Анализ экспериментальных данных

компьютерный лабораторный практикум
для студентов I курса специальности «нанотехнологии и микросистемная техника»

ПЗ № 2. Свойства выборок случайных чисел. Обнаружение промахов

Разработал д.т.н., проф. В.А. Годлевский

2.1. Формирование рабочего поля

Подготовим рабочее поле в Excel для выполнения следующих упражнений. Удобно будет, если мы для текущего задания выделим новый лист книги Excel

1) Внизу слева на рабочем листе таблицы рядом с надписью **Лист 1** нажмем на знак **+** . Рядом появится Новый лист таблицы **Лист 2**. Для большей информативности переименуем листы в названия практических занятий:

Лист 1 → **ПЗ 1**

Лист 2 → **ПЗ 2**

Получим вот что

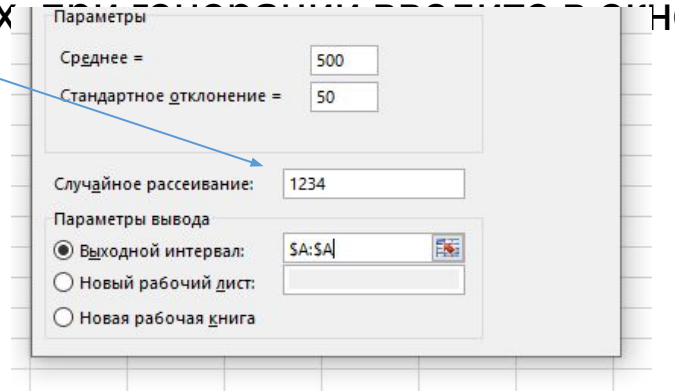
2) Открываем лист ПЗ 2. Нам понадобятся опять два столбца со случайными числами.

Давайте сгенерируем эти столбцы по тем же индивидуальным параметрам, что и на прошлом занятии.

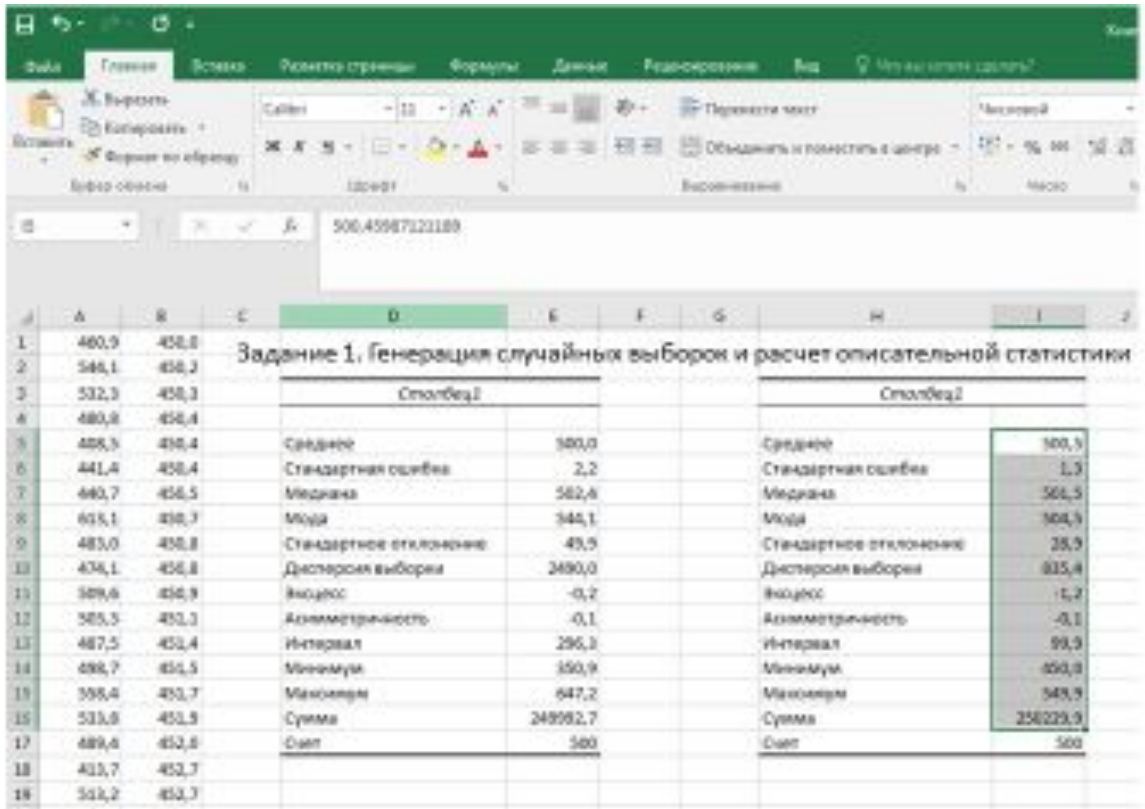
Столбец А с нормальным законом, В – с равномерным. Заодно, закрепим навыки генерации данных.

Чтобы столбцы отличались от предыдущих при генерации вводите в окно **Случайное рассеяние** любое целое число. Скажем, 12345

Данные столбцов округлим до десятых и отсортируем по возрастанию



3) Вспомним еще, как строить таблицы описательной статистики: разместим две таблицы рядом со столбцами это будет выглядеть как-то так.



Задание 1. Генерация случайных выборок и расчет описательной статистики

	Столбец2			Столбец2	
	Среднее	300,0		Среднее	300,5
	Стандартная ошибка	2,2		Стандартная ошибка	1,3
	Медиана	582,6		Медиана	566,5
	Мода	544,1		Мода	568,9
	Стандартное отклонение	49,9		Стандартное отклонение	28,9
	Дисперсия выборки	2490,0		Дисперсия выборки	835,4
	Эксцесс	-0,2		Эксцесс	-1,2
	Асимметричность	-0,1		Асимметричность	-0,1
	Интервал	296,3		Интервал	99,9
	Минимум	150,9		Минимум	450,0
	Максимум	647,2		Максимум	549,9
	Сумма	248982,7		Сумма	258229,9
	Счет	500		Счет	500

4) Теперь получим новый параметр. Используя данные таблиц, рассчитаем для наших выборок **ВАРИАЦИЮ.**

$v = \frac{\sigma}{\bar{x}}$, где σ – стандартное отклонение, $\bar{x}$ - среднее

Результаты вычислений записываем под таблицами в таким виде

$$v_A = \frac{\sigma_A}{\bar{x}_A} \cdot 100\% = \frac{45,4}{340,5} = 13,3\%; \qquad v_B = \frac{\sigma_B}{\bar{x}_B} \cdot 100\% = \frac{38,9}{300,5} = 12,7\%;$$

Если вариация больше 5%, точность результата обычно признают неудовлетворительной. Запишем соответствующий вывод.

5) Проверка нормальности выборки по асимметрии и эксцессу

Чтобы ответить на вопрос о том, можно ли признать выборку нормальной, сравнивают выборочные значения

асимметрии и эксцесса (из таблицы описательной статистики) с их ожидаемыми значениями U_A и U_E .

$$U_A = \sqrt{\frac{6(N-1)}{(N+1)(N+3)}} \qquad U_E = \sqrt{\frac{24N(N-2)(N-3))}{(N+1)^2(N+3)(N+5)}}$$

При N = 500 Ua = 0,109; Ue = 0,21

Вычисляем U_A и U_E для вашего N, заполняем таблицу, делаем выводы

Анализируе-мая выборка	Асим-метрия A (табли-чная)	Ожида-емая асим-метрия U_A	Вывод по асимметрии	Эксцесс, E (табли-чный)	Ожида-емый эксцес с U_E	Вывод по эксцессу
A	-0,252*	0,109	$A < U_A$ Признается нормальным	1,500	0,332	$E >> U_E$ Не признается нормальным
B	0, 112	0,205	$A < U_A$ Признается нормальным	0,020	0,21	$E >> U_E$ Признается нормальным

7) Отбраковывание (цензурирование) выпадающих значений по правилу 3 σ

По этому критерию отбрасывают подозрительные значения x*, лежащие за пределами интервала $\bar{x} \pm 3\sigma$. Это правило применяют в случае, если N < 6. Для выборок большего объема вероятность появления выпадающих результатов возрастает, и тогда рекомендуется расширять интервал цензурирования. Ширину интервала для отбраковки выпадающих результатов можно брать из таблицы.

Поскольку в наших индивидуальных вариантах объемы выборок находятся в диапазоне N = 101…1000, применяем для цезурирования границы отбрасывания $\bar{x} \pm 4{,}5\ \sigma$.

Проверим по этому правилу крайние значения наших выборок:

A) Для выборки A : X_{amin} и X_{amax}

B) Для выборки B: X_{bmin} и X_{bmax}

Результаты проверки сводим в таблицу

N	Интервал цензурирования
< 6	$\lvert x* - \bar{x} \rvert < 3\sigma$
6…100	$\lvert x* - \bar{x} \rvert < 4\sigma$
101…1000	$\lvert x* - \bar{x} \rvert < 4.5\sigma$
1001…10000	$\lvert x* - \bar{x} \rvert < 5\sigma$

Результаты цензурирования выборок А и В по границам диапазона $\bar{x} \pm 4{,}5\ \sigma$.

Выборка	$\bar{x}$	Крайние значения		σ	4,5 σ	$\bar{x}$ + 4,5σ.	$\bar{x}$ - 4,5σ.	Выводы о наличии выпадающих значений
		Min	Max					
A	500,42	316,59		49,96	224,82		275,6	Нет
			648,75			725,24		Нет
B	500,76	450,01		28,99	130,45		370,31	Нет
			549,85			631,21		Нет

Вывод: В двух проверенных выборках выпадающих крайних значений не обнаружено

6) **Расчеты доверительных интервалов**

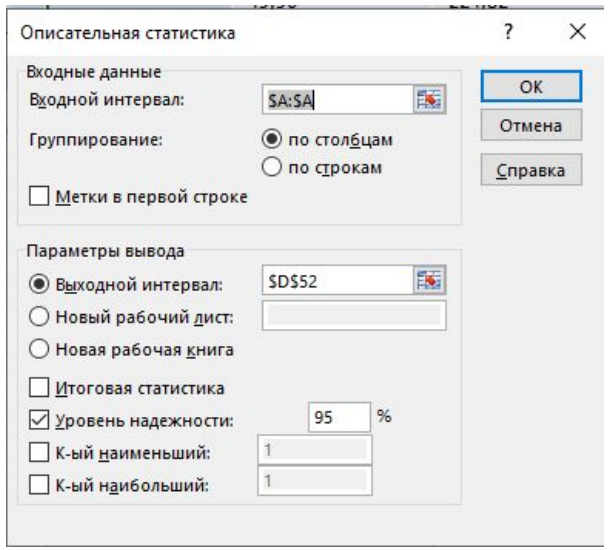
Для нормально распределенной выборки можно рассчитывать доверительные интервалы.

Доверительным называют симметричный интервал вида $\bar{x} \pm \Delta$, для которого указана вероятность попадания отсчета в этот интервал. Эту вероятность называют *доверительным коэффициентом* или *коэффициентом надежности*. В рассчитанном нами наборе параметров описательной статистики присутствует полуширина доверительного интервала Δ. Она дана в графе таблицы **Уровень надежности** (ρ), где ρ — доверительный коэффициент. По умолчанию расчет производят для ρ = 0,95.

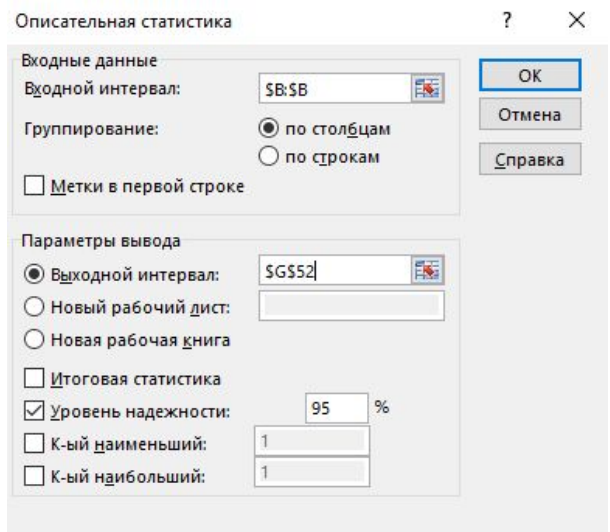
Ширина доверительного интервала связана со значением доверительного коэффициента: чем больше ρ, тем шире доверительный интервал. Если речь идет об повторных измерениях некоторой величины, то можно утверждать, что *чем выше надежность измерения, тем ниже его точность.*

Проверим это утверждение на наших выборках. Выполним процедуру **Описательная статистика** для столбцов A и B при следующих установках

Для столбца A



Для столбца B



Далее таким же образом, с помощью сокращенных таблиц описательной статистики, рассчитаем ширины доверительных интервалов для ρ = 99% и 99,9%. Данные сведем в таблицу

	Среднее	Стандартное отклонение	Полуширина доверительного интервала Δ при доверительной вероятности ρ		
			0,95	0,99	0,999
Столбец А	500	50	4,39	5, 77	7,40
Столбец Б	500	30	2,55	3,35	4,29

Выводы: 1) При повышении уровня доверительной вероятности ширина доверительного интервала увеличивается

2) Для большинства измерений принято принимать доверительную вероятность ρ = 95%.

2) Результат измерения любой величины должен быть представлен тремя числами,

например: $x = 500 \pm 50$ ($\rho = 95\%$)

7. **Проверка подозрительного значения по критерию Шовене**

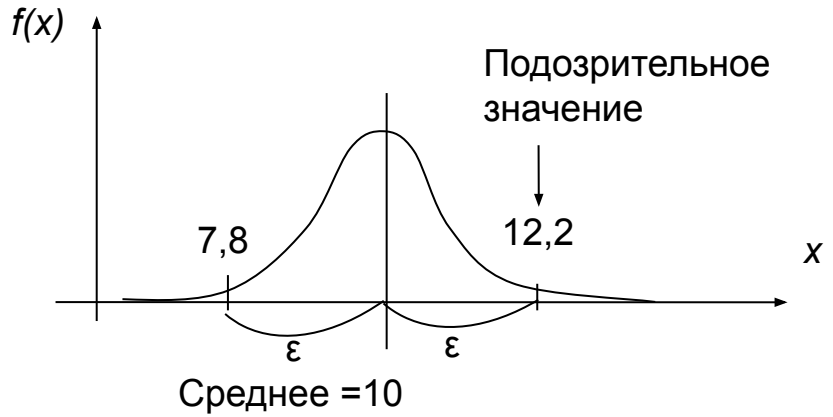
Для того, чтобы проверить подозрительное значение x* с помощью критерия Шовене, требуется рассчитать вероятность P появления столь же плохого или худшего, чем x* результата при объеме выборки N.

Затем эту вероятность сравнивают с критерием Шовене - 0.5. И если P < 0,5, то ∞членов выборки.

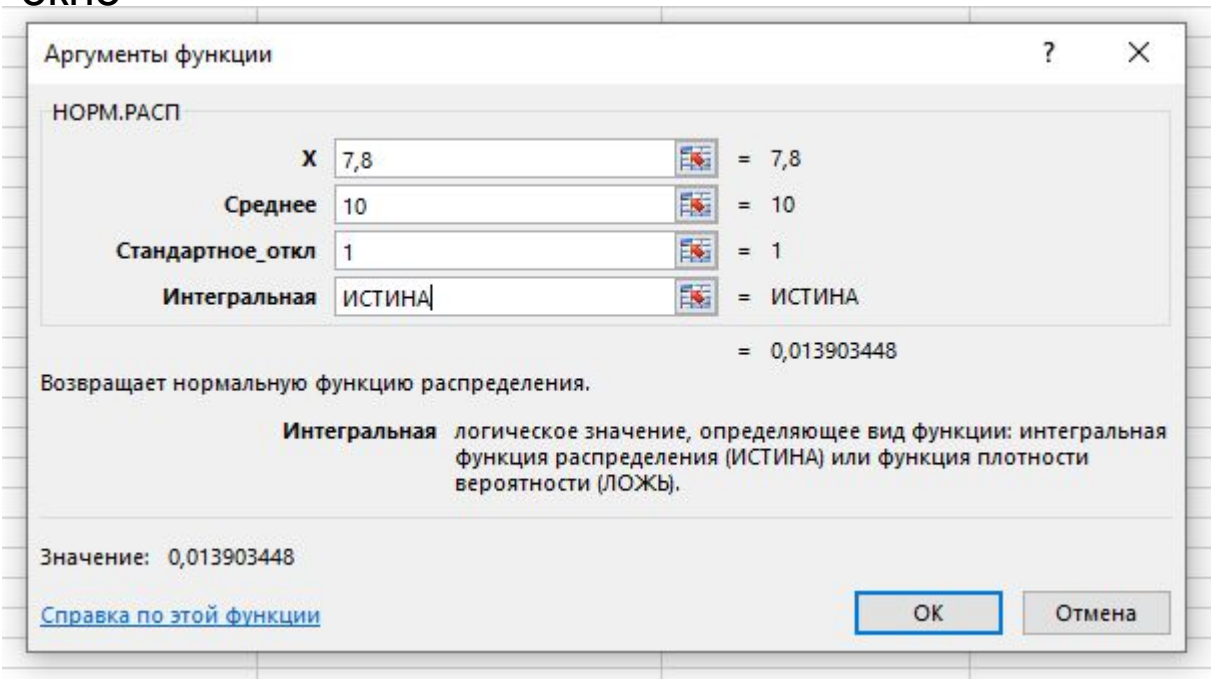
Пример. Пусть дана выборка с параметрами N = 5; $\bar{x} = 10$, σ = 1 .
Один из членов выборки x* = 12,2 является подозрительным. Требуется определить, не является ли число 12,2 промахом

1. Определяем границы интервала для подозрительного значения .
   Это будет симметричный интервал 7.8…12.2 (см. рисунок)

2. Вычисляют интеграл вероятности для интервала - ∞ …7.8. Для этого в Excel вызывают функцию НОРМ.РАСП Нужно зайти в библиотеку функций, раздел **Статистические**



Получим такое диалоговое окно



В нашем случае в диалоговое окно вводят:
- в графу **X** — нижнюю границу интервала: 7,8.
- в графы **Среднее** и **Стандартное отклонение** — параметры нашей выборки ;
- в графу **Интегральная** — ИСТИНА

Получаем в результате расчета вероятность того, что любое последующее измерение попадет в интервал
- ∞ …7.8 (то есть вычислена площадь под левым «хвостом» кривой распределения.

Эта вероятность равна $P_1 = 0.014$.

3. Поскольку нам нужна «двухсторонняя» вероятность того, что измерение будет *такое же или худшее, чем наш подозрительный результат*, вычислим эту вероятность $P_2 = P_1 * 2 = 0.028$. (Это площадь под двумя «хвостами кривой распределения).

4. Вычисляем величину критерия Шовене, чтобы определить, какова вероятность появления такого же или худшего результата во всей серии из 5 измерений. Эта величина составит P5 = P2*5 = 0.028*5 = 0.14:

5. Поскольку эта величина меньше, чем критериальное значение 0,5, **данное подозрительное значение 12, 2 признаем промахом и удаляем его из выборки.**

Решаем задачу по отбраковке подозрительного значения

Индивидуальные варианты для решения задачи по применению критерия Шовене приведены ниже в таблице

Исходные данные для индивидуальных заданий

№	ФИО		Объем выборки N	Стандартное отклонение, σ	Подозрительное значение X*
1	Нефедов	15	5	1,5	8,2
2	Уханова	17	7	1,7	22,5
3	Котомина	19	9	1,9	27
4	Акыев	22	12	2,0	5
5	Васильева	25	15	2,5	33
6	Джумадурдыева Акнабат	27	17	2,75	26
7	Карпов	30	13	3,0	23
8	Розыев	32	14	3,2	42
10	Хыдыров	37	18	3,7	50
11		40	20	4,0	56