Министерство образования и науки Российской Федерации
Магнитогорский Государственный технический университет
имени Г. И. Носова

Кафедра технологий обработки материалов

ОБРАБОТКА И АНАЛИЗ ВЫБОРКИ

Методические указания

Магнитогорск
2017

1.ЦЕЛИ РАБОТЫ

Изучение закономерностей всякой случайной величины возможно только по выборке значений из ее генеральной совокупности. Поэтому для получения достоверных сведений о случайной величине прежде всего необходимо, чтобы выборка не содержала грубых погрешностей (значений, которые для данной случайной величины не характерны).

Если грубые погрешности в выборке отсутствуют, то достоверность найденных значений параметров распределения будет определяться особенностями соответствующих выборочных оценок. Наилучшие результаты обеспечиваются, если оценки обладают свойствами состоятельности, несмещенности и эффективности.

Цели работы:

- Изучение методики выявления и отсеивания грубых погрешностей в выборке.
- Приобретение навыков выявления и отсеивания грубых погрешностей выборки в среде электронных таблиц *MS Excel*.
- Изучение особенностей выборочных оценок параметров распределения случайной величины.
- Приобретение навыков определения выборочных оценок в среде электронных таблиц *MS Excel*.

2.КРАТКИЕ СВЕДЕНИЯ ИЗ СТАТИСТИКИ

Наиболее полно закономерности случайной величины X отображаются функцией $F(x)$ и плотностью $f(x) = dF(x)/dx$ распределения вероятности. На практике также применяются числовые характеристики, среди которых различают:

* характеристики положения (центральной тенденции);
* характеристики рассеяния (вариации);
* характеристики формы распределения (асимметрия и эксцесс).

2.1.Характеристики положения

Математическое ожидание. Математическое ожидание M_X представляет собой такое значение случайной величины, около которого сосредоточены все возможные ее значения. Наилучшей оценкой математического ожидания (т. е. и состоятельной, и несмещенной, и эффективной) является выборочное среднее, рассчитываемое как среднее арифметическое:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \approx M_X . \qquad (1)$$

Медиана. Медиана Me - это такое значение случайной величины, что для 50% ее возможных значений выполняется условие $x < Me$, а для других 50% выполняется условие $x > Me$. На графике функции распределения медиана есть абсцисса, которой соответствует значение $F(x)$=0,5. На графике плотности распределения медиана есть абсцисса, которая делит площадь под кривой $f(x)$ на две равные части.

Для нахождения медианы по выборке необходимо сначала выполнить ее ранжирование (расположить ее элементы) по возрастанию. Если объем выборки есть нечетное число, то оценкой медианы будет элемент, расположенный в середине ряда:

$$Me \approx x_{(n+1)/2} . \qquad (2)$$

Если объем выборки есть четное число, то:

$$Me \approx \frac{x_{n/2} + x_{n/2+1}}{2} . \qquad (3)$$

Мода. Мода Mo - значение случайной величины, вероятность появления которого наибольшая. На графике плотности вероятности мода есть абсцисса, соответствующая максимуму кривой $f(x)$. При оценке моды по выборке Mo принимают равным тому значению случайной величины, которое встречается в выборке наиболее часто.

2.2.Характеристики рассеяния

Дисперсия. Дисперсия Dx - математическое ожидание квадратов отклонений значений случайной величины от ее математического ожидания:

$$D_X = M\left[(x - M_X)^2\right]. \qquad (4)$$

На основании выборки дисперсию случайной величины оценивают следующими характеристиками: *дисперсия распределения*

s^2 и *выборочная дисперсия* $\bar{s}^2$. Указанные выборочные оценки дисперсии рассчитываются по формулам:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 ; \qquad (5)$$

$$\bar{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 . \qquad (6)$$

Выборочная дисперсия (6) является эффективной, несмещенной и состоятельной оценкой при любом объеме выборки. Для дисперсии распределения (5) состоятельность также обеспечивается при любом объеме выборки, но несмещенность и эффективность достигаются только при n > 30.

Среднее квадратическое отклонение (стандартное отклонение, стандарт). Среднее квадратическое отклонение случайной величины есть корень квадратный из ее дисперсии:

$$\sigma_X = \sqrt{D_X} . \qquad (7)$$

Наилучшей выборочной оценкой σ_X является *выборочное среднее квадратическое отклонение (выборочное стандартное отклонение)*:

$$\bar{s} = \sqrt{\bar{s}^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \approx \sigma_X . \qquad (8)$$

Размах. Размах R - разность между наибольшим и наименьшим значениями случайной величины, обнаруженными в выборке:

$$R = x_{max} - x_{min} . \qquad (9)$$

Коэффициент вариации. Коэффициент вариации V_x характеризует, какую долю от математического ожидания случайной величины составлет ее среднее квадратическое отклонение:

$$\vartheta_X = \frac{\sigma_X}{M_X} \approx V_x = \frac{\bar{s}}{\bar{x}} . \qquad (10)$$

Будучи безразмерной величиной, коэффициент вариации позволяет сравнивать степень рассеяния различных случайных величин. Его значение может быть представлено как в относительных единицах, так и в процентах. В последнем случае результат, полученный расчетом по формуле (9), необходимо умножить на 100.

2.3.Характеристики формы распределения

Коэффициент асимметрии A**.** Характеризует степень несимметричности распределения относительно его математического ожидания:

$$A = \frac{\eta_3}{\sigma^3} \approx \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\bar{s}_x} \right)^2, \qquad (11)$$

где η_3 - центральный момент третьего порядка.

Положительная асимметрия ($A > 0$) указывает на отклонение значений случайной величины в сторону значений, меньших математического ожидания. Отрицательная асимметрия ($A < 0$) означает, что значения случайной величины имеют тенденцию к отклонению в сторону значений, больших математического ожидания.

Коэффициент эксцесса E**.** Характеризует относительную остроконечность или сглаженность распределения изучаемой случайной величины по сравнению с нормальным распределением, для которого $E = 3$:

$$E \approx \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4 \right] - \frac{3(n-1)}{(n-2)(n-3)}. \qquad (12)$$

Положительный эксцесс ($E > 0$) означает, что распределение изучаемой случайной величины более остроконечное, чем нормальное. Отрицательный эксцесс ($E < 0$) указывает на относительную сглаженность изучаемого распределения по сравнению с нормальным.

2.4.Интервальные оценки

Интервальные оценки характеризуют ошибку оценивания истинного значения ξ некоторой характеристики случайной величины с помощью соответствующей выборочной оценки ξ^*. Их применение необходимо потому, что каждая выборочная оценка сама по себе является случайной величиной с некоторым распределением вероятности.

При интервальном оценивании используют *доверительную вероятность*, *уровень значимости*, *доверительный интервал* и *доверительные границы*.

Доверительная вероятность – вероятность события, заключающегося в том, что ошибка оценивания истинного значения некоторого параметра распределения случайной величины его выборочной оценкой не превышает величины Δ_ξ:

$$p = Prob\left(\left|\xi - \xi^*\right| \le \Delta_\xi\right). \tag{13}$$

Уровень значимости – вероятность события, заключающегося в том, что ошибка оценивания истинного значения некоторого параметра распределения случайной величины его выборочной оценкой превышает величину ε_ξ:

$$\alpha = Prob\left(\left|\xi - \xi^*\right| > \Delta_\xi\right). \tag{14}$$

Доверительная вероятность и уровень значимости связаны соотношением:

$$p + \alpha = 1. \tag{15}$$

Доверительный интервал – интервал значений выборочной характеристики, внутри которого истинное значение оцениваемого параметра находится с заданной доверительной вероятностью.

Доверительные границы – значения выборочной оценки, представляющие собой границы доверительного интервала:

$$\xi_1^* = \xi^* - \Delta_\xi; \tag{16}$$

$$\xi_2^* = \xi^* + \Delta_\xi; \tag{17}$$

В практике обработки и анализа числовой информации наиболее часто встречается задача интервального оценивания выборочного среднего $\bar{x}$. При ее решении исходят из того, что $\bar{x}$ как случайная величина имеет нормальное распределение с математическим ожиданием $Mx = \bar{x}$ и средним квадратическим отклонением:

$$s_x = \frac{s}{\sqrt{n}}. \tag{18}$$

Величина s_x имеет также специальное название - *стандартная ошибка выборочного среднего*.

Доверительные границы для выборочного среднего:

$$\Delta_x = \pm\ s_x\, t[\alpha\, ; n-1], \tag{19}$$

где $t[\alpha\, ; n-1]$ - табличное значение (квантиль) распределения Стьюдента при уровне значимости α и числе степеней свободы $\nu = n-1$.

Зная доверительные границы выборочного среднего, можно утверждать, что с вероятностью $p = 1-\alpha$ анализируемая случайная величина имеет значение $\overline{x} \pm \Delta_x$.

2.5.Отсеивание грубых погрешностей (проверка однородности выборки)

Выборка из генеральной совокупности получается в результате использования тех или иных средств измерения. Вследствие поломки прибора или по недосмотру исследователя среди элементов выборки могут оказаться результаты, содержащие грубую ошибку (погрешность). Ошибочные значения в дальнейшем анализе искажают информацию об исследуемом явлении. Поэтому перед тем, как использовать выборочные данные для каких-либо выводов, необходимо исследовать их на наличие грубых погрешностей.

Для анализа закономерностей случайной величины по выборке необходимо также иметь гарантию, что полученная выборка сделана из генеральной совокупности именно того параметра, который интересует исследователя, и в ней отсутствуют элементы из какой-либо иной генеральной совокупности. Выборку, которая содержит элементы только одной генеральной совокупности, принято называть *однородной*, а выявление «чужеродных» элементов – *проверкой однородности*.

И обнаружение грубых погрешностей, и проверка однородности по существу сводятся к одной и той же задаче – среди элементов выборки необходимо выявить такие, которые не характерны для исследуемой генеральной совокупности. Известно достаточно много методов решения этой задачи. Рассмотрим метод, рекомендуемый Е. Н. Львовским [1].

Сначала необходимо определить среднее выборочное $\bar{x}$ и несмещенное среднее квадратическое отклонение $\bar{s}$.

Затем среди элементов выборки необходимо выявить результат, который дает наибольшее по абсолютной величине отклонение от среднего выборочного:

$$d_{max} = max\{|x_i - \bar{x}|\}, \tag{20}$$

Далее вычисляются статистики τ_{max} и $\tau[\alpha;n-2]$ для уровней значимости 5 и 0,1%:

$$\tau = \frac{d_{max}}{\bar{s}}; \tag{21}$$

$$\tau[\alpha;n-2] = \frac{t[\alpha;n-2]\sqrt{n-1}}{\sqrt{(t[\alpha;n-2])^2 + (n-2)}}, \tag{22}$$

где $t[\alpha;n-2]$ - табличное число Стьюдента (квантиль распределения Стьюдента) для уровня значимости α и числе степеней свободы $\nu = n-2$.

Значения статистик сравниваются и делается вывод:

$\tau_{max} \leq \tau[0{,}05;n-2]$	- элемент выборки не является аномальным;
$\tau[0{,}05;n-2] < \tau_{max} \leq \tau[0{,}001;n-2]$	- возможно, что элемент выборки является аномальным;
$\tau_{max} > \tau[0{,}001;n-2]$	- элемент выборки является аномальным.

При обнаружении аномального элемента его необходимо исключить из выборки и повторить все предыдущие действия. Если же в отношении аномальности элемента выборки есть сомнения (второй случай соотношения статистики τ_{max} и $\tau[\alpha;n-2]$), то его можно оставить или отсеять по усмотрению исследователя.

3.ВЫПОЛНЕНИЕ РАБОТЫ

Работа выполняется на одном рабочем листе книги *MS Excel*.

Пример оформления листа приведен на рис. 1. Здесь представлены результаты анализа и обработки выборки, полученной по результатам наблюдений за толщиной полос, прокатываемых на широкополосном стане горячей прокатки при его настройке на номинал 2,5 мм.

Исходные данные (в примере - ячейки B2:B27) вводятся с клавиатуры.

3.1.Выявление и отсеивание грубых погрешностей в выборке

Сначала в ячейках G3, G4 и G5 с использованием статистических функций:

СЧЕТ(B3:B27), СРЗНАЧ(B3:B27) и СТАНДОТКЛОН(B3:B27)

программируется вычисление объема выборки (на рабочем листе обозначено как n), среднего выборочного (Xср) и выборочного среднеквадратического отклонения (S).

Внимание! Здесь и далее аргументы статистических функций и адреса отдельных ячеек записаны в соответствии с расположением данных в примере, приведенном на рис.1.

Затем в ячейках D3:D27 программируется расчет отклонения текущего значения случайной величины от среднего выборочного:

$$\left| d_i \right| = \left| x_i - \overline{x} \right|.$$

Например, в ячейке D3:

=ABS(B3-G4).

После этого в ячейку G6 вводится формула, обеспечивающая выбор наибольшего из отклонений:

=МАКС(D3:D27).

Далее в ячейки G7:G11 вводятся формулы для получения данных, которые позволят оценить отклонение dmax. Для рассматриваемого примера:

Ячейка	*Формула*
G7	=G6/G5
G8	=СТЬЮДРАСПОБР(0,05;G3-2)
G9	=СТЬЮДРАСПОБР(0,001;G3-2)
G10	=G8*КОРЕНЬ(G3-1)/(КОРЕНЬ(G8^2+(G3-2)))
G11	=G9*КОРЕНЬ(G3-1)/(КОРЕНЬ(G9^2+(G3-2)))

В ячейку F14 необходимо запрограммировать автоматический вывод результата оценивания dmax. С этой целью можно использовать логическую функцию ЕСЛИ(). Для рассматриваемого примера:

=ЕСЛИ(G7<G10;”ОШИБКИ НЕТ”;ЕСЛИ(G7>G11;”ОШИБКА!”;”МОЖЕТ БЫТЬ”))

	A	B	C	D	E	F	G	H	I	J	K	L
1		**ВЫБОРКА**		**ОТКЛОНЕНИЕ**		**ПРОВЕРКА**			**ВЫБОРОЧНЫЕ ХАРАКТЕРИСТИКИ**			
2		h, мм		|di|,мм		**ОДНОРОДНОСТИ**			**"ОПИСАТЕЛЬНАЯ СТАТИСТИКА"**			**Функции**
3		2,50		0,002		n	24		*h, мм*			***Excel***
4		2,47		0,027		Xср	2,498					
5		2,51		0,012		S	0,025		Среднее	2,4975	мм	2,50
6		2,50		0,002		dmax	0,047		Стандартная ошибка	0,00519	мм	0,005
7		2,45		0,047		τ_{max}	1,869		Медиана	2,5	мм	2,50
8		2,48		0,018		t [0,5;n-2]	2,074		Мода	2,5	мм	2,50
9		2,52		0,023		t [0,001;n-2]	3,792		Стандартное отклонение	0,02541	мм	0,025
10		2,49		0,007		τ [5;n–2]	1,939		Дисперсия выборки	0,00065	мм2	0,0006
11		2,50		0,002		τ [0,1;n–2]	3,015		Эксцесс	-0,50664	-	-0,507
12		2,47		0,027					Асимметричность	0,04553	-	0,046
13		2,46		0,038		**ОШИБКИ НЕТ**			Интервал	0,09	мм	0,09
14		2,51		0,012					Минимум	2,45	мм	2,45
15		2,54		0,043					Максимум	2,54	мм	2,54
16		2,50		0,002					Сумма	59,94	мм	59,94
17		2,46		0,038					Счет	24	-	24
18		2,53		0,032					Уровень надежности(95,0%)	0,01073	мм	0,011
19												
20		2,54		0,043					**СРАВНЕНИЕ**			
21		2,49		0,007					**ХАРАКТЕРИСТИК РАССЕЯНИЯ**			
22		2,49		0,007								
23		2,48		0,018					S*^2	0,00062	мм2	
24		2,50		0,002					S*	0,02487	мм	
25		2,54		0,043					Различие S*^2 и S^2	4,35	%	
26		2,51		0,012					Различие S* и S	2,11	%	
27		2,50		0,002								

Рис. 1.Пример оформления рабочего листа при анализе выборки

При получении результатов «ОШИБКА!» или «МОЖЕТ БЫТЬ» значения x_i и $\left| d_i \right|$, расположенные в строке, в которой |di|=dmax, необходимо удалить.

3.2.Определение выборочных характеристик

После обеспечения однородности выборки вычислить выборочные характеристики с применением инструмента «Описательная статистика» (описание работы с инструментом приведено в приложении). В примере на рис. 1 результаты работы инструмента представлены в ячейках I3:J18. Используя клавиатуру, их следует дополнить размерностями каждой из выборочных характеристик (ячейки К5:К18).

Затем вычислить каждую из характеристик с использованием функций *MS Excel* и полученные значения отформатировать с использованием формата ячейки типа «Числовой». В примере соответствующие значения представлены в ячейках L5:L18.

3.3.Сравнение смещенных и несмещенных характеристик рассеяния.

В заключение работы сравнить характеристики рассеяния s^2 и s с характеристиками $\bar{s}^2$ и $\bar{s}$. В примере выполнения работы (рис. 1) эти характеристики обозначены как S^2, S, S*^2 и S*, а соответствующие данные расположены в ячейках с J23 по J26. Значения s^2 и s определяются в ячейках J23 и J24 с использованием статистических функций ДИСПР(B3:B27) и СТАНДОТКЛОНП(B3:B27). Значения несмещенных оценок $\bar{s}^2$ и $\bar{s}$ расположены среди результатов работы инструмента «Описательная статистика» (в ячейках J10 и J9 соответственно). Различия между смещенными и несмещенными оценками дисперсии и среднего квадратического отклонения определены в ячейках J25 и J26. Для их вычисления здесь запрограммированы формулы:

= 100*(J10-J23)/J23 и = 100*(J9-J24)/J9.

4.СОДЕРЖАНИЕ ВЫВОДОВ ПО РАБОТЕ

В выводах по работе необходимо ответить на следующие вопросы:

1. Сколько элементов выборки и с какой целью были исключены?

2. Чему равно истинное значение случайной величины?
3. Наблюдаются ли асимметрия и эксцесс распределения случайной величины? Если наблюдаются, то о чем это свидетельствует?
4. Наблюдается ли различие между значениями смещенных и несмещенных оценок дисперсии и среднего квадратического отклонения? Чему они равны? Чем объясняется наличие или отсутствие различий?

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Львовский Е.Н, Статистические методы построения эмпирических формул: Учебн. пособие для втузов, - 2-е изд., перераб. и доп.- М.:Высш. шк . , 1988. - 239 с.
2. Математическая статистика / Иванова В. М., Калинина В. Н., Нешумова Л. А. и др. – М.:Высшая школа, 1981. – 371 с.
3. Четыркин Е. М., Калихман И. Л. Вероятность и статистика. – М.: Финансы и статистика, 1982. – 320 с.
4. Ахназарова С.Л,, Кафаров В.В. Организация эксперимента в химии и химической технологии.- М.: Высшая школа , 1978, - 319 с.
5. Замков О. О., Толстопятенко А. В., Черемных Ю. Н. Математические методы в экономике. – М.: МГУ им. М. В. Ломоносова, издательство «ДИС», 1997. – 368 с.
6. Колемаев В. А., Староверов О. В., Турундаевский В. Б. Теория вероятности и математическая статистика. – М.:Высш. шк., 1991. – 400 с.

Содержание

ВЫЧИСЛЕНИЕ ВЫБОРОЧНЫХ ХАРАКТЕРИСТИК В *MS EXCEL*

В *MS Excel* для вычисления выборочных характеристик имеется набор *статистических функций*. Весь комплекс выборочных характеристик можно получить с использованием инструмента «Описательная статистика», входящий в пакет «Анализ данных», который относится к *надстройкам*, включенным в комплекс *MS Excel*.

Запуск инструмента «Описательная статистика» осуществляется последовательным выбором пунктов из меню различных уровней:

<**Сервис**> **/** <**Анализ данных**> **/** <**Описательная статистика**>.

На экране откроется диалоговое окно (рис. П.1).

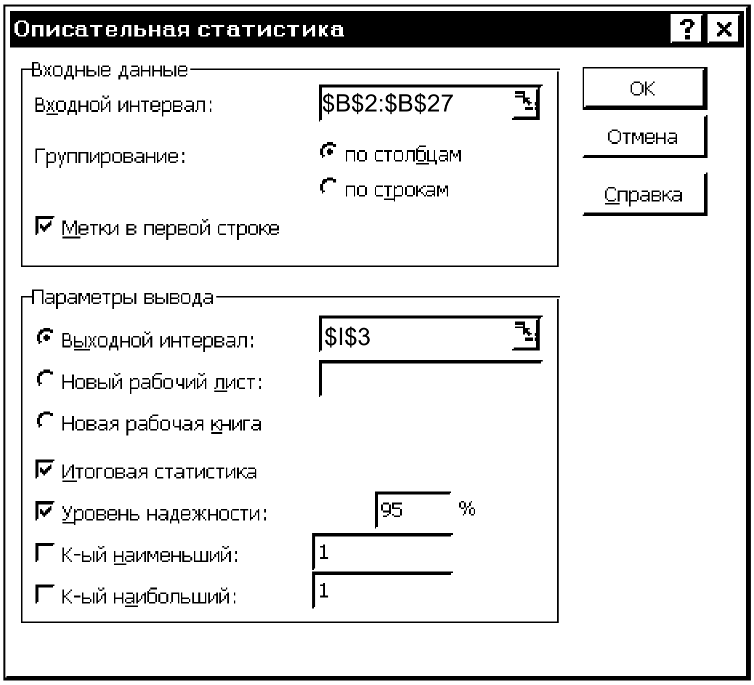

Рис. П1. Диалоговое окно инструмента «Описательная статистика»

Для использования инструмента «Описательная статистика» сначала необходимо охарактеризовать исходные (входные) данные и задать параметры, управляющие выводом результатов.

Входной диапазон. Область рабочего листа, содержащая анализируемые данные. На рис. П.1 задан входной диапазон для данных, пример обработки которых приведен на рис. 1.

Группирование. Указатель особенностей расположения данных во входном диапазоне. Если данные сгруппированы в столбцы, установите переключатель в положение «По столбцам». Если данные сгруппированы в строки, установите переключатель в положение «По строкам».

Метки в первом столбце / Метки в первой строке. Указатель наличия условного обозначения анализируемой величины среди обрабатываемых данных. Если эта опция задана, то данное, записанное в первой ячейке входного диапазона, воспринимается как условное обозначение и при анализе не учитывается. Например, при группировке данных по столбцам условным обозначением будет считаться данное, расположенное в первой строке каждого из них. При группировке по строкам в качестве условного обозначения будут восприниматься данные, расположенные в первом столбце каждой из них. **Внимание!** Если анализируемые данные не снабжены условным обозначением, то задание опции «*Метки …* » приведет к потере данного из первой ячейки входного диапазона, что повлечет ошибку в расчете выборочных параметров.

Выходной диапазон, Новый лист, Новая книга. Указатели расположения результатов работы инструмента «Описательная статистика». При выборе опции *Выходной диапазон* результаты выводятся на активный рабочий лист активной книги *MS Excel*. Они будут размещены слева направо и сверху вниз, начиная с ячейки, адрес которой задан в окне (в примере из прил. 1 - начиная с ячейки I3). При выборе опций *Новый лист* или *Новая книга* результаты выводятся в новый рабочий лист активной книги или в первый лист другой книги *MS Excel*. В обоих случаях результаты будут размещены начиная с ячейки A1.

Итоговая статистика. Указатель, с помощью которого задается команда на вывод минимального набора характеристик выборки. В этот набор входят: *Среднее, Стандартная ошибка (среднего), Медиана, Мода, Стандартное отклонение, Дисперсия выборки, Эксцесс, Асимметричность, Интервал, Минимум, Мак-*

симум, Сумма, Счет. Наименования некоторых из перечисленных характеристик отличаются от принятых в отечественной практике обработки и анализа числовой информации. Их смысл поясняется ниже:

Наименование в *MS Excel*	**Обозначение**	**Формула**	**Расчет в *MS Excel***
Среднее	$\overline{x}$	(1)	=СРЗНАЧ(B3:B27)
Стандартная ошибка	s_x	(18)	=J8/КОРЕНЬ(J16)
Медиана	Me	(2), (3)	=МЕДИАНА(B3:B27)
Мода	Mo	-	=МОДА(B3:B27)
Стандартное отклонение	$\overline{s}$	(8)	=СТАНДОТКЛОН(B3:B27)
Дисперсия выборки	$\overline{s}^2$	(6)	=ДИСП(B3:B27)
Эксцесс	E	(12)	=ЭКСЦЕСС(B3:B27)
Асимметричность	A	(11)	=СКОС(B3:B27)
Интервал	R	(9)	=J14-J13
Минимум	x_{min}	-	=МИН(B3:B27)
Максимум	x_{max}	-	=МАКС(B3:B27)
Сумма	$\sum_{i=1}^{n} x_i$	-	=СУММ(B3:B27)
Счет	n	-	=СЧЁТ(B3:B27)
Наибольший(2)	-	-	=НАИБОЛЬШИЙ(B3:B27;2)
Наименьший(2)	-	-	=НАИМЕНЬШИЙ(B3:B27;2)
Уровень надежности (95%)	Δ_x	(19)	=СТЬЮДРАСПОБР(0,05;J16-1)*J5

Уровень надежности. Указатель, с помощью которого к характеристикам выборки, входящим в набор *Итоговая статистика*, добавляется значение доверительных границ для среднего выборочного. Доверительные границы вычисляются при доверительной вероятности, заданной в окне рассматриваемой опции (в примере из прил. 1 задано p = 95%).

К-ый наибольший, К-ый наименьший. Указатели, с помощью которых к характеристикам итоговой статистики добавляются, соответственно, k-е наибольшее и наименьшее значение из выборки.