

## Статистические оценки параметров генеральной совокупности

### Точечная оценка

Для полного изучения генеральной совокупности помимо закона её распределения во многих случаях приходится выполнять оценку параметров статистической совокупности (например, генеральной средней, генеральной дисперсии, моды и др.) или устанавливать связи между варьирующимися признаками (корреляционные связи).

Числовые характеристики генеральной совокупности называются *генеральными параметрами*, или просто *параметрами*. Отметим, что зачастую в распоряжении исследователя имеются лишь данные выборки, на основании которых и производится оценка искомого количественного генерального признака. Как правило, характеристики, получаемые по различным выборкам, отличаются друг от друга. Вообще говоря, ни при каком объёме выборки нельзя определить точное значение неизвестного параметра, можно лишь найти его приближенное значение, которое является *оценкой* неизвестного параметра по выборке.

**Оценка параметра** – определенная числовая характеристика, полученная из выборки. Когда одно отдельное значение используется для оценки параметра, то такая оценка называется *точечной оценкой генерального параметра* (средняя, дисперсия и др.). Всякая оценка является функцией результата наблюдений, так как чем больше и представительнее выборка, тем оценка будет более точной.

Выборочная совокупность в одних случаях организуется для определения среднего значения признака в генеральной совокупности (средний доход, средняя продолжительность жизни, средний урожай и т. п.); в других – для определения доли членов генеральной совокупности, обладающих интересующим аналитика признаком (процент всхожих семян определённого сорта, доля женских особей в выборке и т. п.)

**Генеральной средней**  $\bar{X}$  (математическим ожиданием  $\mu$ ) дискретной статистической совокупности называется число, которое является центром рассеяния (варьирования) для всех значений изучаемого признака.

**Выборочная** (эмпирическая) **средняя** (от лат. *empirio* – опыт) определяется как среднее арифметическое всех выборочных значений признака:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

или, если значения отдельных вариантов повторяются:

$$\bar{x} = \frac{\sum_{i=1}^k (n_i \cdot x_i)}{n}, \quad (2)$$

где  $n$  – объём выборки,  $k$  – количество отличающихся друг от друга вариантов,  $n_i$  – количество значений вариант  $x_i$  в выборке.

**Генеральная средняя** (теоретическая средняя) дискретного распределения находится как среднее арифметическое всех значений изучаемого признака:

$$\mu = \bar{X} = \frac{\sum_{i=1}^N x_i}{N} \text{ или } \mu = \sum_{i=1}^N x_i \cdot \omega_i, \quad (3)$$

где  $N$  – объём генеральной совокупности.

Отметим, что оценкой генеральной средней  $\bar{X}$  является выборочная средняя  $\bar{x}$ .

**Пример.** Возьмём две выборки и сравним данные подсчёта количества стеблей у 10 растений двух сортов ячменя:

сорт <i>A</i>	2	2	3	3	3	3	3	3	4	4
сорт <i>B</i>	1	1	2	2	3	3	3	4	5	6

*Решение.* Найдём выборочные средние (2):

• для сорта *A* она равна

$$\bar{x}_A = \frac{2+2+3+3+3+3+3+3+4+4}{10} = 3 \text{ (стебля)},$$

• для сорта *B* она равна

$$\bar{x}_B = \frac{1+1+2+2+3+3+3+4+5+6}{10} = 3 \text{ (стебля)}.$$

Средние выборочные обоих сортов одинаковы, в то же время нетрудно заметить, что большинство вариантов сорта *A* имеет значение, равное средней, а остальные – близкое к нему; во втором же сорте имеются варианты, резко отличающиеся от среднего значения, то есть изучаемый признак у сорта *A* варьирует значительно меньше, чем у сорта *B*, и в этом отношении сорта отличаются друг от друга.

Степень варьирования в выборке лучше всего характеризовать путём сравнения значения каждой из вариантов со средней выборочной, то есть посредством выборочной дисперсии и выборочного среднего квадратического отклонения.

**Выборочная дисперсия**  $\sigma^2$  равна среднему арифметическому квадратов отклонений всех значений вариант выборки от выборочной средней:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad (4)$$

или, если значения вариант повторяются:

$$\sigma^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n} = \frac{\sum_{i=1}^k (x_i)^2 \cdot n_i}{n} - (\bar{x})^2, \quad (5)$$

где  $n$  – объём выборки,  $k$  – количество отличающихся друг от друга вариант,  $n_i$  – количество значений вариант  $x_i$  в выборке.

**Генеральной дисперсией**  $\sigma_A^2$  называется величина, характеризующая степень рассеяния значений изучаемого признака относительно центра – генеральной средней или математического ожидания.

Генеральная дисперсия находится как среднее арифметическое квадратов отклонений всех значений изучаемого признака  $x_i$  в генеральной совокупности от генеральной средней  $\bar{X}$ :

$$\sigma_G^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}, \quad (6)$$

или, если значения вариант повторяются:

$$\sigma_G^2 = \frac{\sum_{i=1}^k (x_i - \mu)^2 \cdot n_i}{N}, \quad (7)$$

$N$  – объём генеральной совокупности,  $k$  – количество отличающихся друг от друга вариант,  $n_i$  – количество значений вариант  $x_i$  в совокупности.

Отметим, что оценкой генеральной дисперсии  $\sigma_A^2$  является соответствующий выборочный параметр  $\sigma^2$ , а выборочное среднее квадратическое отклонение  $\sigma$  рассматривается как оценка соответствующего генерального параметра  $\sigma_G$ .

Выборочные оценки параметров генеральной совокупности, как правило, по абсолютной величине не совпадают с соответствующими генеральными параметрами. Несовпадение между оцениваемым параметром и оценкой называется *статистической ошибкой параметра*, или *ошибкой репрезентативности*. При этом следует отметить, что с ростом объёма выборки ошибка репрезентативности уменьшается. Статистические ошибки возникают исключительно при осуществлении отбора элементов из генеральной совокупности, и их нельзя отождествлять с ошибкой измерения признака, например с инструментальной ошибкой, так как это ошибки, зависящие от объёма

и способа осуществления выборки. Следовательно, статистические отклонения, возникающие исключительно в процессе изъятия выборки, по существу являются одной из оценочных характеристик генеральной совокупности. Величина статистической ошибки измеряется **средним квадратическим отклонением**:

$$\sigma_{\Gamma} = \sqrt{\sigma_{\Gamma}^2}, \quad (8)$$

а при нормальном распределении признака – в единицах нормированного отклонения. Оценочным параметром для  $\sigma_{\Gamma}$  является **выборочное среднее квадратическое отклонение  $\sigma$** :

$$\sigma = \sqrt{\sigma^2}. \quad (9)$$

Оно является не только числовой оценкой варьирования признака в выборке, но и служит мерой отклонения отдельных вариантов при оценке генеральных параметров.

**Замечание.** Если вариационный ряд составлен по интервалам, то при нахождении средней арифметической, дисперсии, среднего квадратического отклонения вместо значений  $x_i$  в формулы (1) ÷ (9) подставляют середины соответствующих интервалов.

Рассмотрим другие числовые характеристики вариационного ряда.

**Мода (Mo)** – это значение признака, наиболее часто встречающееся в вариационном ряду. Например, при составлении годового плана потребности животноводческой фермы в комбикорме важнее знать не среднемесячное потребление, а моду, то есть месяц, в котором чаще требуется данный корм.

**Медианой (Me)** является значение признака ряда, относительно которого вариационный ряд делится на две равные по числу вариантов части.

**Замечание.** Если число значений признака  $n$  нечётное, то медианой является числовое значение ранжированного ряда, которое находится в центре распределения, то есть по обе стороны от значения признака, соответствующего медиане, находится одинаковое число значений ряда распределения. При чётном  $n$  за числовое значение  $Me$  принимается второе (правое) значение из двух центральных значений признака.

**Пример.** В ходе измерения роста (см) группы из 11 студентов полученные результаты записаны в порядке их поступления в виде простого статистического ряда:

165; 178; 169; 158; 175; 186; 173; 168; 181; 172; 167.

Найти моду и медиану.

**Решение.** Рассмотрим ранжированный ряд:

158; 165; 167; 168; 169; **172**; 173; 175; 178; 181; 186.

Для данного ряда мода не существует, а медиана (средний рост) равна  $Me = 172$  (см).

**Пример.** На определённом участке леса после выборочного обследования высоты муравейников были получены следующие данные (в см): 52, 53, 33, 29, 38, 42, 31, 53, 47, 36, 43, 47, 47, 40, 36, 54, 52, 58, 48, 43. Найти выборочную среднюю, выборочное среднее квадратическое отклонение, моду и медиану.

*Решение.* Используя исходные данные, составим ранжированный ряд:

29, 31, 33, 36, 36, 38, 40, 42, 43, 43, 47, 47, 47, 48, 52, 52, 53, 53, 54, 58.

Данные выборочного обследования занесем в таблицу:

$x_i$	29	31	33	36	38	40	42	43	47	48	52	53	54	58
$n_i$	1	1	1	2	1	1	1	2	3	1	2	2	1	1

где  $x_i$  – высота муравейника,  $n_i$  – количество муравейников.

Найдем выборочную среднюю (2):

$$\bar{x} = \frac{1}{20} \cdot (29 \cdot 1 + 31 \cdot 1 + 33 \cdot 1 + 36 \cdot 2 + 38 \cdot 1 + 40 \cdot 1 + 42 \cdot 1 + 43 \cdot 2 + 47 \cdot 3 + 48 \cdot 1 + 52 \cdot 2 + 53 \cdot 2 + 54 \cdot 1 + 58 \cdot 1) = 44,1 \text{ (см)}.$$

Найдем выборочную дисперсию:

$$\sigma_g^2 = (29^2 + 31^2 + 33^2 + 36^2 \cdot 2 + 38^2 + 40^2 + 42^2 + 43^2 \cdot 2 + 47^2 \cdot 3 + 48^2 + 52^2 \cdot 2 + 53^2 \cdot 2 + 54^2 + 58^2) / 20 - (44,1)^2 = 66,49 \text{ (см}^2\text{)}.$$

Выборочное среднее квадратическое отклонение равно  $\sigma_g = \sqrt{\sigma_g^2} = \sqrt{66,49} \approx 8,2$  (см).

В нашем случае мода  $Mo$  и медиана  $Me$  равны 47 см.

Таким образом, средняя высота муравейников составляет 44,1 см, наиболее часто встречаются муравейники высотой в 47 см.

**Замечание.** Если вариационный ряд составлен по интервалам значений генеральной совокупности, то мода находится по следующей формуле:

$$M_o = x_{Mo} + \Delta x \cdot \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}, \quad (10)$$

где  $x_{Mo}$  – начало модального интервала (интервала, имеющего максимальную частоту),  $\Delta x$  – длина модального интервала,  $n_i$  – частота модального интервала,  $n_{i-1}$  и  $n_{i+1}$  – частоты соответственно предшествующего и последующего за модальным интервалов. В свою очередь, медиана вычисляется по следующей формуле:

$$M_e = x_{Me} + \Delta x \cdot \frac{n/2 - T_{i-1}}{n_i}, \quad (11)$$

где  $x_{Me}$  – начало медианного интервала (интервала, в котором содержится срединный элемент),  $\Delta x$  – длина медианного интервала,  $n$  – объем выборки,  $T_{i-1}$  – сумма частот интервалов, предшествующих медианному,  $n_i$  – частота медианного интервала.

**Пример.** Пусть дан ряд распределения кальция (мг%) в сыворотке крови шимпанзе:

Классы по содержанию кальция в сыворотке крови	Средние значения в классе	Частоты	Накопленные частоты
8,6-9,3	9,0	2	2
9,4-10,1	9,8	6	8
10,2-10,9	10,6	15	23
11,0-11,7	11,4	23	46
11,8-12,5	12,2	25	71
12,6-13,3	13,0	17	88
13,4-14,1	13,8	7	95
14,2-14,9	14,6	5	100
Сумма	–	100	100

Найти моду и медиану для данного статистического ряда распределения.

*Решение.* Найдем моду. Модальный класс 11,8–12,5, ему соответствует частота, равная 25, длина интервала составляет 0,8. По формуле (10) имеем:

$$M_o = x_{Mo} + \Delta x \cdot \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}} = 11,8 + 0,8 \cdot \frac{25 - 23}{2 \cdot 25 - 23 - 17} = 11,8 + 0,16 = 11,96 \text{ (мг\%)}.$$

Находим медиану. В силу того что классов чётное число 8, то за медианный класс берём интервал 11,8–12,5 (пятый по счёту, так как  $8/2+1=5$ ),  $T_{i-1}=46$ . Следовательно, по формуле (13) имеем:

$$M_e = x_{Me} + \Delta x \cdot \frac{n/2 - T_{i-1}}{n_i} = 11,8 + 0,8 \cdot \frac{100/2 - 46}{25} = 11,8 + 0,128 = 11,93 \text{ (мг\%)}.$$

Таким образом, наиболее часто встречающееся содержание кальция в крови составляет 11,96 мг%, а среднее значение – 11,93 мг%.

**Замечание.** Некоторые выборки имеют не одну, а две или более моды. Это указывает на то, что в них объединён качественно разнородный материал.

Для того чтобы оценка параметра генеральной совокупности была корректной (ошибка не должна превышать установленного значения), как правило, используются математические формулы, вид которых определяется объективными законами теории вероятностей и требованиями практики.

### Требования к оценкам

Установлено, что для получения хорошего приближения при оценке параметров генеральной совокупности  $\mu$ ;  $\sigma^2$  и  $\sigma$  соответствующие параметры выборочной совокупности должны отвечать следующим требованиям: *несмещённости, состоятельности и эффективности.*

**Несмещённость.** Точечная оценка называется несмещённой, если математическое ожидание выборочного параметра при любых объёмах выборки «достаточно близко» или совпадает с его генеральным значением.

В частности, выборочная средняя  $\bar{x}$  является несмещённой оценкой генеральной средней, так как  $M(\bar{x}) = \mu$ . При этом выборочная дисперсия и среднее квадратическое отклонение являются смещёнными оценками соответствующих параметров генеральной совокупности ( $\sigma_{\bar{A}}^2 \neq \sigma^2$  и  $\sigma_{\bar{A}} \neq \sigma$ ). Следовательно, *несмещённая* оценка (устранение систематической ошибки) выборочной дисперсии равна:

$$s^2 = \sigma_{\bar{A}}^2 \cdot \frac{n}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}. \quad (12)$$

Отражением того, что исправленная выборочная дисперсия является *наилучшей* оценкой соответствующего генерального параметра, является приближённое равенство:  $\sigma_{\bar{A}}^2 \approx s^2$ .

**Состоятельность.** Точечная оценка называется состоятельной, если с ростом числа наблюдений (объёма выборки:  $n \rightarrow N$ ) выбранная оценка стремится к истинному значению оцениваемого параметра (то есть дисперсия оценки стремится к нулю). Данное определение основано на законе больших чисел (теорема Чебышева).

Оценки, обладающие свойствами несмещённости и состоятельности, при ограниченном числе опытов могут отличаться дисперсиями. При этом чем меньше дисперсия оценки, тем меньше вероятность грубой ошибки при определении приближенного значения параметра. Поэтому необходимо, чтобы дисперсия оценки была минимальной. Оценку, обладающую таким свойством, называют *эффективной*.

## Средняя и предельная ошибки

Следствием предельных законов теории вероятностей является и тот факт, что выборочные средние  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n$ , вычисленные на основании параметров  $n$  независимых выборок одной и той же нормально распределённой генеральной совокупности, имеют дисперсию в  $n$  раз меньше, чем каждая из них. Тогда величина средней ошибки репрезентативности выборочной средней может быть представлена.

Средняя ошибка:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}. \quad (13)$$

**Предельная ошибка выборки:**

$$\Delta = \bar{x} - \mu = t \cdot s_{\bar{x}}, \quad (14)$$

где  $t$  определяется как аргумент интегральной функции Лапласа (**Приложение III**) по заданной вероятности  $p$  при условии  $\Phi(t) = \frac{p}{2}$ .

**Пример.** Исследуется определённый сорт картофеля по такому признаку, как масса клубня (96 клубней). Все данные сведены в одну специальную таблицу (см. задачу из главы 1 данного раздела):

Интервалы	0,26- 0,28	0,29- 0,31	0,32- 0,34	0,35- 0,37	0,38- 0,4	0,41- 0,43	0,44- 0,46
Средние значения в интервале	0,275	0,305	0,335	0,365	0,395	0,425	0,455
Относительные частоты, $\omega_i$	0,03	0,19	0,32	0,19	0,18	0,08	0,01

*Решение.* Найдем среднее арифметическое данного мерного признака (2).

$$\begin{aligned} \bar{x} = & 0,275 \cdot 0,03 + 0,305 \cdot 0,19 + 0,335 \cdot 0,32 + 0,365 \cdot 0,19 + \\ & + 0,395 \cdot 0,18 + 0,425 \cdot 0,08 + 0,455 \cdot 0,01 = 0,3524 \text{ (кг)}. \end{aligned}$$

Найдём «исправленное» среднее квадратическое отклонение данного мерного признака, используя формулы (5) и (12).

$$\begin{aligned} \sigma_e^2 = & [ (0,275)^2 \cdot 0,03 + (0,305)^2 \cdot 0,19 + (0,335)^2 \cdot 0,32 + (0,365)^2 \cdot 0,19 + \\ & + (0,395)^2 \cdot 0,18 + (0,425)^2 \cdot 0,08 + (0,455)^2 \cdot 0,01 ] / 96 - (0,3524)^2 = 0,00159 \end{aligned}$$

$$s^2 = \sigma_g^2 \cdot \left( \frac{n}{n-1} \right) = 0,00159 \cdot \frac{96}{96-1} = 0,00161,$$

$$s = \sqrt{s^2} = \sqrt{0,00161} \approx 0,0401.$$

Далее определим ошибку среднего арифметического (13).

$$s_{\bar{x}} = \frac{s}{\sqrt{96}} = \frac{\pm 0,0401}{\sqrt{96}} \approx \pm 0,0041 \text{ (кг)}.$$

Таким образом, средняя масса клубня определённого сорта в генеральной совокупности находится в пределах  $0,3524 \pm 0,0041$ , то есть от 0,3483 кг до 0,3565 кг.

Найдём предельную ошибку выборки (14) с доверительной вероятностью 0,95 (с надёжностью 95%), так как значение интегральной функции Лапласа равно  $\Phi(t) = \frac{0,95}{2} = 0,475$ , то ему соответствует по таблице из **Приложения III** значение аргумента, равное  $t=1,96$ . Тогда предельная ошибка будет равна

$$\Delta = t \cdot s_{\bar{x}} = 1,96 \cdot 0,0041 = 0,008036 \text{ (кг)}.$$

Следовательно, предельная ошибка выборки приблизительно равна 8 г.

### Интервальная оценка

Обозначим в общем случае через  $\Theta^*$  выборочную оценку параметра  $\Theta$  генеральной совокупности ( $\mu$ ,  $\sigma^2$  и  $\sigma$ ). Теория статистической оценки рассматривает помимо точечной оценки и оценку интервальную.

*Интервальная оценка – числовой интервал, определяемый двумя числами  $(\Theta_1^*, \Theta_2^*)$ , содержащий неизвестный параметр генеральной совокупности.* Интервальная оценка более информативная мера, чем точечная оценка (это множество точечных оценок, зависящее от результатов наблюдений).

Если для произвольного числа  $\varepsilon > 0$  выполняется неравенство:

$$|\Theta - \Theta^*| < \varepsilon,$$

то положительное число  $\varepsilon$  характеризует **точность оценки**. В интервальной оценке устанавливается **доверительная вероятность (надёжность)**, с которой эта оценка накроет неизвестный параметр, то есть это вероятность  $p$ , с которой выполняется неравенство  $|\Theta - \Theta^*| < \varepsilon$ .

Статистические методы позволяют получать лишь те интервальные оценки, доверительная вероятность  $p$  которых близка к единице. Наиболее часто доверительную вероятность принимают равной 0,9; 0,95; 0,99; 0,999. При разработке стандартов используется степень надёжности 0,99.

Интервальной оценкой обычно является так называемый доверительный интервал  $(\Theta^* - \varepsilon; \Theta^* + \varepsilon)$ , длина которого равна  $2\varepsilon$  и является функцией результатов наблюдений.

*Доверительным интервалом* называют найденный по данным выборки интервал  $(\Theta^* - \varepsilon; \Theta^* + \varepsilon)$ , который покрывает параметр  $\Theta$  с заданной надёжностью  $p$ .

**1. Доверительный интервал для математического ожидания при известном  $\sigma$ .** В некоторых случаях среднее квадратическое отклонение  $\sigma$  ошибки измерения, а вместе с нею и самого измерения бывает известно.

**Пример.** Если измерения производятся одним и тем же прибором при одних и тех же условиях, то  $\sigma$  для всех измерений одно и то же и обычно бывает известно. При этом с надёжностью  $p$  можно утверждать, что доверительный интервал  $\left(\bar{x} - \frac{t \cdot \sigma}{\sqrt{n}}; \bar{x} + \frac{t \cdot \sigma}{\sqrt{n}}\right)$  покрывает неизвестный параметр с точностью оценки

$\varepsilon = \frac{t \cdot \sigma}{\sqrt{n}}$ . Здесь число  $t$  определяется из равенства  $\Phi(t) = \frac{p}{2}$ , где  $\Phi(t)$  – интегральная функция Лапласа (см. **Приложение II**).

**Пример.** Признак  $X$  распределен в генеральной совокупности нормально с известным  $\sigma = 0,40$ . Найти по данным выборки доверительный интервал для математического ожидания с доверительной вероятностью  $p = 0,99$ , если  $n = 20$ ,  $\bar{x} = 6,34$ .

*Решение.* Для  $\Phi(t) = \frac{p}{2} = \frac{0,99}{2} = 0,495$  по **Приложению II** найдем  $t = 2,58$ , следовательно,  $\varepsilon = \frac{t \cdot \sigma}{\sqrt{n}} = 2,58 \cdot \frac{0,40}{\sqrt{20}} \approx 0,23$ . Таким образом, доверительный интервал  $(6,34 - 0,23; 6,34 + 0,23)$ , то есть  $(6,11; 6,57)$  покрывает математическое ожидание с надёжностью  $0,99$ .

**2. Доверительный интервал для математического ожидания при неизвестном  $\sigma$ .** Пусть случайная величина  $X$  имеет нормальное распределение с неизвестными нам параметрами  $\mu$  и  $\sigma$ , и не зависящее от них. В этом случае используется распределение Стьюдента и интервальной оценкой математического ожидания генеральной совокупности с заданной надёжностью  $p$  будет являться интервал  $\left(\bar{x} - t_p(f) \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_p(f) \cdot \frac{s}{\sqrt{n}}\right)$  с границей доверительного интервала (точность оценки)  $\varepsilon = t_p(f) \cdot \frac{s}{\sqrt{n}}$ .

Параметр  $t_p(f)$  (коэффициент Стьюдента) для  $f = n - 1$  и доверительной вероятности  $p$  находят по таблицам распределения Стьюдента или вычисляют с помощью ПЭВМ (см. **Приложение IV**),  $s$  – исправленная оценка генерального среднего квадратического отклонения, находится по формуле (9).

**Пример.** Распределение кальция в сыворотке крови шимпанзе характеризуется следующими показателями. Число животных равно 30, среднее значение распределения кальция в сыворотке крови шимпанзе для данной выборки составляет 12,4 мг%. Несмещённое выборочное среднее квадратическое отклонение равно 1,3 мг%. Найти 90%-й доверительный интервал для среднего значения распределения кальция в сыворотке крови шимпанзе, соответствующего генеральной совокупности.

*Решение:* Из таблицы коэффициентов Стьюдента (**Приложение IV**) находим  $t_{0,9}(30 - 1) = 1,697$ . Таким образом, граница доверительного интервала равна  $\varepsilon = t_p(f) \cdot \frac{s}{\sqrt{n}} = 1,697 \cdot \frac{1,3}{\sqrt{30}} \approx 0,4$ .

Истинное значение распределения кальция в сыворотке крови шимпанзе с доверительной вероятностью 0,90 заключено в интервале  $\mu = 12,4 \pm 0,4$  или (12; 12,8). Следовательно, можно быть на 90% уверенными, что среднее значение распределения кальция в сыворотке крови шимпанзе находится между 12 мг% и 12,8 мг%.

**3. Доверительный интервал для среднего квадратического отклонения  $\sigma$ .** С надежностью  $p$  можно утверждать, что доверительный интервал  $(s - sq; s + sq)$  покрывает неизвестный параметр с точностью оценки  $\varepsilon = sq$ , где  $s = \sigma \sqrt{\frac{n}{n-1}}$ . В **Приложении I** приведена таблица  $q = q(p; n)$ , где  $n$  – объем выборки.

**Пример.** Признак  $X$  распределен в генеральной совокупности нормально. Найти доверительный интервал для  $\sigma$  с надежностью  $p = 0,95$ , если  $n = 20$ ,  $s = 0,40$ .

*Решение.* Для надежности  $p = 0,95$  и  $n = 20$  находим в **Приложении I**  $q = 0,37$ . Далее  $s \cdot q = 0,40 \cdot 0,37 = 0,148$ . Доверительный интервал  $(0,40 - 0,15; 0,40 + 0,15)$ , то есть  $(0,25; 0,55)$  покрывает  $\sigma$  с надежностью 0,95.

## Литература

1. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стенли. – М.: Прогресс, 1976. – 496 с.
2. Гланц С. Медико-биологическая статистика / С. Гланц. - М.: Практика, 1998. - 459 с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман. – М.: Высш. шк., 2003. – 479 с.
4. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи) / Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 67 с.
5. Новиков Д. А. Статистические методы в медико-биологическом эксперименте (типовые случаи) / Д. А. Новиков, В. В. Новочадов. – Волгоград: Изд-во ВГМУ, 2005. – 84 с.
6. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилков. – Киров: Изд-во ВГГУ, 2015. – 158 с.

# Приложения

## Приложение I

Таблица значений  $q = q(p, n)$

$n$	$p$	0,95	0,99	0,999
5		1,37	2,67	5,64
6		1,09	2,01	3,88
7		0,92	1,62	2,98
8		0,80	1,38	2,42
9		0,71	1,20	2,06
10		0,65	1,08	1,80
11		0,59	0,98	1,60
12		0,55	0,90	1,45
13		0,52	0,83	1,33
14		0,48	0,78	1,23
15		0,46	0,73	1,15
16		0,44	0,70	1,07
17		0,42	0,66	1,01
18		0,40	0,63	0,96
19		0,39	0,60	0,92
20		0,37	0,58	0,88
25		0,32	0,49	0,73
30		0,28	0,43	0,63
35		0,26	0,38	0,56
40		0,24	0,35	0,50
45		0,22	0,32	0,46
50		0,21	0,30	0,43
60		0,188	0,269	0,38
70		0,174	0,245	0,34

80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,221
200	0,099	0,136	0,185
250	0,089	0,120	0,162

## Приложение II

### Значения интегральной функции Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

x	0	1	2	3	4	5	6	7	8	9
<b>0,0</b>	0,0000	0,0039	0,0079	0,0119	0,0159	0,0199	0,0239	0,0279	0,0318	0,0358
<b>0,1</b>	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
<b>0,2</b>	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
<b>0,3</b>	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
<b>0,4</b>	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
<b>0,5</b>	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
<b>0,6</b>	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
<b>0,7</b>	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
<b>0,8</b>	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3079	0,3106	0,3133
<b>0,9</b>	0,3159	0,3185	0,3212	0,3238	0,3263	0,3289	0,3314	0,3339	0,3364	0,3389
<b>1,0</b>	0,3413	0,3437	0,3461	0,3484	0,3508	0,3531	0,3554	0,3576	0,3599	0,3621
<b>1,1</b>	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3769	0,3789	0,3810	0,3829
<b>1,2</b>	0,3849	0,3869	0,3887	0,3906	0,3925	0,3944	0,3962	0,3979	0,3997	0,4014
<b>1,3</b>	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
<b>1,4</b>	0,4192	0,4207	0,4221	0,4236	0,4250	0,4264	0,4278	0,4292	0,4305	0,4318
<b>1,5</b>	0,4331	0,4344	0,4357	0,4369	0,4382	0,4394	0,4406	0,4417	0,4429	0,4440
<b>1,6</b>	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4544

<b>x</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>1,7</b>	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
<b>1,8</b>	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
<b>1,9</b>	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
<b>2,0</b>	0,4773	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
<b>2,1</b>	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
<b>2,2</b>	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4889
<b>2,3</b>	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,490863	0,4911	0,4913	0,4916
<b>2,4</b>	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
<b>2,5</b>	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
<b>2,6</b>	0,4953	0,4955	0,4956	0,4957	0,4959	0,4959	0,4961	0,4962	0,4963	0,4964
<b>2,7</b>	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
<b>2,8</b>	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
<b>2,9</b>	0,4981	0,4982	0,4983	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
<b>3,0</b>	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
<b>3,1</b>	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
<b>3,2</b>	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
<b>3,3</b>	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
<b>3,4</b>	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997
<b>3,5</b>	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
<b>3,6</b>	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
<b>3,7</b>	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
<b>3,8</b>	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999

x	0	1	2	3	4	5	6	7	8	9
3,9	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,0	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,1	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,2	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,3	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,4	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,5	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,6	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
5,0	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Для вычисления значения функции при отрицательных значениях аргумента используется соотношение  $\Phi(-x) = -\Phi(x)$ .

 В Microsoft Excel для расчёта интегральной функции Лапласа используется формула **НОРМ.СТ.РАСП(x)-0,5**.

Значения аргумента интегральной функции Лапласа

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} dx$$

$\alpha$	$p$	$\Phi(t)$	$t$
0,01	0,99	0,495	2,575829
0,02	0,98	0,490	2,326348
0,03	0,97	0,485	2,170090
0,04	0,96	0,480	2,053749
0,05	0,95	0,475	1,959964
0,10	0,9	0,450	1,644854
0,20	0,8	0,400	1,281552
0,30	0,7	0,350	1,036433
0,40	0,6	0,300	0,841621
0,50	0,5	0,250	0,674490

**Замечание**



В Microsoft Excel для расчёта аргумента интегральной функции Лапласа  $t$  используются формулы:  $p=1-\alpha$ ,  $\Phi=p/2$ ,  $t=\text{НОРМ.СТ.ОБР}(\Phi+0,5)$ .

Критические значения  $t_{\text{крит}}(p, k)$  распределения Стьюдента

$k$	$\alpha(p)$			$k$	$\alpha(p)$		
	0,1 (0,9)	0,05 (0,95)	0,01 (0,99)		0,1 (0,9)	0,05 (0,95)	0,01 (0,99)
<b>1</b>	6,31375	12,70615	63,65590	<b>18</b>	1,73406	2,10092	2,87844
<b>2</b>	2,91999	4,30266	9,92499	<b>19</b>	1,72913	2,09302	2,86094
<b>3</b>	2,35336	3,18245	5,84085	<b>20</b>	1,72472	2,08596	2,84534
<b>4</b>	2,13185	2,77645	4,60408	<b>21</b>	1,72074	2,07961	2,83137
<b>5</b>	2,01505	2,57058	4,03212	<b>22</b>	1,71714	2,07388	2,81876
<b>6</b>	1,94318	2,44691	3,70743	<b>23</b>	1,71387	2,06865	2,80734
<b>7</b>	1,89458	2,36462	3,49948	<b>24</b>	1,71088	2,0639	2,79695
<b>8</b>	1,85955	2,30601	3,35538	<b>25</b>	1,70814	2,05954	2,78744
<b>9</b>	1,83311	2,26216	3,24984	<b>26</b>	1,70562	2,05553	2,77872
<b>10</b>	1,81246	2,22814	3,16926	<b>27</b>	1,70329	2,05183	2,77068
<b>11</b>	1,79588	2,20099	3,10582	<b>28</b>	1,70113	2,04841	2,76326
<b>12</b>	1,78229	2,17881	3,05454	<b>29</b>	1,69913	2,04523	2,75639
<b>13</b>	1,77093	2,16037	3,01228	<b>30</b>	1,69726	2,04227	2,74998
<b>14</b>	1,76131	2,14479	2,97685	<b>40</b>	1,68385	2,02107	2,70446
<b>15</b>	1,75305	2,13145	2,94673	<b>60</b>	1,67065	2,00030	2,66027
<b>16</b>	1,74588	2,11990	2,92079	<b>120</b>	1,65765	1,97993	2,61742
<b>17</b>	1,73961	2,10982	2,89823	$\infty$	1,64484	1,95996	2,57583

**Замечания**



В Microsoft Excel для расчёта критических значений распределения Стьюдента используется функция: **СТЮДЕНТ.ОБР.2X( $\alpha$ ;  $k$ )**, где  $\alpha = 1 -$