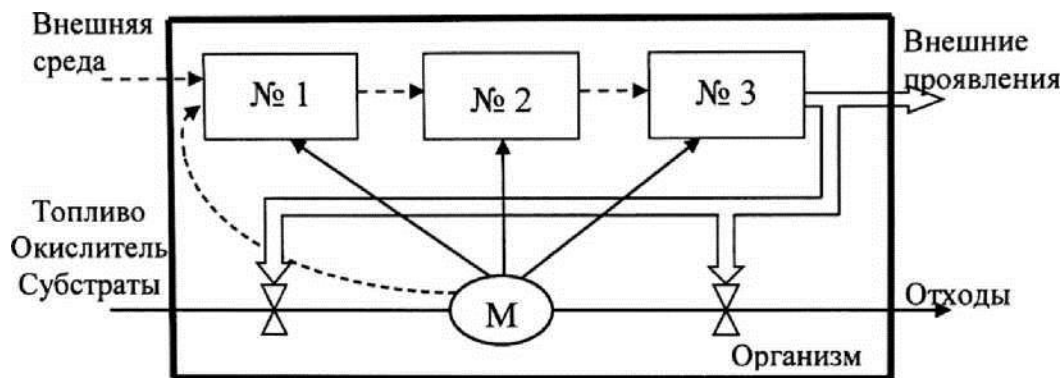
2 Укажите какие подсистемы входят в состав кибернетической подсистемы:

- а) эффекторная подсистема;
- б) управляющая подсистема;
- в) рецепторная подсистема;
- г) управляемая подсистема.

3 Отметьте название воздействий на биологическую систему в целом или на ее подсистемы физических управляющих агентов, не изменяющих количества вещества биологического объекта:

- а) вещественное управление;
- б) энергетическое управление;
- в) информационное управление.

4 Укажите название подсистемы №1 кибернетической подсистемы, обозначенной на рисунке:



а) рефлексивная подсистема;

б) рецепторная подсистема;

в) управляющая подсистема;

г) эффекторная подсистема;

5 Выделите формулировку первого принципа организации различных систем:

а) деятельность организма — это последовательное

функционирование разнообразных подсистем регулирования,

подчиненное главной функции - обеспечению выживания в условиях изменяющейся внешней среды;

б) деятельность организма — это одновременное функционирование разнообразных подсистем регулирования, подчиненное главной функции - обеспечению выживания в условиях изменяющейся внешней среды;

в) деятельность организма - это параллельное функционирование разнообразных подсистем регулирования, подчиненное главной функции - обеспечению выживания в условиях изменяющейся внешней среды.

Контрольные вопросы:

1. Дайте понятие “организма” с позиции системного анализа.
2. Что такое информационное управление и как оно осуществляется?
3. Что такое энергетическое управление и как оно осуществляется?
4. Как контролируются процессы анаболизма и катаболизма?
5. Опишите метаболическую подсистему организма
6. Опишите кибернетическую подсистему организма
7. Нарисуйте и опишите обобщенную структуру живого организма

Практическое занятие №4

«Анализ вариационных рядов при выявлении характера распределения изучаемых явлений»

Характер распределения изучаемых явлений, как правило, выявляют при анализе вариационных рядов, которые в силу этого носят ещё название рядов распределения.

Вариационный ряд – это ряд числовых измерений определенного признака, отличающихся друг от друга по своей величине, расположенных в определенном порядке.

Вариационный ряд состоит из вариантов (V) и соответствующих им частот (p). Вариантом (V) называют каждое числовое значение изучаемого признака. Частота (p) – абсолютная численность отдельных вариантов в совокупности, указывающая, сколько раз встречается данная варианта в вариационном ряду.

Если исследователь имеет не более 30 наблюдений, то достаточно все значения признака расположить в нарастающем или в убывающем порядке (от максимальной варианты до минимальной или наоборот) и указать частоту каждой варианты. При большом числе наблюдений (более 30) вариационный ряд рекомендуется сгруппировать.

Построение сгруппированного ряда складывается из нескольких этапов: 1) определение количества групп; 2) определение интервала между группами; 3) определение начала, середины и конца группы; 4) распределение данных наблюдений по группам; 5) графическое изображение вариационного ряда.

I этап: определение количества групп в вариационном ряду. В связи с тем, что количество групп зависит от числа наблюдений, совершенно ясно, что чем больше число наблюдений, тем больше может быть групп. На основании специальных расчетов с учетом этой взаимосвязи составлена специальная таблица, в которой указывается, при каком числе наблюдений сколько должно быть групп в вариационном ряду, чтобы характерные особенности распределения изучаемого явления не были завуалированы.

II этап: определение величины интервала (i) между группами. Обязательным требованием при построении вариационного ряда является соблюдение единого

интервала. Определяя величину интервала между группами, амплитуду вариационного ряда (разность между максимальным и минимальным значениями вариантов) делят на число групп.

III этап: определение начала, середины и конца группы. Прежде всего необходимо определить середину для первой группы. В нашем примере максимальная варианта равна 64. Поскольку середина группы должна делиться на величину интервала, то, соблюдая это требование, за середину первой группы следует взять варианту, равную 65, которая будет ближайшей к максимальному значению и без остатка разделится на величину интервала, равного «а». Середины для других групп находят следующим образом: от середины каждой предыдущей группы отнимают величину интервала. Так, если середина первой группы – 65, то середина второй группы равна 60 (65-5), середина третьей группы – 55 (60-5) и т. д. После составления ряда из величин, принятых за середину группы – 65, 60, 55, 50 и т. д., нужно определить границы (начало и конец) этих групп.

Определяя начало группы, к её середине прибавляют величину $(i - 1)/2$, вычитая же её из середины, получают конец группы. В нашем примере

$(i - 1)/2 = (5-1)/2 = 2$. Прибавив 2 к середине первой группы, получим 67 (начало группы), её концом будет: $65-2=63$.

IV этап: распределение случаев наблюдения по группам.

V этап: графическое изображение вариационного ряда. Для углубленного анализа полученных данных большее значение имеет правильное построение графического изображения вариационных рядов.

Оно позволяет определить моду (соответствует вершине), разнообразие признака в совокупности.

Средние величины – второе свойство статистической совокупности

Средний уровень измеряют с помощью критериев, которые носят название средних величин. Под средней величиной понимают число, выражающее общую меру исследуемого признака в совокупности. Средняя величина как бы выражает то общее,

что характерно для признака в данной совокупности.

Общепотребительными являются три вида средних величин: мода (M_o), медиана (M_e), средняя арифметическая (M).

Мода (M_o) – соответствует величине признака, которая чаще других встречается в данной совокупности. Иначе говоря, за моду принимают варианту, которой соответствует наибольшее количество частот (P) вариационного ряда.

Медиана (M_e) – величина признака, занимающая срединное положение в вариационном ряду. Она делит ряд на две равные части по числу наблюдений. Для определения медианы надо найти середину ряда. При четном числе наблюдений за медиану принимают среднюю величину из двух центральных вариантов. При нечетном числе наблюдений медианой будет срединная (центральная) варианта, которая определяется так: $(n+1)/2$ или, например, $(25+1)/2=13$. Это означает, что середина ряда приходится на тринадцатую варианту с начала ряда или тринадцатую варианту с конца ряда.

Средняя арифметическая величина опирается на все наблюдения и рассчитывают её несколькими способами в зависимости от численности вариант, характера вариационного ряда и наличия вычислительной техники.

Основными способами расчета M_e является: среднеарифметический способ и способ моментов (условных отклонений).

Среднеарифметический способ применяется для вычисления средней арифметической простой и средней арифметической взвешенной.

Средняя арифметическая простая – вычисляется из вариационного ряда, в котором каждая варианта встречается только один раз (для всех вариант $p=1$); средняя арифметическая взвешенная вычисляется из вариационного ряда, в котором отдельные варианты встречаются различное число раз: $p \geq 1$.

Способ моментов. Применяя этот способ, среднюю арифметическую рассчитывают по формуле:

$$M = A + i \frac{\sum a \times p}{n}$$

Эта формула технически упрощает расчеты, особенно в тех случаях, когда варианты состоят из многозначных чисел, а совокупность – из большого числа наблюдений.

$$\sum a \times p = -10 \text{ кг};$$

Этапы расчета M по способу моментов: 1) за условную среднюю A рекомендуется принять варианту, чаще других повторяющуюся в вариационном ряду, например, $A=M_0=62$ кг, так как 62 кг было у 9 юношей из 25; 2) определяем a – условное отклонение от условной средней; для этого из каждой варианты вычитаем условную среднюю: $a = (V - A)$, например,

$a=64-62=+2$; 3) умножаем условное отклонение (a) на частоту (p) каждой варианты и получаем произведения ($a \times p$); 4) получаем сумму:

$$\sum ap / n = (-10) / 25 = -0,4 \text{ кг};$$

5) определяем среднее отклонение от условной средней (момент первой степени):

6) определяем интервал между группами вариант: $i = 1$ кг;

7) рассчитываем среднюю арифметическую по способу моментов:

$$M = A + I \frac{\sum a \times p}{n} = 62 \text{ кг} - 1 \times 0,4 \text{ кг} = 61,6 \text{ кг}$$

Таким образом, можно сделать вывод о том, что изучаемые нами юноши имеют среднюю массу тела 61,6 кг.

Средняя арифметическая одним числом характеризует совокупность, обобщая то, что свойственно всем её вариантам, поэтому она имеет ту же размерность, что и каждая из вариантов.

Средняя арифметическая величина обладает тремя свойствами.

1. Средняя занимает срединное положение в вариационном ряду. В строго симметричном ряду $M=M_0=M_e$.

2. Средняя является обобщающей величиной и за средней не видны случайные колебания, различия в индивидуальных данных, она вскрывает то типичное, что характерно для всей совокупности.

3. Сумма отклонений всех вариантов от средней равна нулю: $(V-M)=0$.

Это происходит потому, что средняя величина превышает размеры одних вариантов и меньше размеров других вариантов.

Разнообразие признака – третье свойство статистической совокупности

Величина того или иного признака неодинакова у всех членов совокупности, несмотря на её относительную однородность. Например, в группе детей, однородной по возрасту, полу и месту жительства, рост каждого ребенка отличается от роста сверстников.

В этом проявляется разнообразие, колеблемость признака в изучаемой совокупности.

Статистика позволяет охарактеризовать это специальными критериями, определяющими уровень разнообразия каждого признака в той или иной группе. К таким критериям относятся лимит (lim), амплитуда ряда (A_m), среднее квадратическое отклонение (σ^2) и коэффициент вариации (C_v). Так как каждый из этих критериев имеет свое самостоятельное значение, то следует остановиться на них отдельно.

Лимит (lim) определяется крайними значениями вариантов в вариационном ряду. Например, если масса тела изменяется от 55 до 105 кг, то

$$M_T = 55 \div 105 \text{ кг.}$$

Амплитуда (A_m) – разность крайних вариантов.

Наиболее полную характеристику разнообразия признака в совокупности дает так называемое среднее квадратическое отклонение, обозначаемое греческой буквой «сигма» – σ .

Существуют два способа расчета среднего квадратического отклонения: среднеарифметический и способ моментов. При среднеарифметическом способе расчета применяют формулу:

$$\sigma = \sqrt{\frac{\sum d \times d}{n - 1}}$$

где d истинное отклонение вариант от истинной средней ($V-M$). Эта формула используется при небольшом числе наблюдений ($n \leq 30$), когда в вариационном ряду все частоты $p=1$. Коэффициент вариации (C_v) является относительной мерой разнообразия, так как исчисляется как процентное отношение среднего квадратического отклонения (σ) к средней арифметической величине (M). Формула коэффициента вариации:

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

Для ориентировочной оценки степени разнообразия признака пользуются следующими градациями коэффициента вариации. Если коэффициент составляет более 20%, то отмечают сильное разнообразие; при 20-10% – среднее, и если коэффициент менее 10%, то считают, что разнообразие слабое.

Коэффициент вариации применяют при сравнении степени разнообразия признаков, имеющих различия в величине признаков или неодинаковую их размерность. Допустим, необходимо сравнить степень разнообразия массы тела у новорожденных и 7-летних детей. Понятно, что у новорожденных она всегда будет меньше, чем у семилетних детей, так как меньше их индивидуальная масса. Среднее квадратическое отклонение будет меньше там, где меньше величина самого признака (σ новорожденных = $\pm 0,35$ кг, а σ семилетних мальчиков = $\pm 3,88$ кг). В этом случае для определения различия в степени разнообразия необходимо ориентироваться не на среднее квадратическое отклонение (σ), а на относительную меру разнообразия – коэффициент вариации C_v .

Большое значение коэффициент вариации также имеет для оценки и сопоставления степени разнообразия нескольких признаков с разной размерностью. Например, необходимо оценить какой признак (масса тела, рост, количество лейкоцитов или СОЭ) является наиболее разнообразным.

Среднее квадратическое отклонение связано со структурой ряда распределения признака.

Теорией статистики доказано, что при нормальном распределении в пределах $M \pm \sigma$ находится 68% всех случаев, в пределах $M \pm 2 \sigma$ – 95,5% всех случаев, в пределах $M \pm$

σ – 99,7% всех случаев, составляющих совокупность. Таким образом, $M \pm 3\sigma$ охватывает почти весь вариационный ряд.

Это теоретическое положение статистики о закономерностях структуры ряда имеет огромное значение для практического применения среднего квадратического отклонения. Можно воспользоваться этим правилом для выяснения вопроса о типичности средней величины. Если 95% всех вариантов находятся в пределах $M \pm 2\sigma$, то средняя является характерной для данного ряда и не требуется увеличивать число наблюдений в совокупности. Для определения типичности средней сравнивается фактическое распределение с теоретическим путем расчета сигмальных отклонений.

Практическое значение среднего квадратического отклонения заключается также в том, что, зная M и σ , можно построить вариационные ряды и рассчитать количество одежды и обуви разных размеров, необходимых для детей, подростков, военнослужащих, физическое развитие которых было изучено. Сигму (σ) также используют для сравнения степени разнообразия однородных признаков, например, при сравнении колеблемости (вариабельности) роста юношей 17 лет в городе и сельской местности.

Зная сигму (σ), можно рассчитать коэффициент вариации (C_v), необходимый для сравнения степени разнообразия признаков, выраженных в различных единицах измерения (сантиметрах, килограммах и т.д.). Это позволяет выявить более устойчивые (постоянные) и менее устойчивые признаки в совокупности.

Среднее квадратическое отклонение (σ) используется также для оценки отдельных признаков у каждого индивидуума.

Стандартное отклонение указывает, на сколько сигм (σ) от средней (M) отклоняются индивидуальные измерения.

Среднее квадратическое отклонение также может быть использовано в клинике при разработке проблемы нормы и патологии.

Наконец, среднее квадратическое отклонение является важным компонентом формулы m_m средней ошибки средней арифметической (ошибки

репрезентативности): $m_m = \sigma / \sqrt{n}$, где m – средняя ошибка средней арифметической величины (ошибка репрезентативности), n – число наблюдений.

Задание 1. Составление простого вариационного ряда, определение моды и медианы.

Типовое задание

При изучении веса у 11 обследованных получены следующие результаты: 68, 65, 65, 64, 76, 75, 65, 64, 72, 72, 71 кг.

На основе приведенных данных необходимо:

- составить простой вариационный ряд,
- определить моду,
- определить медиану,
- сформулировать вывод.

Образец выполнения задания

1. Простой вариационный ряд строим, располагая в порядке возрастания значений вариант (можно в порядке убывания):

64 64 65 65 65 68 71 72 72 75 76.

2. Модой является та варианта, которая чаще других встречается в данном вариационном ряду, т. е. мода равна 65 кг.

3. Медиана в данном нечетном ряду ($n=11$) определяем как $n+1/2 = 11+1/2=6$, т. е. 6-я по счету от начала или конца вариационного ряда варианта будет являться медианой. В нашем случае она равна 68 кг.

4. Вывод: У 11 обследованных лиц наиболее часто встречаются вес, равный 65 кг. Медиана в данном примере равна 68 кг. Мода и медиана не равны, что свидетельствует о несимметричности вариационного ряда.

Задание 2. Вычисление средней арифметической простой и средней арифметической взвешенной.

Типовое задание

На основе приведенных данных в задании 1 требуется определить:

1. среднюю арифметическую простую,
2. среднюю арифметическую взвешенную,
3. сделать вывод.

Образец выполнения задания

1. Среднюю арифметическую простую определяем по формуле

$$M = \frac{\sum V}{n} = \frac{68 + 65 + 65 + 64 + 76 + 75 + 65 + 64 + 72 + 72 + 71}{11} = \frac{757}{11} = 68,82 \text{ кг}$$

2. Среднюю арифметическую взвешенную вычисляем по формуле

$$3. M = \frac{\sum V * P}{n} = \frac{(64 * 2) + (65 * 3) + (68 * 1) + (71 * 1) + (72 * 2) + (75 * 1) + (76 * 1)}{11} = \frac{128 + 195 + 68 + 71 + 144 + 75 + 76}{11} = \frac{757}{11} = 68,82 \text{ кг}$$

3. Вывод: Средняя арифметическая величина, рассчитанная различными способами, является одинаковой и равна 68,82 кг.

Задание 3. Определение среднего квадратического отношения и коэффициента вариации.

Типовое задание

На основе данных приведенных в задании 1 необходимо:

1. вычислить среднее квадратическое отношение,
2. вычислить коэффициент вариации,
3. сделать вывод.

Образец выполнения задания

1. Среднее квадратическое отношение определяем по формуле:

$$G = \sqrt{\sum d^2 / n - 1} \quad , \text{ где } d = V - M.$$

Средняя арифметическая (M), полученная в задании 2, равна 68,82 кг. С учетом этого определяем сигму.

№ п/ п	V	d	d ²
1	64	-4,82	23,23
2	64	-4,82	23,23
3	65	-3,82	12,22
4	65	-3,82	12,22
5	65	-3,82	12,22
6	68	-0,82	0,67
7	71	+2,18	4,75
8	72	+3,18	10,11
9	72	+3,18	10,11
10	75	+6,18	38,19
11	76	+7,18	51,55

$$\Sigma d^2 = 23,23 + 23,23 + 12,22 + 12,22 + 12,22 + 0,67 + 4,75 + 10,11 + 10,11 + 38,19 + 51,55 = 198,5$$

$$G = \sqrt{\frac{198,5}{11-1}} = \sqrt{\frac{198,5}{10}} = \sqrt{19,85} = 4,46$$

2. Коэффициент вариации определяем по формуле:

$$C_v = \frac{G}{M} \times 100\% = \frac{4,46}{68,82} \times 100\% = 0,064 \times 100\% = 6,4\%$$

3. На основе полученных результатов можно утверждать о слабом разнообразии веса у обследованных лиц, так как коэффициент вариации менее 10%.

ВАРИАНТ 1.

При определении содержания белка в крови у 12 больных установлено, что количество составило: 80, 60, 64, 72, 68, 90, 80, 68, 68, 71, 120, 68 г/л.

ВАРИАНТ 2.

Частота пульса у 15 новорожденных составила: 120, 140, 160, 140, 110, 140, 120, 110, 100, 115, 105, 108, 112, 108, 140 ударов в минуту.

ВАРИАНТ 3.

На экзамене по общественному здоровью и здравоохранению студентами 5-й группы получены следующие оценки: 5, 4, 4, 3, 3, 2, 2, 5, 4, 3, 4, 5, 4, 4.

ВАРИАНТ 4.

Длительность лечения пневмонии у 13 больных в городской больнице № 1 составила: 15, 18, 13, 21, 23, 18, 16, 17, 18, 20, 21, 18, 14 дней.

ВАРИАНТ 5.

Оборот койки в неврологическом отделении за последние 12 лет изменялся так: 15, 18, 21, 18, 14, 17, 18, 20, 21, 16, 17, 18 больных.

ВАРИАНТ 6.

Систолическое артериальное давление у 15 обследованных больных гипертонической болезнью составило: 160, 180, 150, 165, 200, 180, 170, 200, 220, 180, 150, 210, 220, 200, 180 мм.рт.ст.

ВАРИАНТ 7.

Уровень рождаемости на 18 административных территориях варьировал: 14, 18, 13, 15, 8, 8, 7, 7, 13, 15, 8, 12, 10, 11, 12, 8, 7, 14 случаев на 1000 населения.

ВАРИАНТ 8.

Масса тела у 14 новорожденных составила: 3200, 3100, 3200, 4000, 4100, 3800, 3200, 4000, 3500, 3600, 3800, 2800, 3000, 2500 г.

ВАРИАНТ 9.

Среднегодовая занятость койки в различных отделениях больницы в 2000 году составила: 310, 280, 340, 340, 310, 340, 290, 295, 320, 315, 330 дней.

ВАРИАНТ 10.

Частота сердечных сокращений на 1-ой минуте после выполнения упражнений максимальной анаэробной мощности у спортсменов составила: 170, 180, 160, 170, 165, 170, 164, 160, 170, 182, 148, 150 ударов в минуту.

Контрольные вопросы:

1. Что такое вариационный ряд и из чего он состоит?
2. Из каких этапов складывается построение сгруппированного ряда?
3. Что такое средние величины?
4. Дайте определение понятиям: мода (M_o) и медиана (M_e).
5. Что такое средняя арифметическая простая?
6. Какими свойствами обладает средняя арифметическая величина?
7. Какие существуют способы расчета среднего квадратического отклонения?

Практическое занятие №5

«Оценка достоверности результатов исследования»

Для характеристики изучаемой совокупности широко используют относительные величины.

Относительные величины весьма распространены и постоянно применяются в медицине и здравоохранении. Они необходимы прежде всего для сравнения и сопоставления одной совокупности с другой. Однако это не означает, что абсолютные величины вообще не применяются при анализе. Абсолютные величины сами по себе несут важную информацию о размере того или иного явления: количестве больных, родившихся, числе коек в стационаре и т. д.

Но чаще всего абсолютные величины без преобразования их в относительные показатели имеют ограниченное познавательное значение. Например, в Мантуровском районе было зарегистрировано в 1982 г. 25 случаев дизентерии, а в Дмитриевском районе – 10 случаев этого заболевания. По этим абсолютным величинам нельзя сделать вывод о том, что население Мантуровского района болеет дизентерией чаще, чем Дмитриевского района. Чтобы сделать правильный вывод, необходимо учесть различия в численности населения этих двух районов, что достигается путем преобразования абсолютных величин в относительные величины.

Относительные величины рассчитываются путем отношения (деления) одной абсолютной величины на другую и полученную дробь умножают на 100 (или 1000, 10000 и т. д.). Соответственно этому относительные величины могут быть выражены в процентах (%), промилле (‰) или продецимилле ‰ и т. д.). При этом подбор того или иного множителя связан с тем, что относительные величины целесообразно представлять в целых числах, легко воспринимающихся при анализе.

Различают следующие виды относительных величин: интенсивные, экстенсивные показатели, показатели соотношения и наглядности.

Интенсивный показатель, или показатель частоты, распространенности, указывает на частоту изучаемого явления в среде, непосредственно «производящей» данное

явление.

$$ИП = \frac{\text{абсолютная численность изучаемого явления}}{\text{абсолютная численность среды}} \times 100\% (1000, 10000, 100000)$$

Примерами интенсивных величин являются: уровень рождаемости, смертности, заболеваемость раком, сифилисом и т. д.

Интенсивные показатели применяются в следующих случаях:

- для определения уровня, частоты, распространенности того или иного явления;
- для сравнения ряда различных совокупностей по степени частоты того или иного явления (например, для сравнения уровней рождаемости в разных странах, в разных районах или для сравнения уровней заболеваемости мужчин и женщин, уровней смертности в разных возрастных группах и т. д.);
- для выявления в динамике изменений в частоте явления в наблюдаемой совокупности (например, сдвига в распространенности инфекционных заболеваний населения области Н. за ряд лет и др.).

Интенсивные показатели могут быть наглядно представлены в виде следующих диаграмм: столбиковой, линейной, радиальной, картограммы и картодиаграммы.

Столбиковая диаграмма применяется для иллюстрации однородных, но не связанных между собой интенсивных показателей. Столбиковыми диаграммами изображают статику явления (рис. 1).

Линейная диаграмма применяется для иллюстрации частоты явления, изменяющегося во времени. Линейная диаграмма как бы символизирует непрерывность наблюдения и обычно употребляется для изображения динамики явления. Линейная диаграмма иллюстрирует значения ряда величин, нанесенных в виде точек на систему координат и соединенных линиями (рис. 2).

Радиальная диаграмма является частным видом линейной диаграммы, построенной на полярных координатах. Радиальной диаграммой пользуются при необходимости изобразить графически динамику явления за замкнутый цикл времени (сутки, неделя, год) (рис. 3).

Картограмма – особая географическая карта, на которой отдельные территории заштрихованы с различной интенсивностью соответственно уровню интенсивного показателя. Например, карта Орловской области, где изображен уровень рождаемости или смертности различной интенсивности штриховкой.

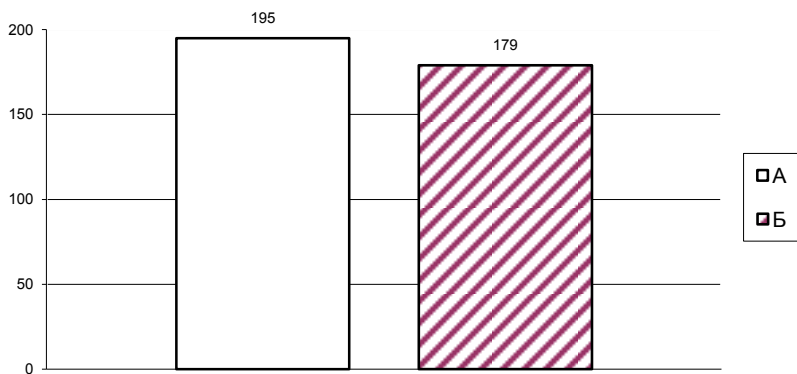


Рис. 1. Уровень заболеваемости сифилисом мужского (А) и женского (Б) населения в 1999 г. в России

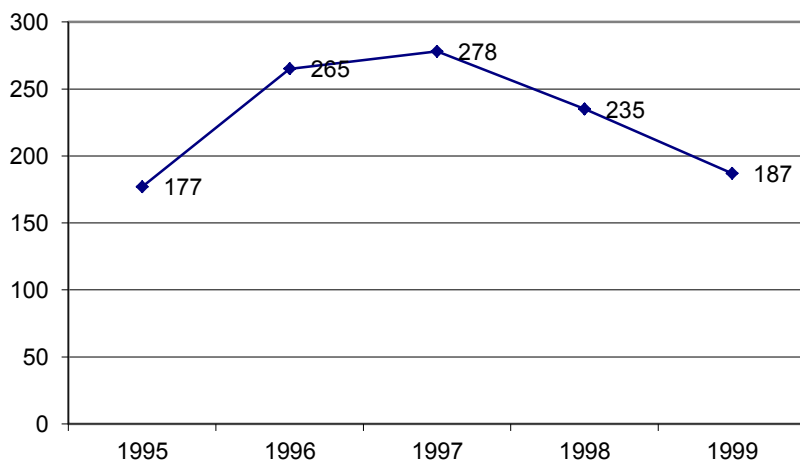


Рис. 2. Динамика заболеваемости сифилисом в России за 1995-1999 гг. (на 100000 населения)

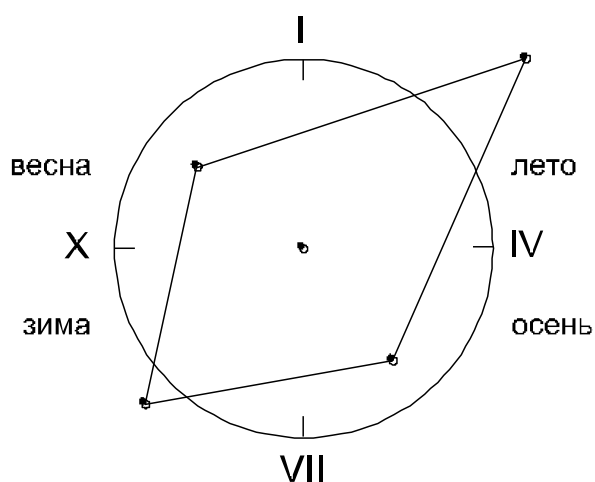


Рис. 3. Уровень травматизма в различные сезоны года в 1998 г. (%)

Картодиаграмма представляет собой сочетание географической карты с диаграммой, чаще всего столбиковой, причем столбики различной величины (соответственно показателю) наносятся на карту и ставятся на той территории, которую они представляют.

Экстенсивный показатель, или показатель структуры распределения, указывает на отношение части к целому, на долю части в целом. Экстенсивный показатель дает представление о количественном распределении составных (структурных) частей в какой-либо одной совокупности.

$$\frac{\text{Часть явления}}{\text{Явление в целом}} \times 100\%$$

Методика расчета экстенсивного показателя. Прежде всего следует составить пропорцию: все заболевшие (целое явление) – 105 = 100%, а заболевшие корью (часть явления) – 30 = x, тогда доля случаев кори среди всех заболеваний составит:

$$(30/105) \times 100\% = 28,6\%$$

Так как экстенсивные показатели характеризуют структуру явления, т. е. дают представление об удельном весе части в целом, то графически они могут быть изображены секторной (рис. 4), внутрестолбиковой (рис. 5, 6) или внутривенточной диаграммой (рис. 7).

Показатель соотношения характеризует отношение между двумя самостоятельными совокупностями (в этом его сходство с интенсивным показателем), причем независимые совокупности не только связаны друг с другом, но и не продуцируют одна другую (в этом отличие показателя соотношения от интенсивного коэффициента). Показателями соотношения являются показатели обеспеченности населения врачами, медсестрами, койками, рассчитанные на 1000, 10000 населения. Их широко используют при планировании здравоохранения.

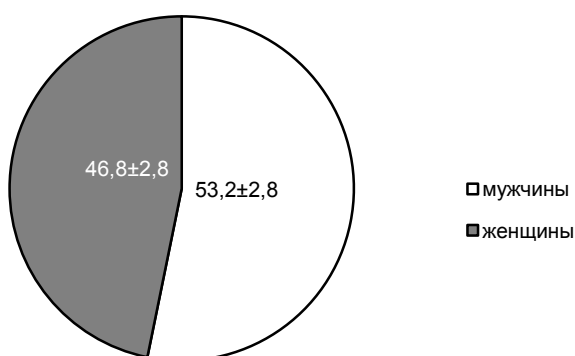


Рис. 4. Удельный вес мужчин среди больных с заболеваниями нервной системы ($P \pm mр, \%$)



Рис. 5. Доля заболеваний периферической нервной системы в структуре неврологической патологии ($P \pm mр, \%$)

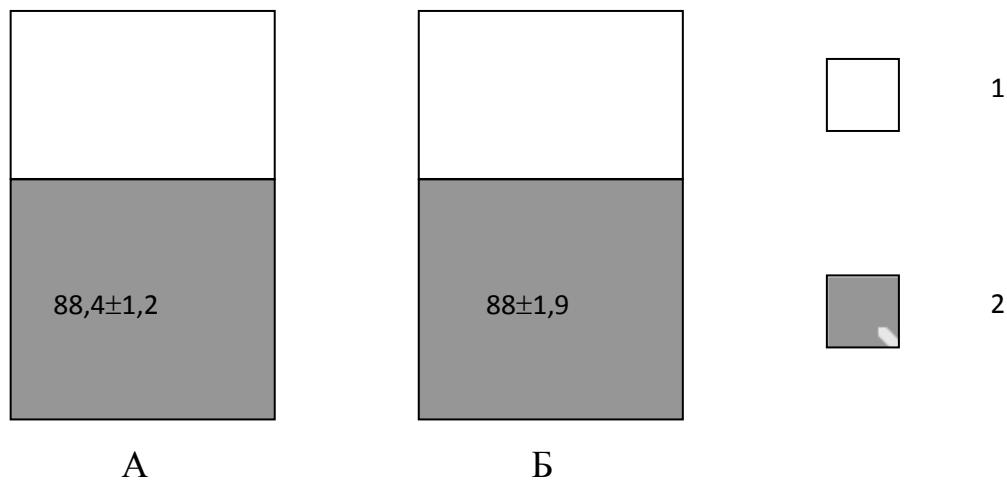


Рис. 6. Удельный вес пациентов с сопутствующей патологией в опытной (А) и контрольной (Б) группах

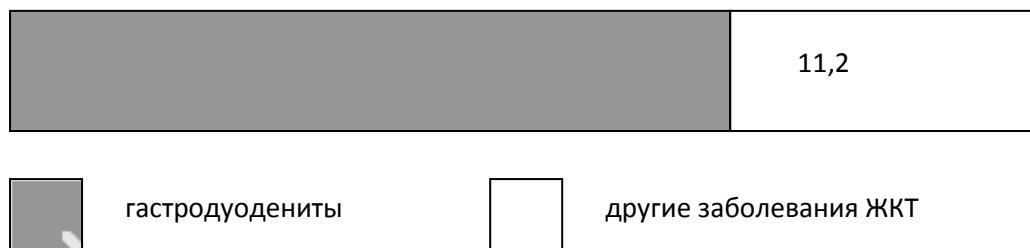


Рис. 7. Структура заболеваний ЖКТ у детей (в %)

Для графического изображения показателей соотношения применяют те же диаграммы, что и для интенсивных (столбиковые и линейные диаграммы, картограммы и картодиаграммы).

Показатели наглядности указывают, на сколько процентов или во сколько раз произошло увеличение или уменьшение сравниваемых величин. Показатели наглядности получают при отношении ряда сравниваемых величин к одной из них, принятой за 100. Как правило, за такую исходную величину берут начальные или конечные числа ряда, чтобы показатели наглядности иллюстрировали тенденцию

повышения или снижения (рис. 8).

В показателях наглядности можно представить абсолютные величины, интенсивные показатели, показатели соотношения, а также средние величины. Принципы графического изображения показателей наглядности такие же, как и интенсивных величин.

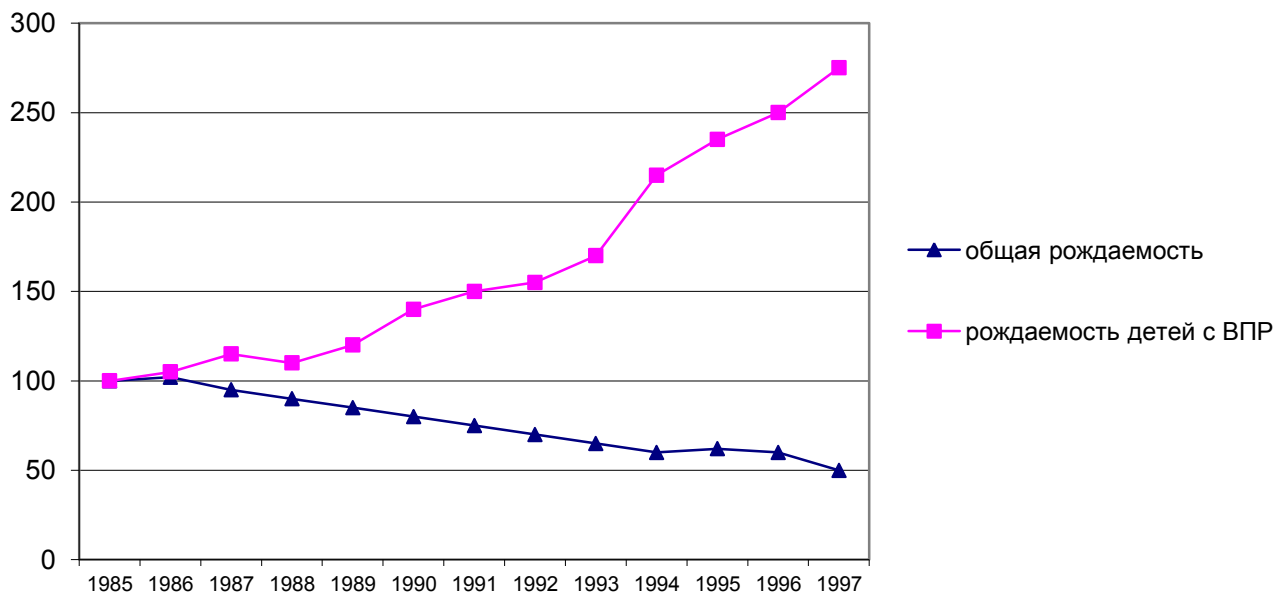


Рис. 8. Показатели наглядности общей рождаемости и рождаемости детей с ВПР (в % к 1985 г.)

Задание 1. Вычисление, графическое изображение и анализ относительных величин.

Типовое задание

Таблица 1. Число обращений и посещений по поводу мочекаменной болезни в амбулаторные лечебно-профилактические учреждения, численность взрослого населения различного возраста, число врачей в городе Б в 2000 году (абс. число).

Возраст, годы	Число обращений	Число посещений	Численность населения	Число врачей
18-29	102		51900	

30-49	148		48299	
50-69	220		159588	
70 и более	47		7600	
ИТОГ О:	517	3850	267387	1016

На основе приведенных данных в табл. 1 необходимо:

1. рассчитать интенсивные, экстенсивные показатели, показатели соотношения и наглядности,
2. выполнить графическое изображение интенсивных, экстенсивных величин, показателей соотношения и наглядности,
3. провести анализ полученных данных и сделать соответствующие выводы.

Образец выполнения задания

Вычисление показателей:

1. Интенсивные показатели.

Уровень обращаемости в амбулаторные ЛПУ по поводу мочекаменной болезни среди взрослого населения города Б в 2000 году и в различных возрастных группах

$$\begin{aligned} \text{Уровень обращаемости} \\ \text{всего взрослого населения} \end{aligned} = \frac{\text{Число всех обращений}}{\text{Численность населения города}} \times 100000 = \frac{517}{267387} \times 100000 = 193$$

$$\begin{aligned} \text{Уровень} \\ \text{обращаемости} \end{aligned} = \frac{\text{Число обращений лиц 18 – 29 лет}}{\text{Численность населения 18 – 29 лет}} \times 100000 = \frac{102}{51900} \times 100000 = 196$$

$$\begin{aligned} \text{Уровень} \\ \text{обращаемости} \end{aligned} = \frac{\text{Число обращений лиц 30 – 49 лет}}{\text{Численность населения 30 – 49 лет}} \times 100000 = \frac{148}{48299} \times 100000 = 306$$

$$\begin{aligned} \text{Уровень} \\ \text{обращаемости} \end{aligned} = \frac{\text{Число обращений лиц 50 – 69 лет}}{\text{Численность населения 50 – 69 лет}} \times 100000 = \frac{220}{159588} \times 100000 = 138$$

$$\begin{aligned} \text{Уровень} \\ \text{обращаемости} \end{aligned} = \frac{\text{Число обращений лиц 70 и более лет}}{\text{Численность населения 70 и более лет}} \times 100000 = \frac{47}{7600} \times 100000 = 618$$

Уровень посещаемости

$$\begin{aligned} \text{всего взрослого населения} &= \frac{\text{Число всех посещений}}{\text{Численность населения города}} \times 100\,000 = \frac{3850}{267387} \times 100\,000 = 1440 \\ \text{(на 100 000 человек)} & \end{aligned}$$

1. Экстенсивные показатели

$$\begin{aligned} \text{Доля обращений лиц} &= \frac{\text{Число обращений лиц 18–29 лет}}{\text{Число всех обращений}} \times 100\% = \frac{102}{517} \times 100\% = 19,73\% \\ \text{в возрасте 18-29 лет} & \end{aligned}$$

$$\begin{aligned} \text{Доля обращений лиц} &= \frac{\text{Число обращений лиц 30–49 лет}}{\text{Число всех обращений}} \times 100\% = \frac{148}{517} \times 100\% = 28,63\% \\ \text{в возрасте 30-49 лет} & \end{aligned}$$

$$\begin{aligned} \text{Доля обращений лиц} &= \frac{\text{Число обращений лиц 50–69 лет}}{\text{Число всех обращений}} \times 100\% = \frac{220}{517} \times 100\% = 42,55\% \\ \text{в возрасте 50-69 лет} & \end{aligned}$$

$$\begin{aligned} \text{Доля обращений лиц} &= \frac{\text{Число обращений лиц 70 и более лет}}{\text{Число всех обращений}} \times 100\% = \frac{47}{517} \times 100\% = 9,09\% \\ \text{в возрасте 70 и более} & \end{aligned}$$

3. Показатель соотношения

$$\begin{aligned} \text{Число врачей} &= \frac{\text{Число врачей}}{\text{Число населения города Б}} \times 10\,000 = \frac{1016}{267387} \times 10\,000 = 37,99 \text{ (38)} \\ \text{на 10 000 населения} & \end{aligned}$$

3. Показатель соотношения

$$\begin{aligned} \text{Число врачей} &= \frac{\text{Число врачей}}{\text{Число населения города Б}} \times 10\,000 = \frac{1016}{267387} \times 10\,000 = 37,99 \text{ (38)} \\ \text{на 10 000 населения} & \end{aligned}$$

4. Показатели наглядности (в процентах по отношению к интенсивному показателю – обращаемости в возрасте 50-69 лет, принятому за 100%).

Таблица 2. Показатели наглядности.

Показатели	Возраст, годы			
	18-29	30-49	50-69	70 и более
Интенсивный (на 100000 населения)	196	306	138	618
Наглядности (в % к	142	22	10	448

возрасту 50-69 лет)		2	0	
---------------------	--	---	---	--

Полученные показатели сводим в таблицу.

Таблица 3. Уровень обращаемости, посещаемости по поводу мочекаменной болезни в амбулаторные ЛПУ и обеспеченность врачами взрослого населения города Б в 2000 году.

Возраст, годы	Уровень обращаемости на 10000 взрослых	Структура обращаемости, %	Показатели наглядности, % (повозрастной обращаемости)
18-29	196	19,73	142
30-49	306	28,63	222
50-69	138	42,55	100
70 и более	618	9,09	448
ИТОГО:	193	100,0	-

Уровень посещаемости – 1440

Обеспеченность взрослого населения врачами – 38

Графическое изображение показателей

1. Интенсивные показатели

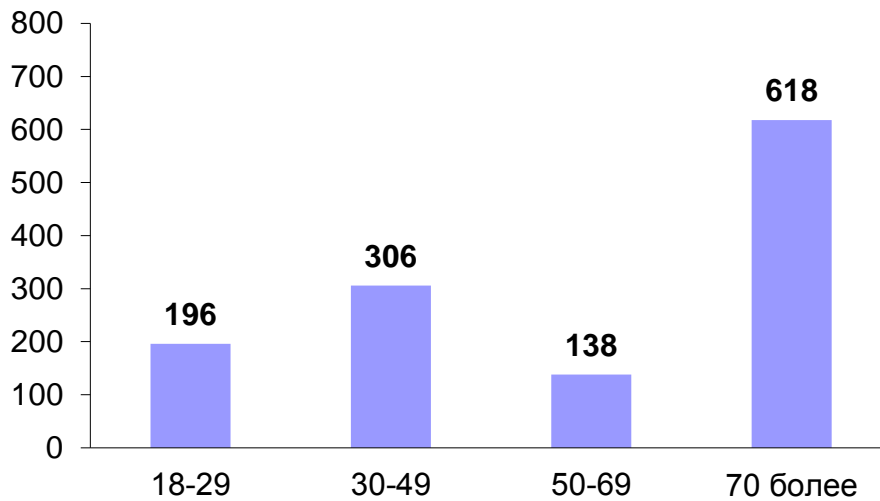


Рис. 9. Уровень обращаемости в амбулаторные ЛПУ взрослого населения города Б в 2000 году по поводу мочекаменной болезни (на 1000 000 населения).

По оси абсцисс - возрастные группы, по оси ординат - уровень обращаемости.

2. Экстенсивные показатели

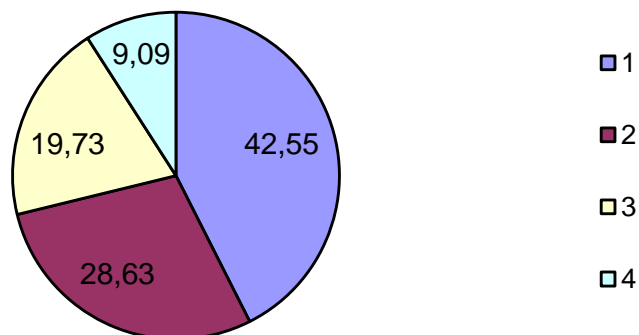


Рис. 10. Структура обращаемости взрослого населения (в %).

1-лица в возрасте 50-69 лет, 2- лица в возрасте 30-49 лет,
3-лица в возрасте 18-29 лет, 4-лица в возрасте 70 и более лет.

3. Показатели соотношения

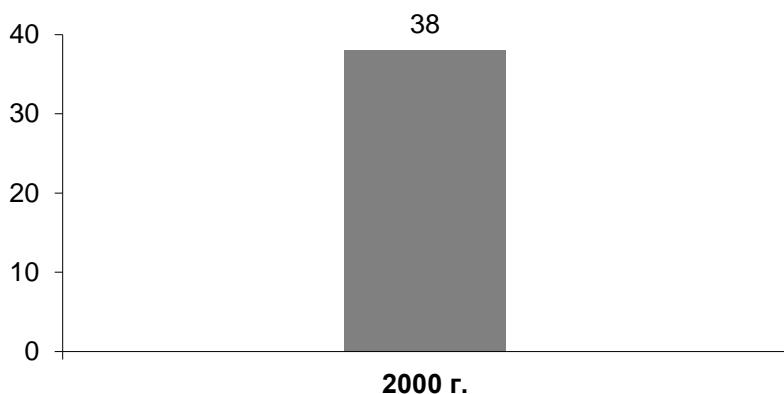
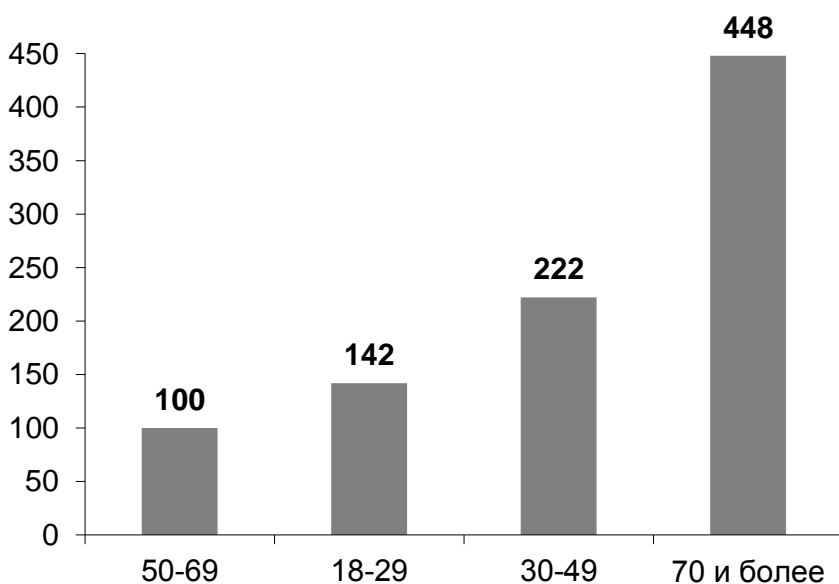


Рис. 11. Обеспеченность взрослого населения города Б врачами в 2000 году (на 10000).

По оси абсцисс - изучаемый год, по оси ординат - показатель обеспеченности

4. Показатели наглядности



5.

Рис. 12. Показатели наглядности обращаемости взрослого населения по поводу мочека-

менной болезни в амбулаторные ЛПУ города Б в 2000 году (в % к возрастной группе

50-69 лет).

По оси абсцисс – возрастные группы, по оси ординат – показатели.

Анализ полученных результатов и выводы

Анализируя полученные результаты можно сделать следующие выводы:

1. Наиболее высокий уровень обращаемости взрослого населения в городе Б в 2000 году по поводу мочекаменной болезни приходится на возрастную группу 70 лет и более, несмотря на то, что удельный вес лиц данного возраста в городе Б минимальный.

2. Самый низкий уровень обращаемости характерен для возрастной группы 50-69 лет, доля которых в структуре населения максимальна.

3. Указанные различия в уровне обращаемости в амбулаторные ЛПУ по поводу мочекаменной болезни подтверждаются значениями показателей наглядности. Показатель наглядности имеет максимальную величину в возрасте 70 и более лет, а минимум – в возрастной категории – 50-69 лет.

4. Обеспеченность врачами взрослого населения города Б в 2000 году высокая и превышает аналогичный показатель по стране.

5. В структуре обратившегося за медицинской помощью взрослого населения преобладают лица в возрасте 50-69 лет.

При наличии соответствующих данных возможно сделать сравнение анализируемых показателей в динамике, с нормативами, принятыми в стране и т. д.

ВАРИАНТ 1.

Таблица 4. Число впервые выявленных случаев сифилиса в кожно-венерологическом диспансере, число посещений, врачей и медсестер и численность обслуживаемого населения (в асб. ч.).

Годы исследования	Численность населения	Число впервые зарегистрированных случаев сифилиса	Число посещений	Число врачей	Число медсестер
1998	300000	750			
1999	298000	730			
2000	299000	810			
2001	296000	850			
ИТОГО:	-	3140	15700	200	600

ВАРИАНТ 2.

Таблица 5. Число впервые выявленных заболеваний гонореей в различных социальных группах, число специализированных коек и врачей дерматовенерологов (абс. ч.).

Группы населения	Численность населения	Число впервые выявленных заболеваний гонореи	Число специализированных коек	Число врачей дерматовенерологов
Рабочие	200000	300		
Служащие	110000	120		
Бизнесмены	400000	250		
Прочие	150000	80		
ИТОГО:	500000	750	120	20

ВАРИАНТ 3.

Таблица 6. Число обращений и посещений в поликлинику, число врачей и численность обслуживаемого населения (абс. ч.).

Возрастные группы, годы	Численность населения	Количество всех обращений в поликлинику	Количество всех посещений в поликлинику	Число врачей
18-29	25000	18000		
30-49	31000	45000		
50-69	35000	51000		
70 и более	9000	18000		
ИТОГО:	100000	132000	1320000	290

ВАРИАНТ 4.

Таблица 7. Число обращений детей в поликлинику по поводу ОРВИ, число педиатров и численность детей в зоне деятельности поликлиники (абс. ч.)

Возраст, годы	Численность детей	Число обращений по поводу ОРВИ	Число педиатров
0-2	3000	9000	
3-5	4000	8000	
6-7	8000	4000	
8-17	25000	15000	
ИТОГО:	40000	36000	50

ВАРИАНТ 5.

Таблица 8. Число обращений детей по поводу травм в травмпункт, число госпитализаций, педиатрических участков и численность обслуживаемого населения (абс. ч.).

Возраст, годы	Численность детей	Число обращений по поводу травм	Число госпитализаций по поводу травм	Число педиатрических участков
0-2	4000	136		
3-5	6000	342		
6-7	12000	1200		
8-17	28000	6322		
ИТОГО:	50000	8000	750	62

ВАРИАНТ 6.

Таблица 9. Число прошедших профосмотры в МСЧ, выявленных больных ИБС, число врачей МСЧ и работающих на ГОКе (абс. ч.)

Профессиональные группы	Число прошедших профилактические осмотры	Число выявленных больных ИБС	Число врачей в МСЧ	Средняя численность работающих на ГОКе

Экскаваторщики	4000	380		
Водители	2000	250		
Горнорабочие	11000	2130		
Управленческий персонал	1000	240		
ИТОГО:	18000	3000	60	18000

ВАРИАНТ 7.

Таблица 10. Число прошедших профосмотры, выявленных больных гипертензией, врачей в МСЧ и работающих на металлургическом комбинате (абс. ч.).

Изучаемые годы	Число прошедших профосмотры	Число выявленных больных гипертензией	Число врачей в МСЧ	Среднегодовая численность работающих на металлургическом комбинате
1998	9200	301		
1999	9000	440		
2000	8800	480		
2001	9000	279		
ИТОГО:	9000	1500	30	9000

ВАРИАНТ 8.

Таблица 11. Число умерших от всех болезней, число врачей и коек в Курске, Орле и Белгороде в 2001 году (абс. ч.).

Изучаемые города	Численность населения	Число умерших человек	Число врачей	Число коек

Курск	450000	7300	1575	3600
Белгород	350000	4200	1420	3500
Орел	330000	4290	1386	3960
ИТОГО:	1130000	15790	4381	11060

ВАРИАНТ 9.

Таблица 12. Число умерших детей на 1-ом году жизни, число родившихся живыми, число педиатров и детей всех возрастов в Курске, Орле, Белгороде (абс. ч.).

Исследуемые города	Численность умерших на 1-ом году жизни	Число родившихся живыми	Число педиатров	Число всех детей
Курск	81	4050	110	110000
Орел	49	3500	115	90000
Белгород	49	3300	100	80000
ИТОГО:	179	10850	325	280000

ВАРИАНТ 10.

Таблица 13. Число посещений в поликлинику по поводу различных заболеваний, численность населения и число врачей в 2001 году (абс. ч.).

Название заболевания	Число посещений в поликлинику	Численность населения	Число врачей
Гипертоническая болезнь	2000		
ИБС	5000		
Травмы	500		
Пневмония	200		
Прочие	2300		
ИТОГО:	10000	120000	48

Контрольные вопросы:

1. Что такое относительные величины?
2. Как рассчитываются относительные величины?
3. Какие различают виды относительных величин?
4. В каких случаях применяются интенсивные показатели?
5. Что такое радиальная диаграмма?
6. Что такое экстенсивный показатель, или показатель структуры распределения?
7. Расскажите о методике расчета экстенсивного показателя.