

***Основы  
био-  
статистики***

**СТАТИСТИКА – самостоятельная общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной.**

Медицинская статистика является отраслевой статистикой, которая изучает вопросы, связанные с медициной, как областью науки и практической деятельности, направленной на сохранение и укрепление здоровья людей, предупреждение и лечение болезней.

**Методы сбора информации, ее изучения, обработки и анализа, имеющиеся в арсенале медицинской статистики используются всеми клиническими и теоретическими дисциплинами нашей отрасли.**

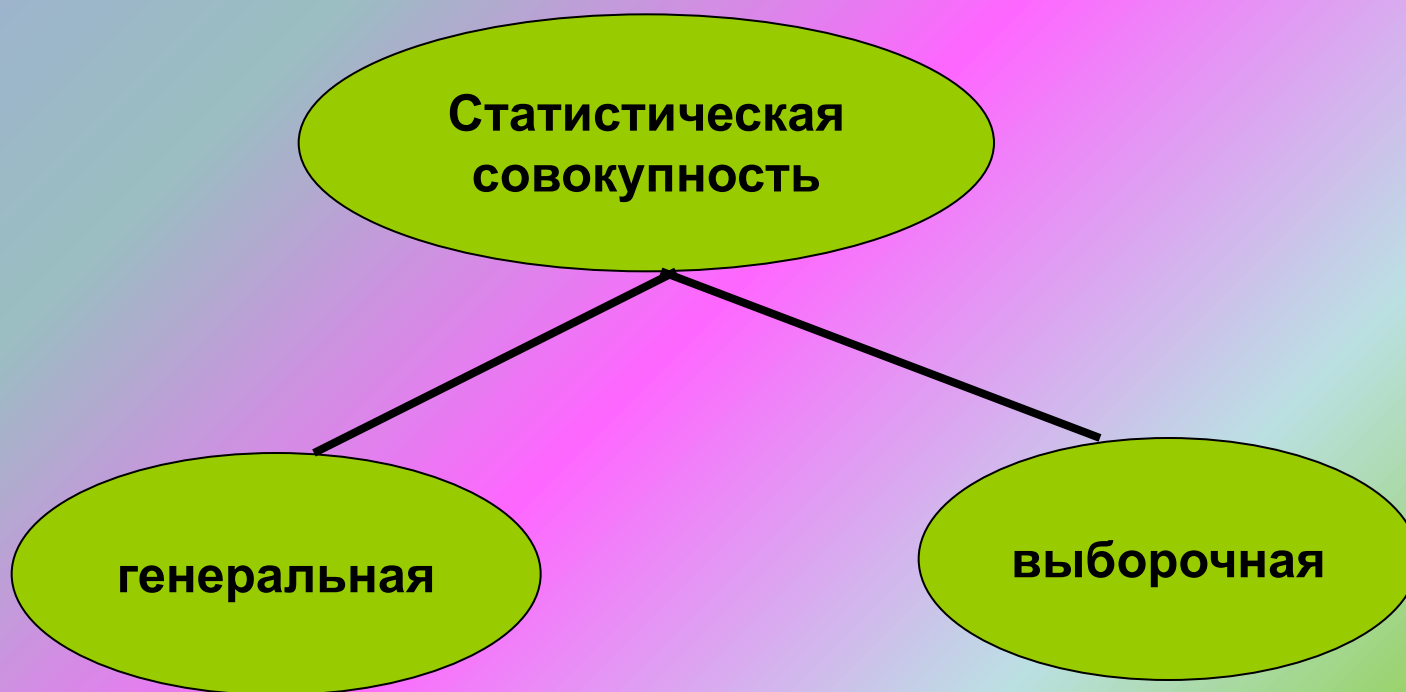
# **Основные разделы медицинской статистики:**

- Статистика здоровья населения,**
- Статистика здравоохранения.**

### **Статистическая совокупность –**

**группа, состоящая из определенного числа относительно однородных элементов (единиц наблюдения), взятых вместе в конкретных границах времени и пространства.**

# Статистика здоровья населения



**Генеральная совокупность –**

**совокупность всех единиц наблюдения в соответствии с целью исследования.**

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



## Статистика здоровья населения

**Выборочная совокупность должна отражать структуру и свойства генеральной, т.е. быть по отношению к ней представительной (репрезентативной).**

# Статистика здоровья населения

Формула определения  
минимального объема  
выборочной совокупности

$$n = \frac{t^2 \times pq}{\Delta^2}$$

Пример:

$$n = \frac{2^2 \times 50,0 \times 50,0}{10^2} = 100$$

**n** — требуемое число наблюдений,  
**t** — критерий достоверности,  
**p** — показатель распространенности,  
**q = 1 - p** или **(100 - p)**,  
**Δ** — предельная ошибка.  
**Δ = t × m**, где  
m — ошибка показателя

# Статистика здоровья населения

## Бесповторный отбор единиц наблюдения:

1. Случайная выборка:
  - метод жеребьевки,
  - метод случайных чисел
2. Механическая выборка
3. Типическая (типологическая) выборка
4. Серийная выборка

## Фрагмент таблицы случайных чисел:

22268	33438	80072	15542	48690
89788	61740	98666	56241	65233
33302	48110	01135	79509	74652
18924	03164	51354	41271	70338
91413	49689	65814	52444	83068
83024	15196	33225	53066	15830
41487	92226	99027	97751	54987
46768	34578	67195	13303	39784
05088	91708	18073	59685	50768
58476	78094	72652	38655	74018
26795	67348	77768	79736	08602

# Статистика здоровья населения

Точность исследования - приближение, с которым получаются подлинные значения показателей.

---

Надежность исследования определяется вероятностью ошибки полученного показателя.

Ошибка показателя

$$m = \sqrt{\frac{P \times q}{n}} \quad P \pm m$$

Достоверность показателя

$$t = \frac{P}{m} \geq 2,0 \quad (95,0\%)$$

$P \geq 0,95$  (95% достоверности)

$p \leq 0,05$  (вероятность ошибки)

# Статистика здоровья населения

Виды относительных показателей:

## 1. Интенсивный (частоты)

$$\text{Смертность} = \frac{\text{Число умерших за год}}{\text{Среднегодовая численность населения}} \times 1000$$

## 2. Экстенсивный:

- альтернативный,

Отношение числа случаев, имеющих признак к числу всех случаев

$$P = \frac{\text{Число больных}}{\text{число здоровых} + \text{число больных}} \times 1000$$

# Статистика здоровья населения

- распределительный

## Отношение между частями и целым

### *Структура заболеваний (в %)*

Гипертоническая болезнь	- 30,0
Ишемич. болезнь сердца	- 20,0
Легочное сердце	- 5,0
Другие болезни сердца	- 7,0
Цереброваскулярные болезни	- 15,0
Б-ни артерий, артериол, капил.	- 5,0
Болезни вен и лимф. сосудов	- 3,0
Др. б-ни системы кровообращения	- 15,0
<b>ВСЕГО.....</b>	<b>100,0</b>

# Статистика здоровья населения

- **указательный**

$$k = \frac{\text{Уровень обращаемости по ранее извест. заб.}}{\text{Уровень первичной заболеваемости}}$$

## Отношение между частями целого

$$\begin{array}{l} \text{Гипертоническая болезнь} \quad 30,0 \\ \text{-----} = \text{-----} = 2,0 \\ \text{Цереброваскулярные болезни} \quad 15,0 \end{array}$$

3. **Соотношения**

## Соотношение двух независимых совокупностей

$$P = \frac{\text{Количество коек}}{\text{Численность населения}} \times 10\,000$$

Характеристика коэффициентов степени хронизации заболеваний по классам болезней

Классы болезней	Заболевания (на 1000)		$K_{xp} = \frac{P_{xp}}{P_{п}}$	Ранг
	Впервые в жизни $P_{п}$	Ранее известные $P_{xp}$		
<b>Всего</b>	<b>466,0</b>	<b>529,0</b>	<b>1,13</b>	
Инфекционные болезни	27,0	18,8	0,67	10
Новообразования	8,4	23,4	2,79	5
Б-ни эндокринной системы	7,0	30,8	4,40	4
Б-ни крови и кроветв. орган.	0,5	2,5	5,00	2
Психические расстройства	21,4	40,3	1,88	7
Б-ни нервной сист. и орг. чувств	52,2	76,1	1,46	8
<b>Б-ни системы кровообращения</b>	<b>15,4</b>	<b>103,0</b>	<b>6,69</b>	<b>1</b>
Б-ни органов дыхания	111,4	60,7	0,54	11
Б-ни органов пищеварения	11,2	51,5	4,60	3
Б-ни мочеполовой системы	35,0	35,6	1,02	9
Б-ни кожи и подкожной клетч.	41,9	5,9	0,14	12
Б-ни костно -мышечной системы	26,6	74,0	2,78	6
Симптомы, признаки и неточно обозначенные состояния	0,9	1,1	1,22	9
Травмы и отравления	107,1	5,3	0,05	13



# Статистика здоровья населения

## Оценка достоверности различий показателей

**Задача:** Провести сравнительный анализ ЗВУТ по двум совхозам (Солонцы и Красноярский)

**С/з Красноярский:** Из 1000 работающих случаев забол. с ВУТ за год было 86.

На 100 работающих -  $(86 \times 100) : 1000 = 8,6$   
 $m = \pm 0,9$

**С/з Солонцы:** Из 700 работающих случаев забол. с ВУТ за год было 92.

На 100 работающих -  $(92 \times 100) : 700 = 13,2$   
 $m = \pm 1,3$

$$t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \geq 2 \text{ (95,5\%)}$$

$$m = \sqrt{\frac{Pq}{n}}$$

$$t = \frac{13,2 - 8,6}{\sqrt{1,3^2 + 0,9^2}} = \frac{4,6}{1,6} = 2,9$$

**Вывод:** ЗВУТ в с/з Солонцы существенно (достоверно) выше, чем в с/з Красноярский.

**Вопрос:** Чем это обусловлено?

Стаж работы	С/з КРАСНОЯРСКИЙ				С/з СОЛОНЦЫ			
	Число работающих	%	Из них заболело	На 100 раб.	Число работающих	%	Из них заболело	На 100 раб.
До 5 л.	300	30,0	18	6,0	50	7,1	2	4,0
6- 10 л.	500	50,0	40	8,0	100	14,3	6	6,0
11-15 л.	150	15,0	16	11,0	300	42,8	39	13,0
>15 л.	50	5,0	12	25,0	250	35,8	45	18,0
<b>Всего</b>	<b>1000</b>	<b>100</b>	<b>86</b>	<b>8,6</b>	<b>700</b>	<b>100</b>	<b>92</b>	<b>13,2</b>

$$P_{\text{сол.}} = \frac{4,0 \times 7,1 + 6,0 \times 14,3 + 13,0 \times 42,8 + 18,0 \times 35,8}{100} = 13,2$$

$$P_{\text{станд.}} = \frac{4,0 \times 30,0 + 6,0 \times 50,0 + 13,0 \times 15,0 + 18,0 \times 5,0}{100} = 7,05$$

## Шкала интегралов вероятностей

	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0080	0160	0239	0319	0399	0478	0558	0638	071
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2050	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6578	6629	6680	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9666	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972

## *ЗАКОН БОЛЬШИХ ЧИСЕЛ*

Частота (выраженная в долях)  $m/k$ -го события  $X$  будет сколь угодно близкой к его вероятности  $p$ , если число испытаний неограниченно возрастает.

## *ЗАКОН РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН*

Функция  $f(x)$ , связывающая значения  $X_i$  переменной случайной величины  $\tilde{X}$  с их вероятностью "  $P$  ", называется законом распределения этой случайной величины.

## *ЗАКОН НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ.*

Вероятность отклонения любой варианты  $X_i$  от центра распределения ( $\mu$ ), является функцией нормированного отклонения ( $t$ ).

- накапливание вариантов в центральных классах и постепенное убывание их численности по мере удаления от центра ряда -

**В МАССЕ ОДНОРОДНЫХ ЕДИНИЦ ОДНОЙ СТАТИСТИЧЕСКОЙ СОВОКУПНОСТИ БОЛЬШАЯ ЧАСТЬ ВАРИАНТ ОКАЗЫВАЕТСЯ СРЕДНЕЙ ВЕЛИЧИНЫ ИЛИ БЛИЗКИ К НЕЙ.**

*ЗАКОН БОЛЬШИХ ЧИСЕЛ*

**ЧАСТОСТЬ  $m/n$  СОБЫТИЯ А БУДЕТ СКОЛЬ УГОДНО БЛИЗКОЙ К ЕГО ВЕРОЯТНОСТИ, ЕСЛИ ЧИСЛО ИСПЫТАНИЙ НЕОГРАНИЧЕННО ВОЗРАСТАЕТ.**

ПРИМЕР:

В ЕМКОСТИ НАХОДИТСЯ **5** БЕЛЫХ И **15** ЧЕРНЫХ ШАРОВ  
– ВСЕГО **20**.

КАКОВА ВЕРОЯТНОСТЬ ТОГО, ЧТО ПЕРВЫЙ ЖЕ  
ВЫНУТЫЙ ШАР ОКАЖЕТСЯ БЕЛЫМ?

$$m = 5$$

$$k = 20$$

$$P = 5/20 = 0,25$$

$$P_1 \dots P_2 \dots P_3 \dots P_k$$

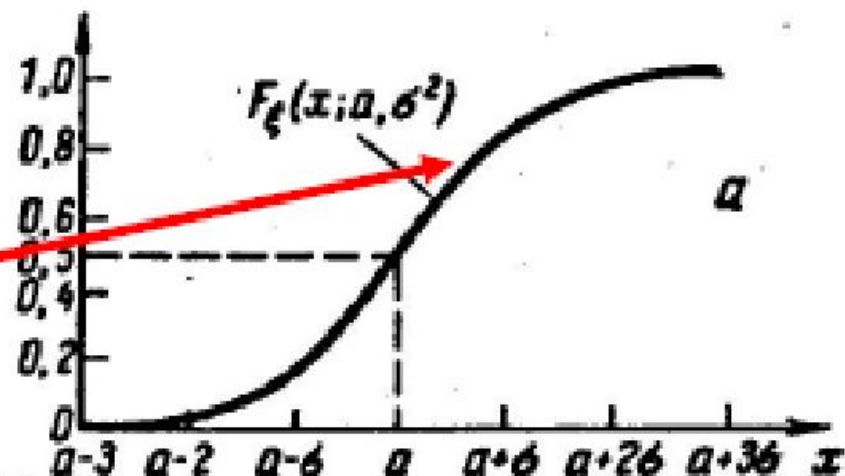
**«...внутренний закон, регулирующий случайные явления, становится видимым лишь тогда, когда они охватываются в больших массах...»**

*К.Маркс*

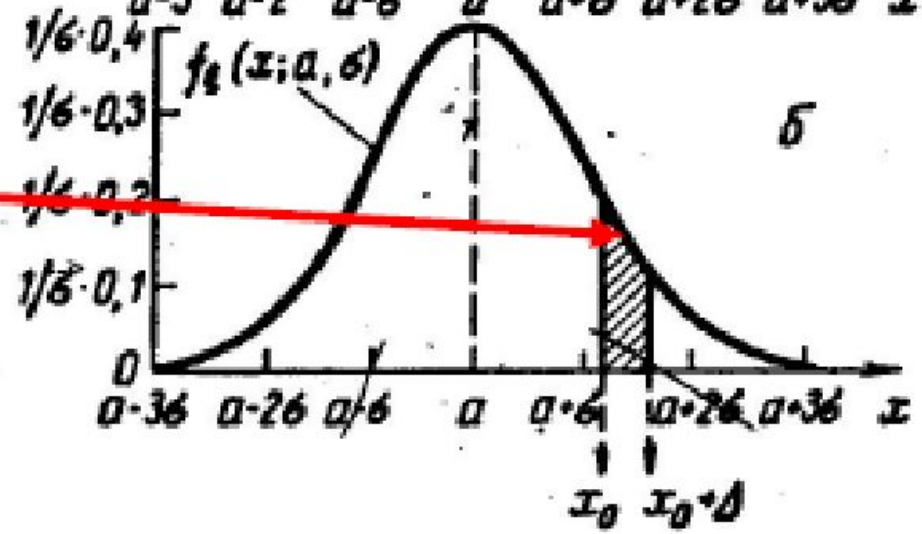
*ЗАКОН РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН*

**Функция  $f(X)$ , связывающая значения  $X_i$  переменной случайной величины  $X$  с их вероятностями  $P_i$ , называется законом распределения этой величины.**

кумулята



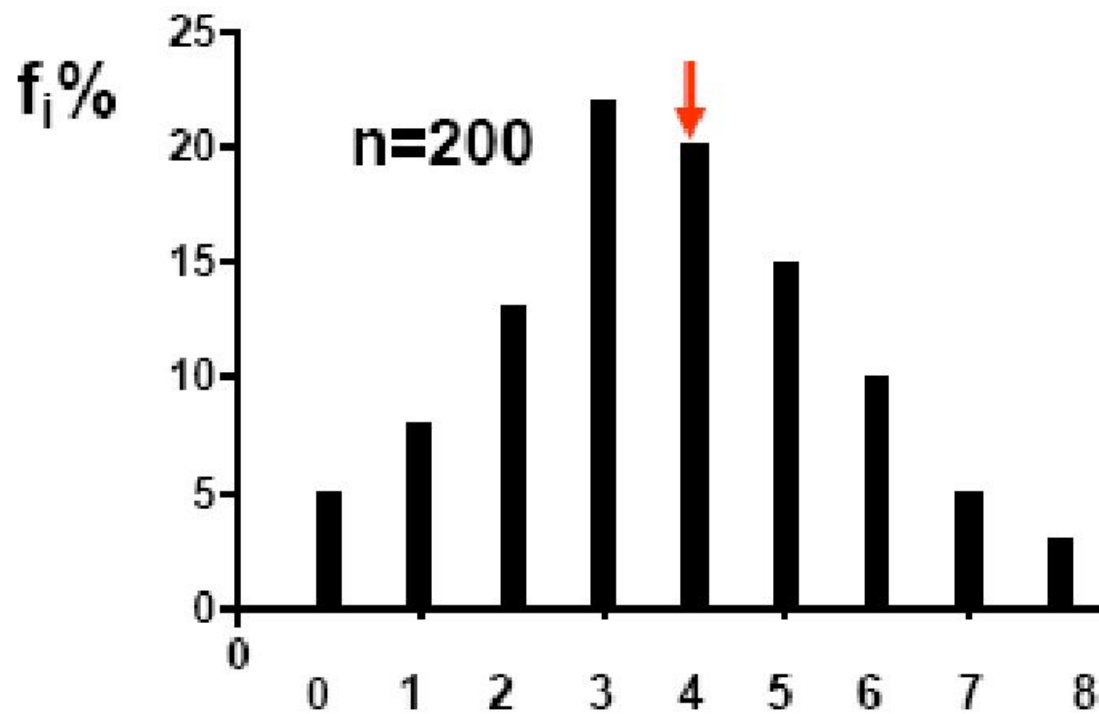
$$\frac{F_i}{n} = P_i$$



$$F_{\varepsilon(X)} = \frac{\lim_{\Delta \rightarrow 0} F_{(X+\Delta)} - F_X}{\Delta} \rightarrow P_X$$



## Дискретный аналог графика плотности распределения



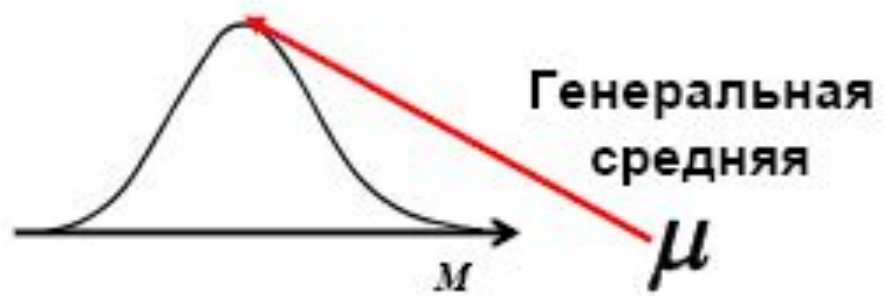
$X$  – число особей в рамке

## МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

$$\sum_{i=1}^k P_i = 1$$

$$\mu = \sum_{i=1}^m P_i X_i$$

$$\bar{X} = \frac{\sum f_i X_i}{n} \rightarrow P_i$$



$n \rightarrow \infty$

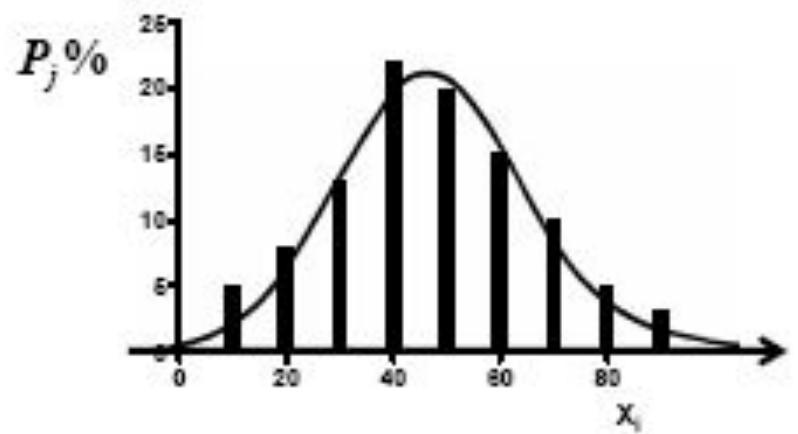
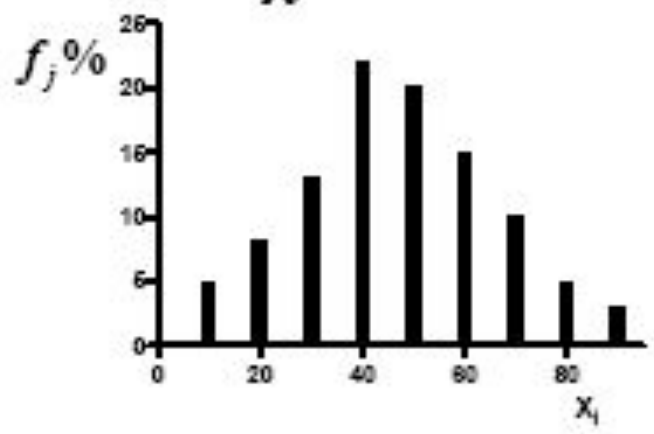
$n \rightarrow \infty$

Математическое ожидание

$$M = \frac{\sum_{i=1}^n X_i}{n}$$

$n \rightarrow \infty$

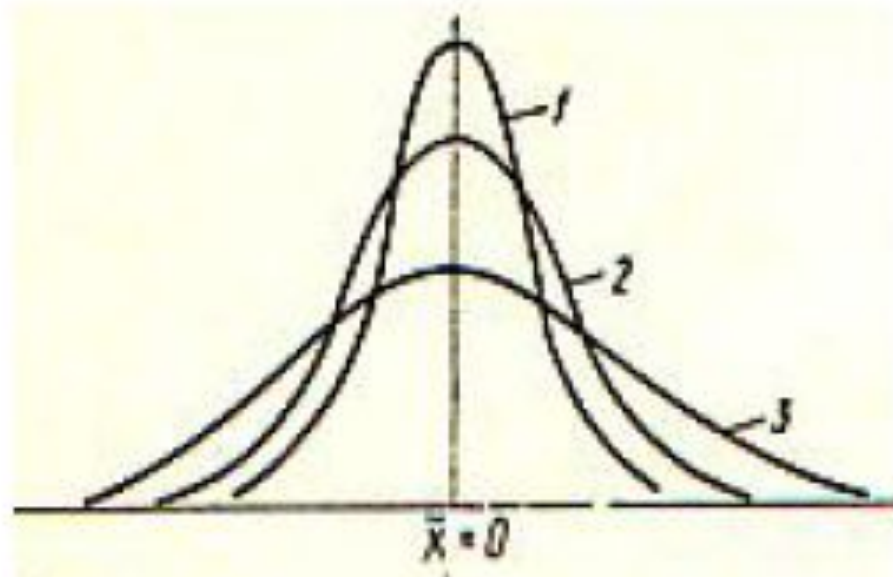
$$\mu = \sum_{j=1}^k P_j$$

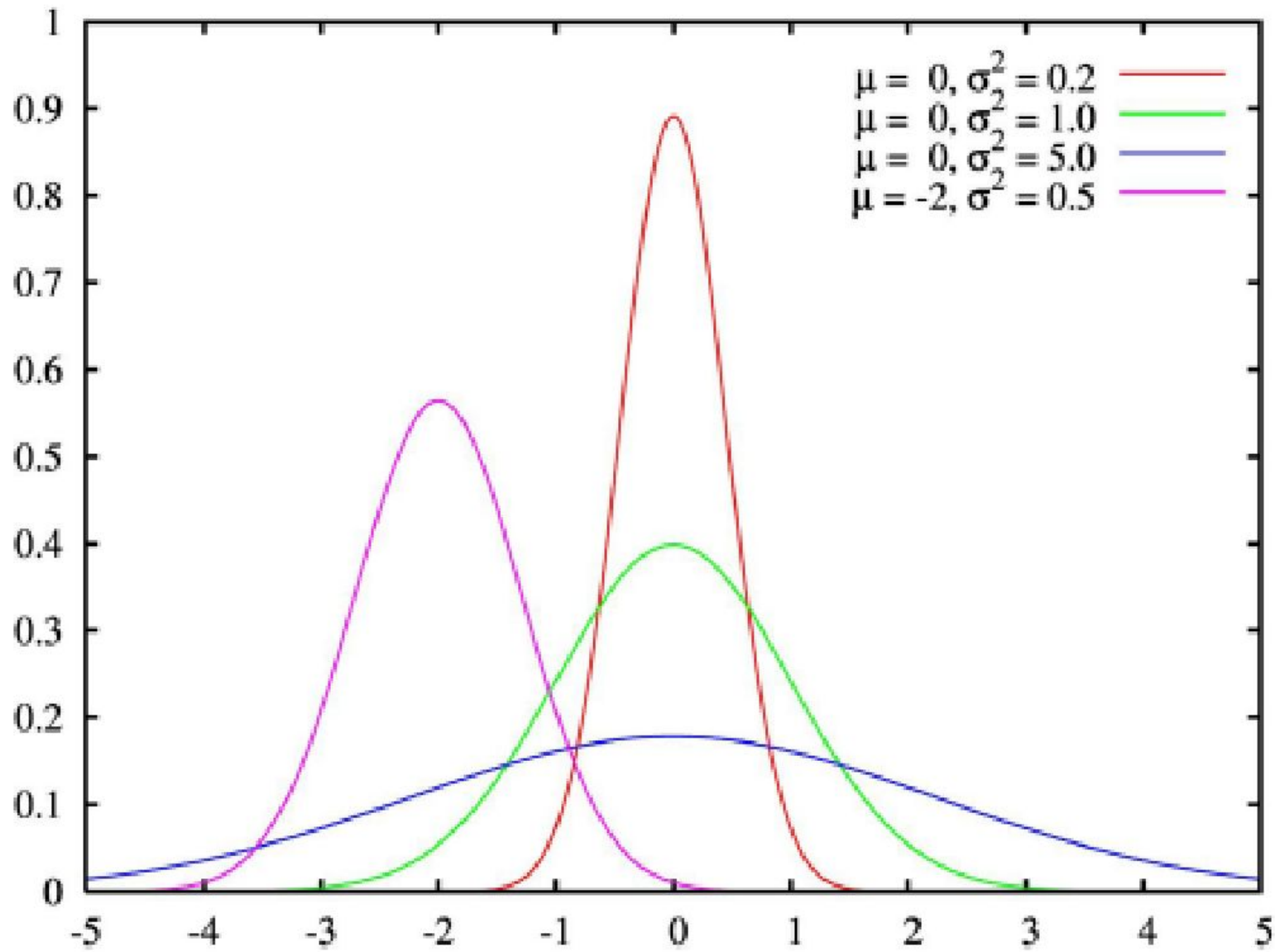


## СВОЙСТВА НОРМАЛЬНОЙ КРИВОЙ

1. - ПРИ ИЗМЕНЕНИИ СРЕДНЕЙ  
ФОРМА КРИВОЙ НЕ МЕНЯЕТСЯ, ОНА  
ПРОСТО СМЕЩАЕТСЯ ПО ОСИ  
АБСЦИСС

2. - ПРИ ИЗМЕНЕНИИ  
СРЕДНЕГО  
КВАДРАТИЧЕСКОГО  
ОТКЛОНЕНИЯ  
МЕНЯЕТСЯ ШИРИНА  
КРИВОЙ





1. ЛЮБУЮ НОРМАЛЬНУЮ КРИВУЮ МОЖНО ПРИВЕСТИ К СТАНДАРТНОЙ ФОРМЕ

$$X_i' = \frac{X_i - M}{s}$$

СТАНДАРТИЗИРОВАННАЯ НОРМАЛЬНАЯ КРИВАЯ

$\mu=0$  и  $\sigma=1$

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

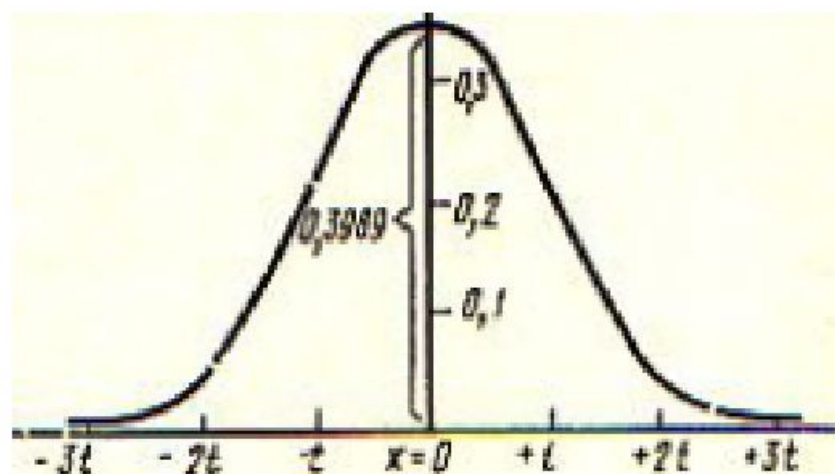


Рис. 11. Стандартизованная форма нормальной кривой (при  $\sigma=1$ )

2. СТАНДАРТНАЯ КРИВАЯ ИМЕЕТ ПЛОЩАДЬ РАВНУЮ 1

$$P ( -t < | X_i - M | < +t ) = 0.6827$$

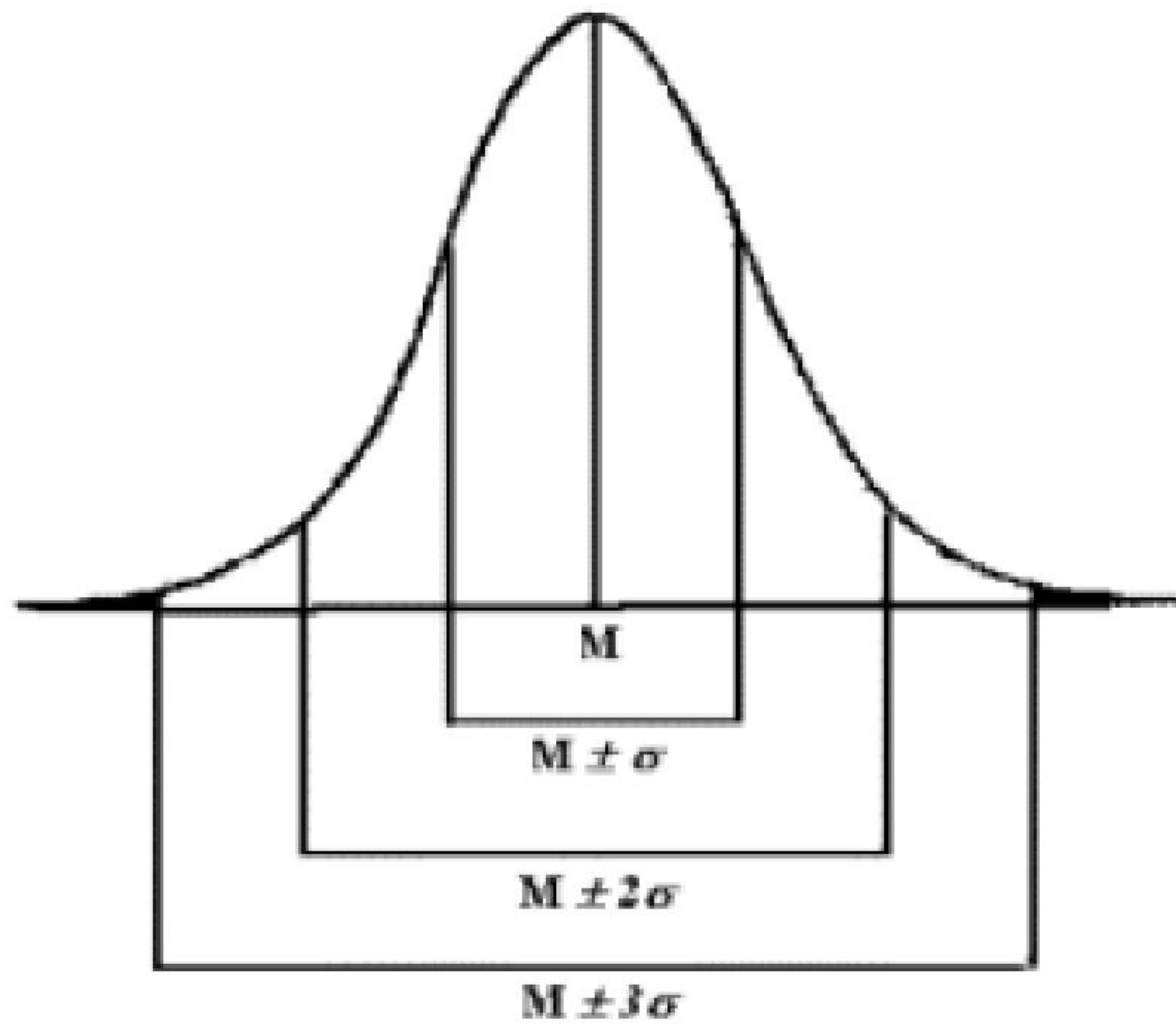
$$P ( -2t < | X_i - M | < +2t ) = 0.9545$$

$$P ( -3t < | X_i - M | < +3t ) = 0.9973$$

ПРАВИЛО ТРЕХ СИГМ:

**99,7%** всех вариантов нормально распределенной совокупности находится в пределах

$$M \pm 3s$$





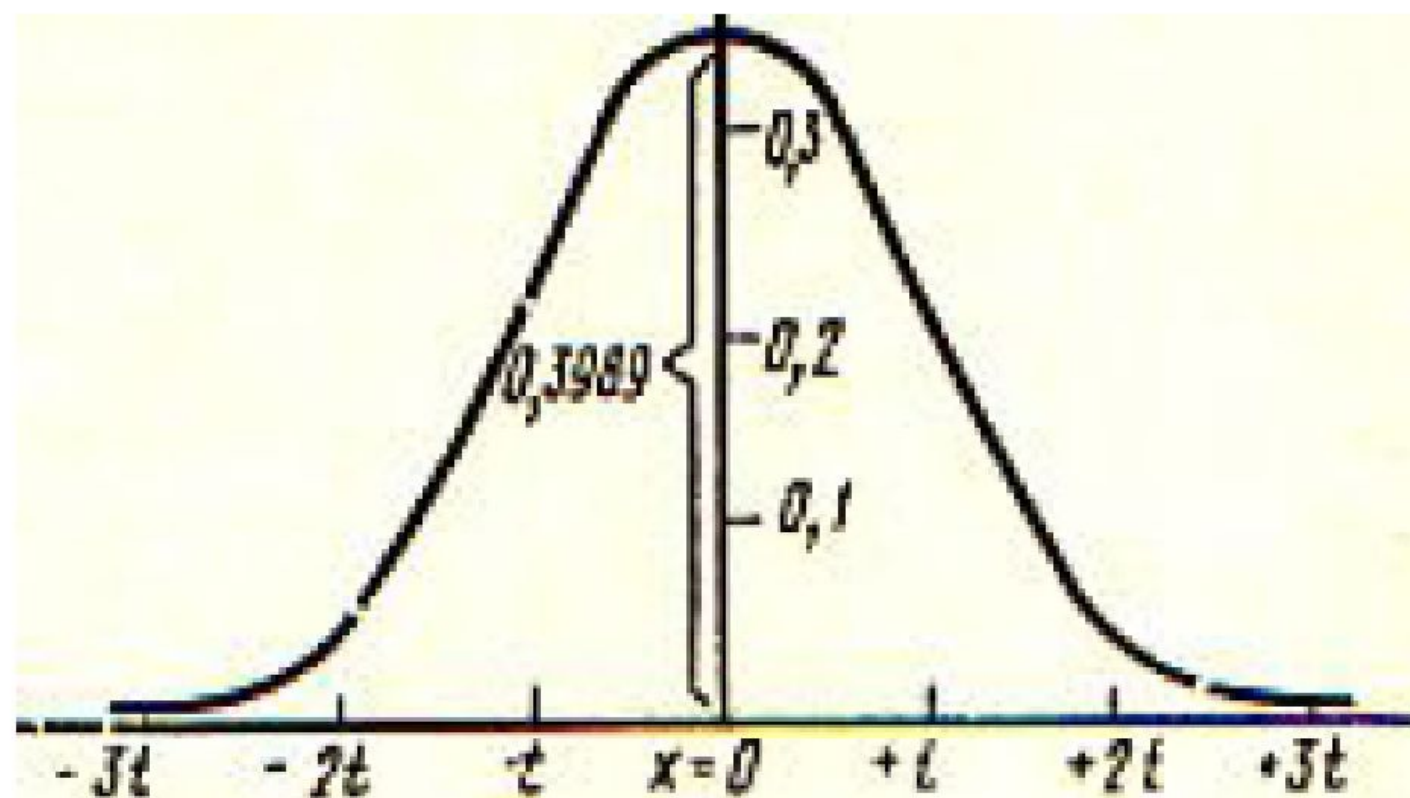


Рис. 11. Стандартизованная форма нормальной кривой (при  $\sigma=1$ )

# Статистика здоровья населения

## Области применения средних величин:

- в клинических и лабораторных исследованиях,
- при оценке физического развития населения,
- при эпидемиологических и гигиенических исследованиях,
- при оценке деятельности медицинских учреждений.

# Повторение изученного

Найди среднее  
арифметическое чисел:

А) 4,4,7

Б) 10,5,5,8

В) 250,50,100,100,500

# Средние величины

1. 26,23,18,25,20,25,30,25,34,19.
2.  $(26+23+18+25+20+25+30+25+34+19):$   
 $10 = 245:10 = 24,5$   
24,5 – среднее арифметическое  
18 – наименьшее время  
34 – наибольшее время

Число, наиболее часто встречающееся в данном ряду, называется модой ряда чисел.

25 – мода ряда чисел.

Ряд может иметь две моды, а может не иметь моды. Например,

47, 46, 50, 52, 47, 49, 52, 55 – имеет две моды:  
47 и 52

59, 68, 66, 70, 67, 71, 74 – этот ряд не имеет моды.

# Способы расчета средней арифметической

1. Простая -  $M = \frac{\sum v}{n}$

2. Взвешенная -  $M = \frac{\sum vp}{n}$

3. По способу моментов -  $M = A + i \frac{\sum dp}{n}$

# Основные критерии разнообразия признаков

**Лимит** – это значения крайних вариантов в вариационном ряду

$$lim = V_{min} \div V_{max}$$

**Амплитуда** – это разница крайних вариантов вариационного а

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

**Среднее квадратическое отклонение** – характеризует рассеивание вариантов вокруг средней величины (внутреннюю структуру совокупности).

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}} \quad \text{среднеарифметический способ}$$

$$\sigma = i \sqrt{\frac{\sum d^2 p}{n} - \left(\frac{\sum dp}{n}\right)^2} \quad \text{способ моментов}$$

# Разнообразиие

26,23,18,25,20,25,30,25,34,19.

Разница между наибольшим и наименьшим числом называется размахом ряда чисел (амплитудой).

$34 - 18 = 16$ , это амплитуда ряда чисел.



# Оценка достоверности результатов исследования

Погрешность средней величины:

$$mM = \frac{\sigma}{\sqrt{n}}, \text{ если } n > 30;$$

$$mM = \frac{\sigma}{\sqrt{n-1}}, \text{ если } n \leq 30.$$

Погрешность относительной величины:

$$m\% \text{ /или } m_p/ = \sqrt{\frac{pq}{n}}, \text{ если } n > 30;$$

$$m\% \text{ /или } m_p/ = \sqrt{\frac{pq}{n-1}}, \text{ если } n \leq 30$$

# Оценка достоверности разницы относительных и средних величин

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} \text{ для средних величин;}$$

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

Разница достоверна при  $t = 2$  и больше, что отвечает вероятности безошибочного прогноза  $P = 95,5 \%$  и больше.

- **Динамический ряд** - это ряд однородных статистических величин, которые показывают изменение какого-то явления во времени
- **Простой динамический ряд** - выраженный абсолютными величинами.
- **Сложный динамический ряд** - выраженный относительными и средними величинами

# Показатели динамического ряда

1. Абсолютный прирост - разница между настоящим и предыдущим уровнями ряда
2. Темп роста - отношение настоящего уровня к предыдущему уровню (в %)
3. Темп прироста - отношение абсолютного прироста к предыдущему уровню (в %)
4. Показатель наглядности

# Методы стандартизации показателей

1. Прямой
2. Непрямой
3. Обратный

Используются в случаях, когда нужно сравнить явление, которое изучается в неоднородной совокупности

# Этапы прямого метода стандартизации

1. Вычисление интенсивных показателей
2. Выбор и вычисление стандарта
3. Вычисление “ожидаемых” величин
4. Вычисление стандартизированных показателей
5. Сравнение простых интенсивных и стандартизированных показателей

# Корреляционный анализ

применяется для определения  
взаимосвязи между явлениями

Корреляционная связь бывает:

- по характеру - прямой и обратной;
- по силе: сильной, средней, слабой

# Способы определения коэффициента корреляции

## 1. Метод рангов (Спирмена)

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

## 2. Метод квадратов (Пирсона)

$$r_{xy} = \frac{\sum |d_x \times d_y|}{\sqrt{\sum d_x^2 + \sum d_y^2}}$$



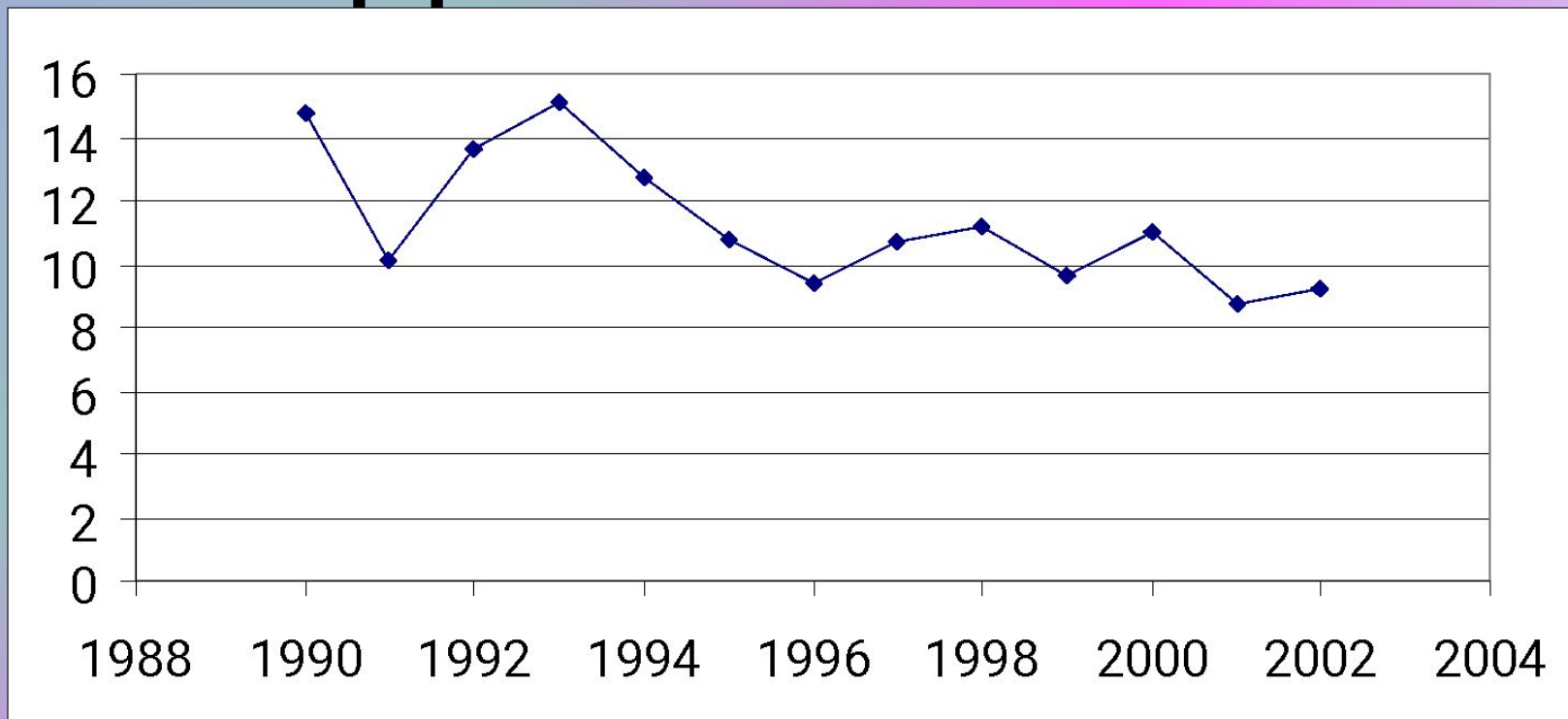
# Оценка силы и характера корреляционной связи

Сила связи	Прямая (+)	Обратная (-)
Полная	+1	
Сильная	От +1 до +0,7	От -1 до -0,7
Средняя	От +0,7 до +0,3	От -0,7 до -0,3
Слабая	От +0,3 до 0	От -0,3 до 0
Связь отсутствует	0	0

# *Виды графических изображений*

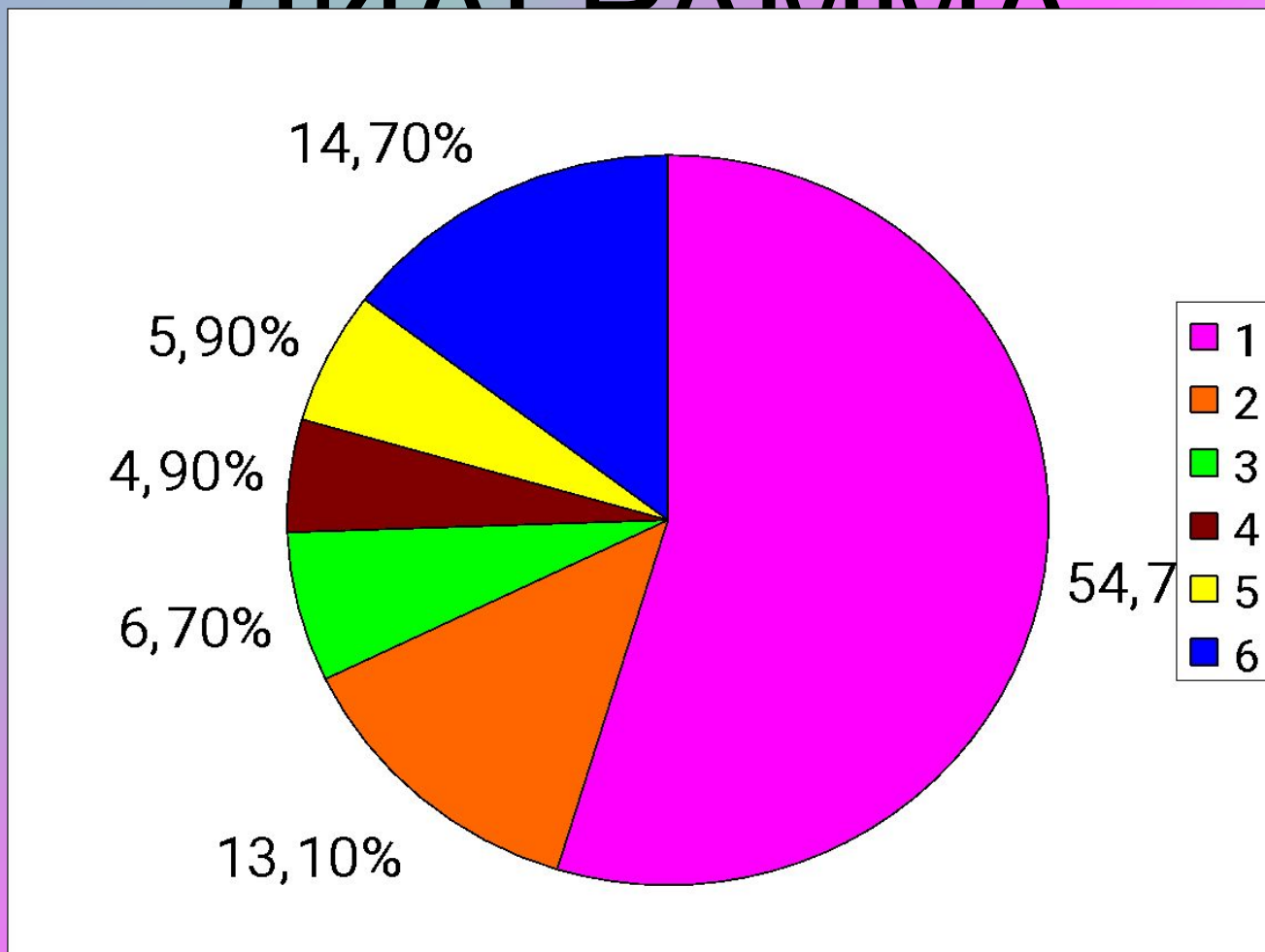
1. Диаграммы
2. Картограммы
3. Картодиаграммы

# ЛИНЕЙНАЯ ДИАГРАММА



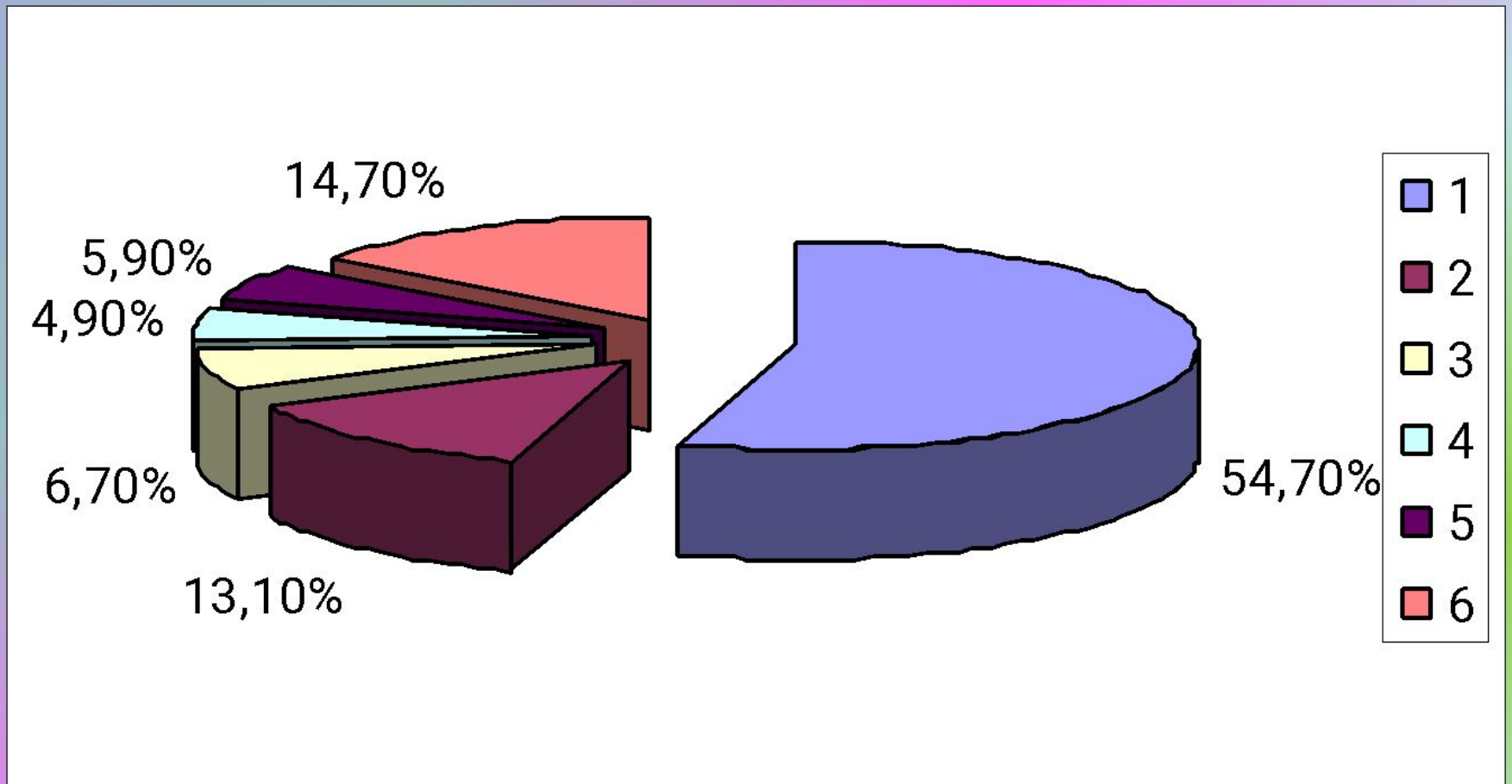
Уровень рождаемости в городе А. с 1990 по 2002 год /на 1000 населения/

# СЕКТОРНАЯ ДИАГРАММА

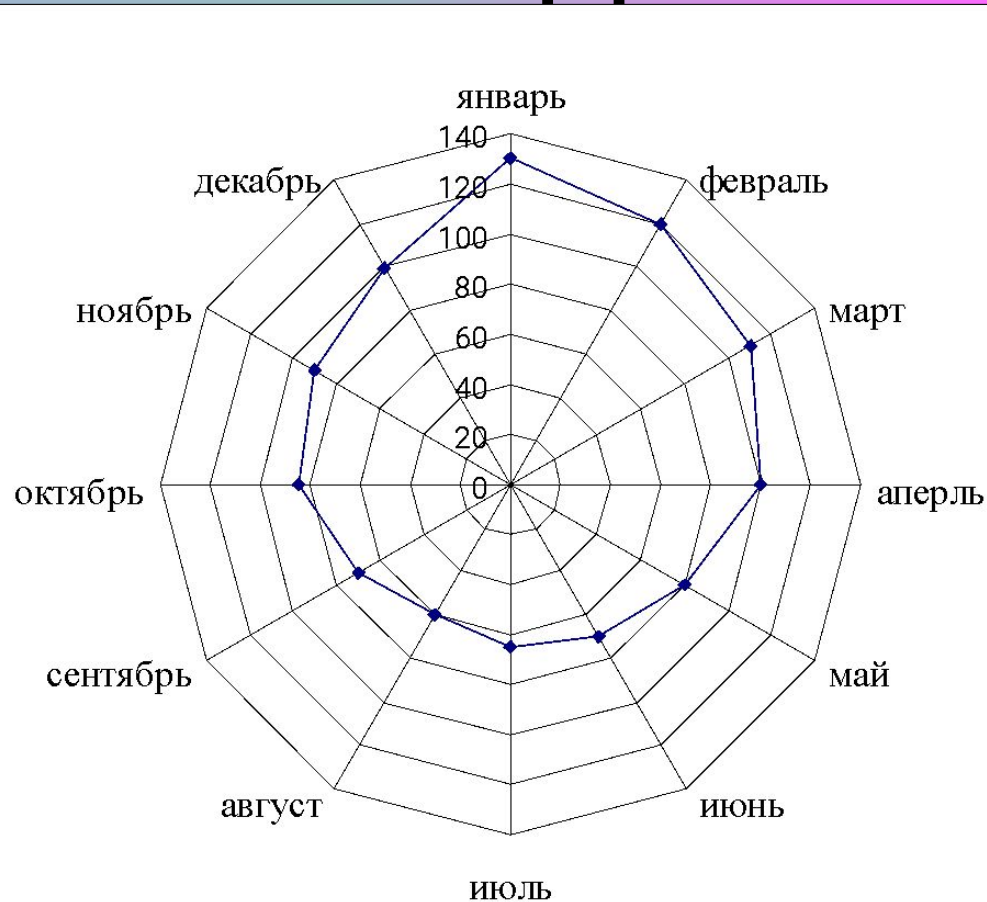


Структура заболеваемости населения в г. А.  
в 2005 г.

# Секторная диаграмма



# РАДИАЛЬНАЯ ДИАГРАММА



Число случаев заболеваемости с временной потерей трудоспособности на 100 работающих на Н-ском комбайновом заводе в 2006 г.



**Спасибо за  
внимание!**

BUKET-EXPRESS.KIEV.UA