

Лабораторная работа №1

Статистическая обработка вариационного ряда

Цель работы: изучить методику статистической обработки вариационного ряда, освоить практические приемы статистической обработки вариационного ряда средствами пакета «Microsoft Excel».

1. Теоретические сведения


Изучение медицинских явлений и поиск присущих им закономерностей, как правило, связано с повторением однородных наблюдений или опытов. При этом исследователя интересуют не отдельные наблюдения, а их обобщенные характеристики, помогающие понять типичные черты изучаемых явлений. Анализируя результаты нескольких серий наблюдений или опытов, исследователь обнаруживает различия в частоте интересующих его признаков, если эти признаки качественные, либо в величине признаков, если их можно оценить количественно. Разброс значений признака называется варьированием. Результат отдельного измерения называется вариантой и обозначается – x .



Во всех случаях обнаружения варьирования исследователю необходимо выяснить, насколько существенно варьирование, случайно оно или нет и каковы факторы, его определяющие. Для решения этих задач необходимо составить вариационный ряд и вычислить его обобщенные характеристики.

Методика составления вариационного ряда. Если число наблюдений (n) небольшое, то варианты достаточно просто ранжировать, т. е. расположить в порядке возрастания их значений. Например, при измерении размеров вируса орнитоза получены следующие величины (в $\mu\text{м}$): 0,34; 0,45; 0,20; 0,29; 0,40. Эти варианты нужно записать в такой последовательности: 0,20; 0,29; 0,34; 0,40; 0,45.

При увеличении числа наблюдений обычно отмечаются повторения отдельных вариантов. В этом случае для построения вариационного ряда необходимо выписать все значения вариантов в порядке возрастания, а затем подсчитать число повторений (частоту – f) каждой варианты и записать их рядом с соответствующими значениями вариантов. Например, исследователем произведено 47 измерений мембранного потенциала мышечной клетки в покое (с точностью до 1 мВ). Составленный вариационный ряд показан в табл. 1



Таким образом, главными составными элементами вариационного ряда являются:

x – варианты - значения варьирующего признака;

f – частоты - число повторений каждой варианты;

n – общее число наблюдений (n равно сумме частот, т. е. $n = \sum f$).

Последовательное суммирование частот образует так называемые накопленные частоты. Последняя накопленная частота представляет собой общее

число наблюдений. Подобным же образом составляется и интервальный вариационный ряд, в котором перечисляются не отдельные варианты, а их группы. Интервальный вариационный ряд следует составлять в тех случаях, когда исследователь имеет дело с большим разнообразием значений вариантов (более 20). Интервалы в таком вариационном ряду целесообразно иметь одинаковыми, т. е. они должны объединять равное число значений вариантов. Интервальные вариационные ряды строятся при изучении как дискретных величин (признаков, выражаемых только целым числом, например число посещений, операций, число эритроцитов, частота пульса и т. д.) так и при исследовании непрерывных величин (признаков, регистрируемых дробными числами, например, рост, вес, биохимические показатели т. п.).

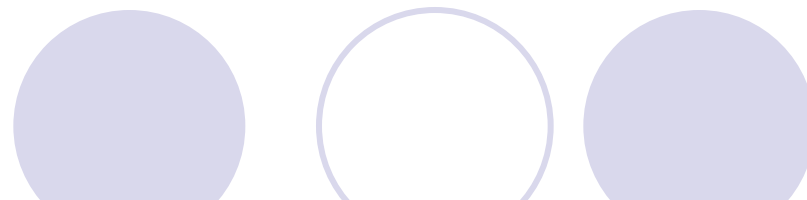



Таблица 1.1 – Результаты измерения потенциала мышечной клетки

Варианта x	Частота f	Накопленные частоты
33	1	1
34	2	3
35	4	7
36	5	12
37	8	20
38	10	30
39	7	37
40	6	43
41	3	46
42	1	47
	$n=47$	



Для графического изображения вариационного ряда применяют полигоны и гистограммы (рис. 1.1.). Полигоны используют для изображения рядов дискретных величин, а гистограммы — непрерывных. При построении полигона на оси абсцисс откладывают значения вариант или их групп, на оси ординат — частоты. Полученные точки соединяют прямыми линиями. При построении гистограммы на оси абсцисс восстанавливают столбики, по высоте соответствующие частотам взятых интервалов, а вся гистограмма приобретает вид суммы прямоугольников.

Графическое изображение вариационного ряда дает ориентировочное представление о законе, которому подчиняется повторяемость вариант, так называемом законе распределения.

Знание закона распределения варьирующих признаков или достаточно достоверное предположение о нем дают возможность исследователю выбрать наиболее правильный и эффективный метод для статистической характеристики имеющихся наблюдений. Если исследуются непрерывные случайные величины и ряд на графике выглядит одновершинной симметричной кривой, то можно предположить, что изучаемые величины подчиняются нормальному закону распределения (см. рис. 1.2.).

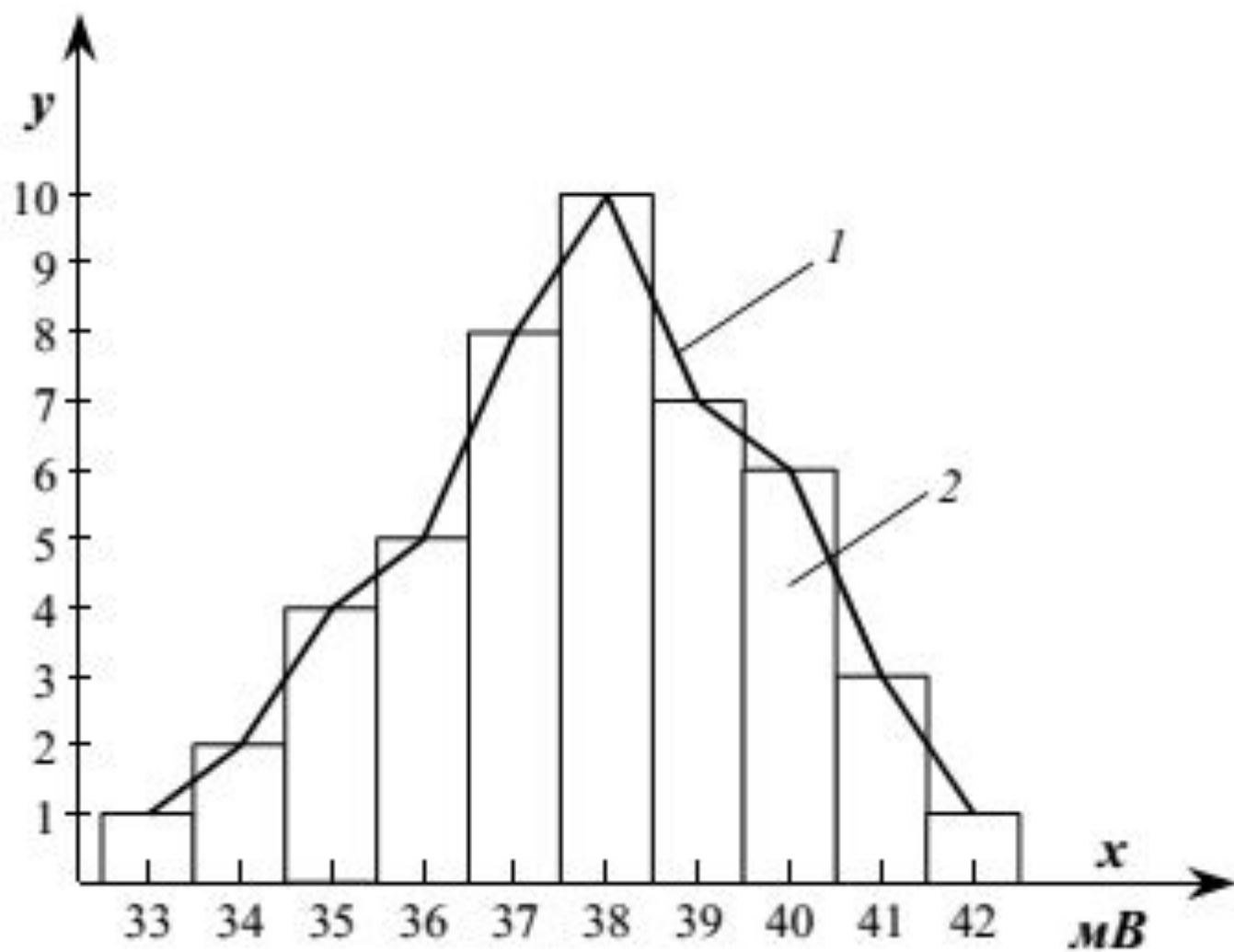
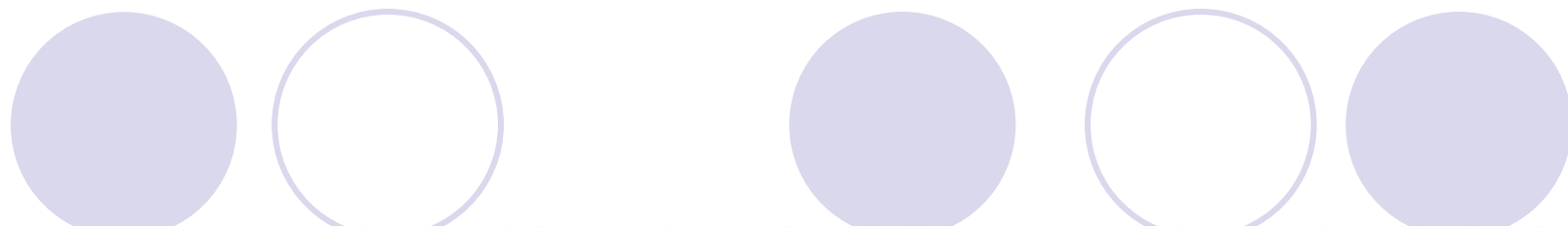


Рис. 1.1. Полигон (1) и гистограмма (2) распределения

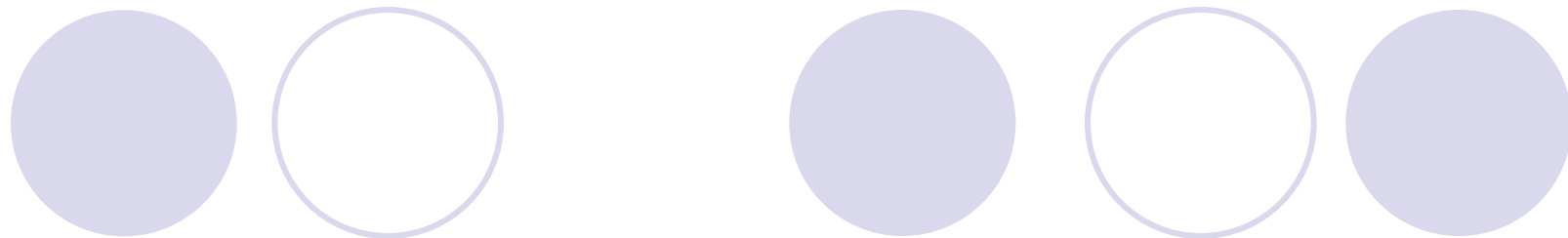


Сводными характеристиками значений вариант служат средняя арифметическая величина, мода, медиана и квартили. Каждая из этих характеристик своеобразна. Они не могут подменить друг друга и лишь в совокупности достаточно полно и в сжатой форме представляют особенности вариационного ряда. Наиболее общей характеристикой всех значений вариант является *средняя арифметическая величина*. Различают среднюю арифметическую простую и взвешенную. Средняя арифметическая простая вычисляется по формуле:

$$\bar{x} = \frac{\sum x}{n} \quad (1.1)$$

В вариационных рядах, где отдельные варианты встречаются с разной частотой (т.е. имеют разный вес) определяется средняя арифметическая взвешенная по формуле:

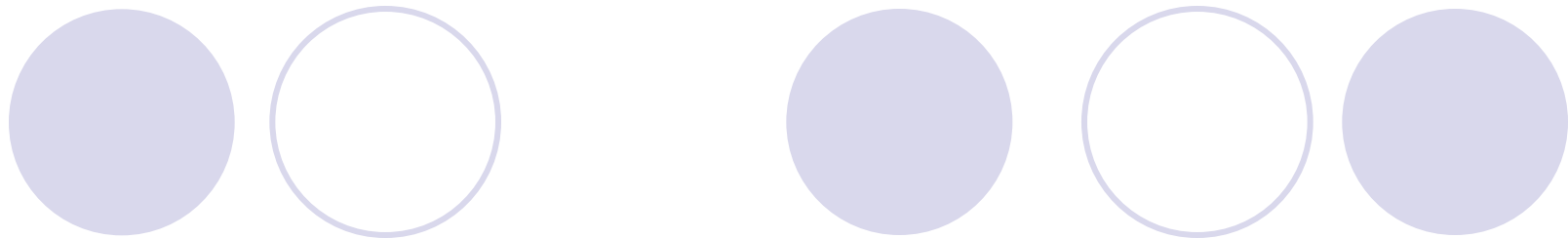
$$\bar{x} = \frac{\sum x \cdot f}{n} \quad (1.2)$$



Как видно из формулы, на величине средней арифметической сказывается влияние всех вариантов входящих в вариационный ряд, причем это влияние прямо пропорционально числу повторений вариантов. Взвешенную среднюю арифметическую величину необходимо вычислять во всех случаях, когда частоты не одинаковы.

В интервальных вариационных рядах при определении средней арифметической величины прежде всего следует определить середины интервалов. Середину интервала при изучении непрерывных величин можно определить как среднюю арифметическую начальных значений двух соседних интервалов. В дискретных рядах середина интервала вычисляется как среднее арифметическое начального и конечного значений данного интервала. Затем значения середин интервалов используют при дальнейших расчетах в качестве вариантов x .

Средняя арифметическая величина обладает следующими свойствами:



1) сумма отклонений от средней равна нулю;

2) при умножении (делении) всех вариантов на один и тот же множитель (делитель) средняя арифметическая умножается (делится) на тот же множитель (делитель);

3) если прибавить (вычесть) ко всем вариантам одно и то же число, средняя увеличится (уменьшится) на то же число.

Эти свойства могут быть использованы для облегчения и упрощения расчета средней арифметической величины.

Первое свойство, например, служит обоснованием расчета средней арифметической по способу моментов:



$$\bar{x} = A + \frac{\sum (x - A) \cdot f}{n} \quad (1.3)$$

где:

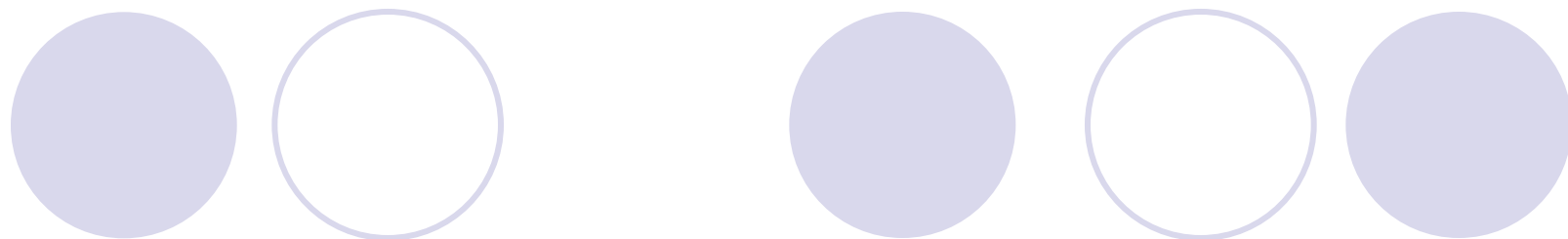
x – середины интервалов вариационного ряда;

A – условная средняя.



Особенно удобно способ моментов использовать при вычислении средней арифметической в интервальном вариационном ряду. Для этого необходимо сначала определить середины интервалов. Величину одной из середин интервала, лучше наиболее часто встречающуюся, следует принять за условную среднюю (A), после чего найти отклонения всех других середин интервалов от этой величины $x - A$. Полученные разности затем необходимо умножить на соответствующие частоты, произведения суммировать и подставить найденную величину $\sum(x - A) \cdot f$ в формулу для вычисления (1.3).

Второе свойство средней арифметической полезно применить при анализе вариационного ряда, состоящего либо из очень больших, либо из очень малых величин. Имеются, например, варианты: 0,0001; 0,0002; 0,0003. Используя это свойство, увеличим их в 10 000 раз, получим величины 1, 2, 3. Средняя арифметическая из них равна 2, а искомая средняя арифметическая в 10000 меньше, т. е. 0,0002.



Модой (M_0) называют значение наиболее часто встречающейся варианты. В примере в табл. 1 это варианта 38 мВ. В интервальном вариационном ряду мода находится как середина того интервала, которому соответствует наибольшая частота.

Более точно мода определяется по формуле:

$$M_0 = X_{M_0} + \Delta \frac{f_{M_0} - f_{M_0 - 1}}{(f_{M_0} - f_{M_0 - 1}) + (f_{M_0} - f_{M_0 + 1})} \quad (1.4)$$

где:

X_{M_0} – начальное значение интервала, содержащего моду;

Δ – ширина интервала;

f_{M_0} – частота вариант в интервале, содержащем моду;

$f_{M_0 - 1}$ и $f_{M_0 + 1}$ – частоты вариант в соседних интервалах.



Как указывалось выше, кривая нормального распределения симметричная и одновершинная. Следовательно, в таком вариационном ряду может быть только одна мода. Если при анализе явления, которое предположительно подчиняется закону нормального распределения, получена, например, несимметричная, двухвершинная (бимодальная) кривая, то следует еще раз проанализировать состав исследуемой группы и, исключив искажающие наблюдения, сделать группу однородной.

Медиана (Me) — значение варианты, делящей вариационный ряд пополам (с каждой стороны от медианы находится половина вариантов).



Квартили (верхний – Q_3 и нижний – Q_1) — значения вариантов, делящих вариационный ряд (вместе с Me) на 4 части. Между Q_1 и Q_3 находится половина всех вариантов. Порядковый номер варианты, являющейся медианой или квартилем, определяется по формулам:

$$Q_1: (n+1) / 4; \quad Me: (n+1) / 2; \quad Q_3: 3 \cdot (n+1) / 4; \quad (1.5)$$

В случае получения дробного значения порядкового номера его округляют до ближайшего целого числа.

Более точный расчет медианы в интервальном вариационном ряду следует производить по формуле:

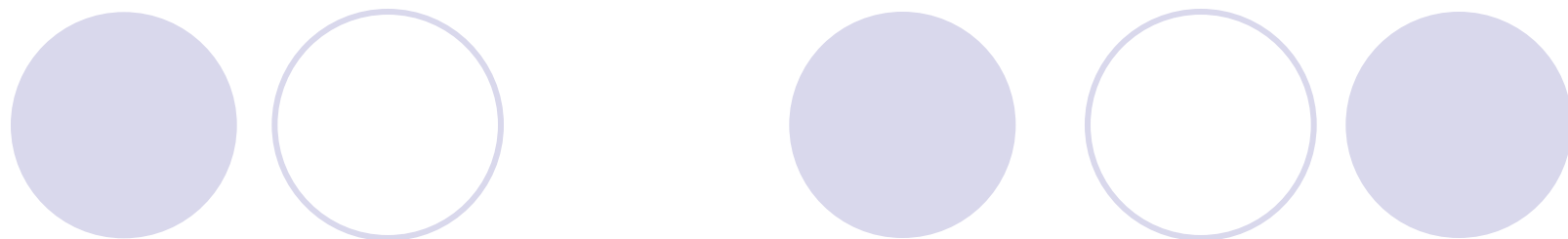
$$Me = X_{Me} + \Delta \frac{n/2 - S_{Me-1}}{f_{Me}}, \quad (1.6)$$

где X_{Me} – начальное значение интервала, содержащего медиану;

Δ – ширина интервала;

S_{Me-1} – накопленная частота до интервала, содержащего медиану;


f_{Me} – частота вариантов в интервале, содержащем медиану.



Размеры Mo и Me не зависят от значений крайних вариантов. В симметричном вариационном ряду они равны между собой и совпадают со значением средней арифметической. Мода особенно важна для характеристики несимметричного ряда. Медианой и квантилями обязательно нужно пользоваться при обработке ряда с открытыми крайними интервалами.

После определения обобщенных характеристик вариационного ряда следует установить его *колеблемость*, т.е. размеры варьирования значений изучаемого признака. Приблизительно о колеблемости можно судить по амплитуде (размаху) вариационного ряда-разности максимальной и минимальной вариант. Более точно колеблемость ряда характеризует среднее квадратическое отклонение (σ), вычисляемое по формуле:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{n}} \quad (1.7)$$



Квадрат среднего квадратического отклонения (σ^2) называется дисперсией. Небольшая величина среднего квадратического отклонения свидетельствует об однородности исследуемой группы наблюдений. Среднюю арифметическую в таком случае следует признать вполне характерной, типичной для данного вариационного ряда. Однако слишком малая величина σ заставляет думать об искусственном подборе наблюдений. При очень большой σ средняя арифметическая в меньшей степени характеризует весь вариационный ряд, что говорит о значительной вариабельности явления или неоднородности исследуемой группы.

Оценка степени рассеяния вариантов около средней может быть произведена с помощью коэффициента вариации, вычисляемого по формуле: $c = \frac{\sigma}{x} \cdot 100\%$

Значения коэффициента вариации менее 10% свидетельствуют о малом рассеянии, от 10 до 20% – о среднем и более 20% – о сильном рассеянии вариант вокруг средней арифметической.

Согласно теории вероятностей в явлениях, подчиняющихся нормальному закону распределения, между значениями средней арифметической, среднего квадратического отклонения и вариантами существует строгая зависимость. Например, 68,3% значений варьирующего признака находятся в пределах $\bar{x} \pm 1\sigma$; 95,5%—в пределах $\bar{x} \pm 2\sigma$ и 99,7%—в пределах $\bar{x} \pm 3\sigma$. Эти соотношения показаны на рис. 1.2. Указанные взаимоотношения средней арифметической, среднего квадратического отклонения и отдельных вариантов иногда называют правилом трех сигм. С помощью этого правила, зная \bar{x} и σ (и предполагая нормальным изучаемое распределение), можно получить представление о вероятных размерах варьирующего признака.

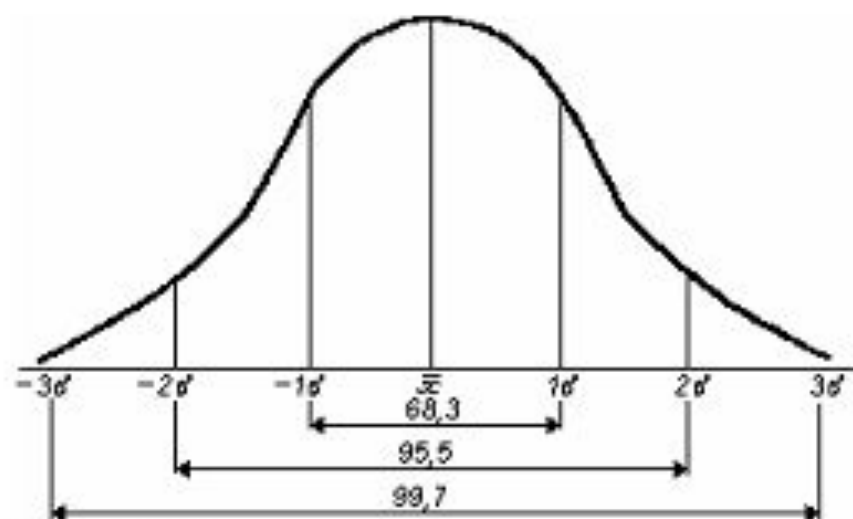
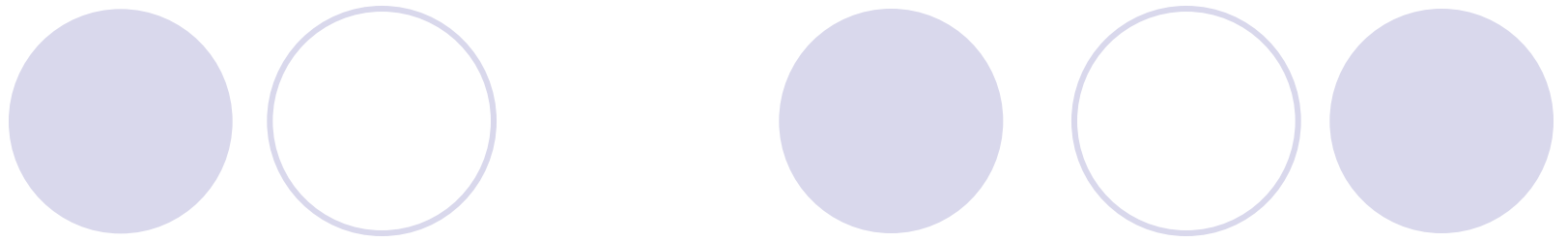



Рис. 1.2. Кривая нормального распределения



Правило трех сигм можно использовать при решении ряда практических задач:


1. Знание значений \bar{x} и σ дает исследователю возможность определить границы средних (нормальных) значений признака. Нормальными обычно рекомендуется считать значения в пределах $\bar{x} \pm 1\sigma$. Иногда пределы нормы определяют с использованием $0,5\sigma$; $1,34\sigma$ и т.п. Решать этот вопрос должен специалист, знающий существо исследуемого явления.

2. Нормированное отклонение $t = \frac{x - \bar{x}}{\sigma}$, позволяет также решить, относится ли данное наблюдение к интересующей нас совокупности. Ответ будет положительным всегда, когда $t < 3$.



2. Порядок выполнения работы

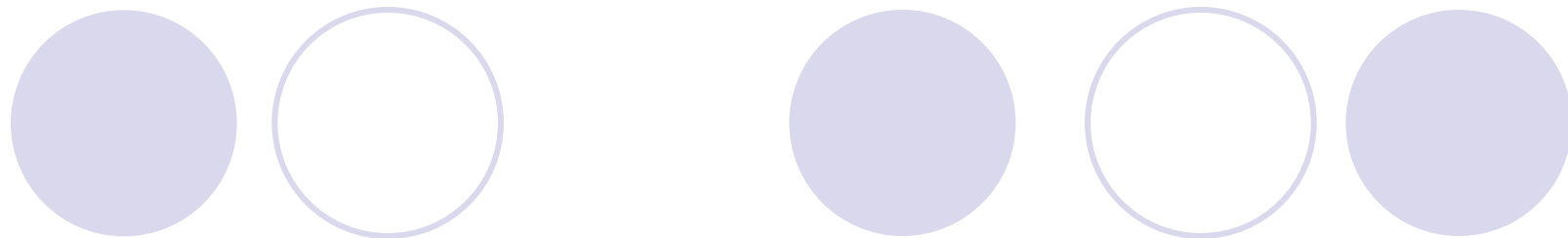
- 2.1. Изучить теоретический материал;
- 2.2. Загрузить файл **Образцы_стат.xls** с исходными данными;
- 2.3. Выполнить практические задания по проверке гипотез;
- 2.4. Получить распечатку результатов выполнения работы;
- 2.5. Оформить отчет по лабораторной работе.






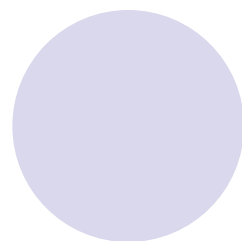
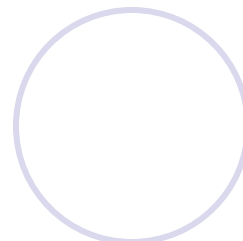
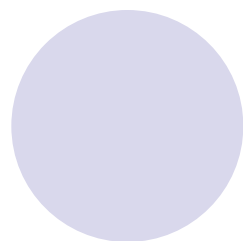
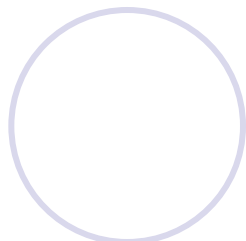
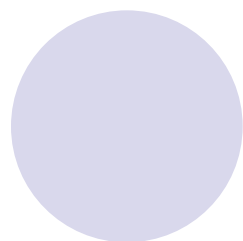
3. Практические задания по статистической обработке вариационного ряда

Задание 1. Проведите анализ данных в рамках описательной статистики с использованием средств **Вставка функций** и **Мастер диаграмм MS Excel**.

1. Запустите MS Excel: **Пуск / Программы /Microsoft Excel** и сохраните созданную при запуске книгу под именем **Примеры_стат** в вашу рабочую папку: **Файл / Сохранить как /** введите в поле **Имя файла** название книги **Примеры_стат**.




2. Переименуйте ярлычок рабочего листа **Лист 1**: двойной щелчок по ярлычку и напечатайте поверх выделения **Статистика_1** / скопируйте из файла **Образцы _стат.xls** с листа **Образец 1_1** исходные данные и оформите таблицу по **Образцу 1** (рис. 1.3.): выделите ячейку **A1** щелчком мыши / введите текст заголовка по образцу и зафиксируйте щелчком по инструменту **Enter**  / расположите заголовок по центру столбцов **A-E**: выделите ячейки **A1:E1** / инструмент **Объединить и поместить в центре**  / отформатируйте заголовок нижней границей при помощи инструмента **Границы**  / аналогично введите и оформите заголовок к таблице со статистикой / введите исходные данные, заголовки строк статистической таблицы и число выборок.






Основные статистические показатели и меры рассеяния

	A	B	C	D	E	F	G	H
1	x, прирост населения в 50 городах						x	
2	27	36	34	46	34		Среднее	
3	28	29	37	41	43		Среднеквадратичное откл.	
4	40	33	50	37	41		Дисперсия	
5	32	29	43	34	32		Медиана	
6	30	43	54	42	47		Мода	
7	35	49	49	54	36		Асимметрия	
8	36	51	36	24	35		Эксцесс	
9	25	33	38	38	36		Наименьшее	
10	29	51	32	36	53		Наибольшее	
11	30	55	44	46	38		Размах	
12							Квартиль 1	
13							Квартиль 3	
14							Кол-во выборок	50

Рис. 1.3. Образец 1

3. Выполните расчеты указанных в статистической таблице параметров, вставляя при помощи средства  **Вставка функции** расчетные формулы как показано на **Образце 2**(рис. 1.4.):


– вставьте формулу для расчета среднего: выделите ячейку **H2** / инструмент **Вставка функции**  / в окне **Мастер функций** в поле **Категории** выберите **Статистические**, в поле **Функция** при помощи полосы прокрутки пролистайте список названий функций, найдите и выберите **СРЗНАЧ** / **ОК** / в окне вставки функции справа от поля **Число 1** кнопка сворачивания  / выделите мышью диапазон ячеек **A2:E11** / в свернутом окне вставки функции кнопка разворачивания  / **ОК**;

– аналогично вставьте остальные формулы из **Образца 2**.



	G	H
1		x
2	Среднее	=СРЗНАЧ(A2:E11)
3	Среднеквадратичное откл.	=СТАНДОТКЛОН(A2:E11)
4	Дисперсия	=ДИСПА(A2:E11)
5	Медиана	=МЕДИАНА(A2:E11)
6	Мода	=МОДА(A2:E11)
7	Асимметрия	=СКОС(A2:E11)
8	Эксцесс	=ЭКСЦЕСС(A2:E11)
9	Наименьшее	=МИН(A2:E11)
10	Наибольшее	=МАКС(A2:E11)
11	Размах	=МАКС(A2:E11)-МИН(A2:E11)
12	Квартиль 1	=КВАРТИЛЬ(A2:E11;1)
13	Квартиль 3	=КВАРТИЛЬ(A2:E11;3)
14	Кол-во выборок	50


Рис. 1.4. Образец 2



4. Сформируйте таблицу частот исследуемой величины, выполнив группировку данных и расчеты в соответствии с **Образцом 3** (рис. 1.5.) непосредственным вводом формул и при помощи средства  **Вставить функцию:**

– введите заголовки строк и столбцов по образцу;

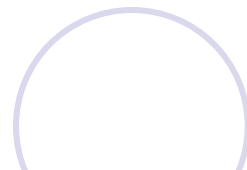
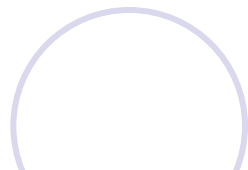
– вставьте формулу для вычисления минимального числа интервалов группирования по эмпирическому соотношению $k \leq 5 \cdot \lg N$ при помощи средства **Вставка функции**: выделите ячейку **B14** / инструмент **Вставка функции**  / в поле **Категории** выберите **Математические** / в поле **Функция** найдите и выберите **ОКРУГЛ** / **ОК** / в окне вставки функции установите курсор в поле **Количество_цифр** и введите **0** (округление до целого числа), установите курсор в поле **Число** и введите **5*** (множитель) / в инструменте выбора функции  кнопка списка и выберите **Другие функции...** / в окне **Мастер функции** выберите функцию **LOG10** из категории **Математические** / **ОК** / в окне вставки функции в поле **Число 1** введите ссылку на ячейку с количеством выборок **H14** / **ОК**;








– вставьте формулу для расчета ширины интервала группирования методом непосредственного ввода: выделите ячейку **B15** / введите знак **=(равно)** и знак **((скобка)** / щелкните ячейку с максимальным значением **H10** / клавиша **F4** для перехода к абсолютной ссылке / введите знак **– (минус)** / щелкните ячейку с минимальным значением **H9** / клавиша **F4** / введите знак **) (скобка)** и знак **/ (наклонная черта)** и щелкните ячейку **B14** с числом интервалов / **Enter**;




	А	В	С
13	Группирование		
14	Мин. кол-во интервалов	=ОКРУГЛ(5*LOG10(Н14),0)	
15	Ширина интервала	=(Н\$10-Н\$9)/В14	
16			
17			
18	Таблица частот		
19	Интервалы, x _{прав}	Частоты	Накопленные частоты
20	=Н\$9+В15	=ЧАСТОТА(А2:Е11;А20:А27)	=В20
21	=А20+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С20+В21
22	=А21+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С21+В22
23	=А22+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С22+В23
24	=А23+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С23+В24
25	=А24+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С24+В25
26	=А25+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С25+В26
27	=А26+В\$15	=ЧАСТОТА(А2:Е11;А20:А27)	=С26+В27

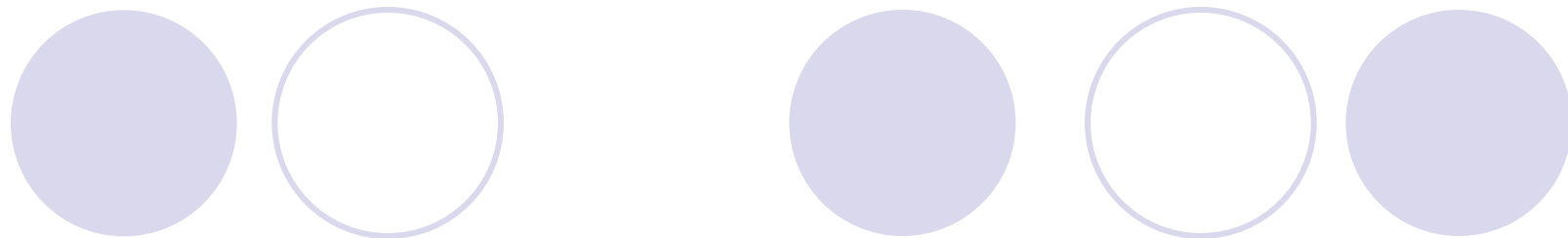
Рис. 1.5. Образец 3



– аналогично вставьте формулы для вычисления правых границ интервалов как показано на **Образце 3** (введите формулы в ячейки **A20**, **A21**, в ячейки от **A22** до **A27** растяните формулу из ячейки **A21** при помощи автозаполнения: после ввода формулы в **A21** укажите на правый нижний угол ячейки **A21** до появления маркера автозаполнения **+(малый черный плюс)**, нажмите левую кнопку мыши и, удерживая ее, протяните выделение до ячейки **A27** и опустите кнопку мыши). Эту же схему можно использовать и для столбца с накопленными частотами;

– вставьте формулу для расчета частот с применением функции массивов: выделите диапазон ячеек **B20:B27** / инструмент **Вставка функции**  / найдите и выберите функцию **ЧАСТОТА** из категории **Статистические** / **ОК** / в окне вставки функции справа от поля **Массив_данных** кнопка сворачивания  / выделите мышью диапазон ячеек исходных данных **A2:E11** / кнопка разворачивания  / справа от поля **Двоичный_массив** кнопка сворачивания  / выделите мышью диапазон ячеек интервалов **A20:A27** / кнопка разворачивания  / клавиши **Ctrl+Shift+Enter** для фиксации функции массива.

5. Постройте гистограмму для исследуемой величины с применением мастера диаграмм: выделите диапазон ячеек с таблицей частот **A19:C27** / инструмент **Мастер диаграмм**  / на вкладке **Нестандартные** в поле **Тип** выберите вариант **График|Гистограмма 2** и кнопка **Далее** / в окне **...источник данных диаграммы** на вкладке **Диапазон данных** включите переключатель **в столбцах** / на вкладке **Ряд** кнопка сворачивания  справа от поля **Подписи оси X** / выделите диапазон ячеек **A20:A27** / кнопка разворачивания  / в поле **Ряд** выберите **Интервалы...** / кнопка **Удалить** / кнопка **Далее>** / в окне **...параметры диаграммы** на вкладке **Заголовки** напечатайте в полях **Название диаграммы** текст **Гистограмма 1**, **Ось X (категорий)** – текст **Интервалы** / на вкладке **Линии сетки** установите флажки **основные линии** в разделах **Ось X** и **Ось Y** / кнопка **Далее>** / в окне **...размещение диаграммы** включите переключатель **имеющемся** / **Готово**.



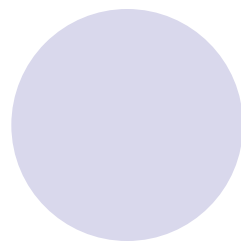
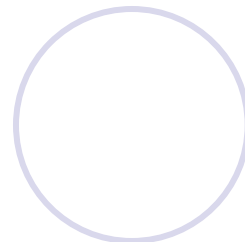
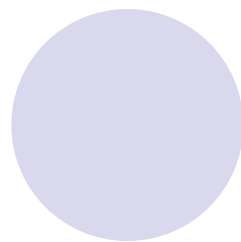
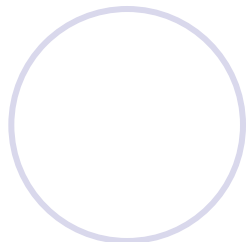
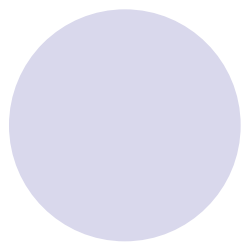
6. Скорректируйте построенную гистограмму: выделите диаграмму щелчком мыши по ней / переместите рамку диаграммы правее таблицы частот / укажите мышью на угловой ограничительный маркер диаграммы до появления указателя в форме двунаправленной стрелки и растяните мышью размеры диаграммы / измените оформление линии накопленных частот / щелкните линию правой кнопкой мыши и выберите команду **Формат рядов данных** контекстного меню / в окне **Формат ряда данных** на вкладке **Вид** установите флажок **Сглаженная линия** в разделе **Линия** и включите переключатель **отсутствует** в разделе **Маркер** и **ОК** / снимите заливку области построения диаграммы: выделите область построения (сетку) щелчком мыши / щелкните кнопку списка инструмента **Цвет заливки**  и прямоугольник **Нет заливки** в палитре.



Задание 2. Проведите аппроксимацию и сглаживание построенных в **Задании 1** гистограмм при помощи построения линий тренда основных типов – линейного, логарифмического, полиномиального, степенного, экспоненциального и скользящего среднего.

1. При необходимости добавьте в конец книги рабочий лист (щелчок правой кнопкой по листу **Статистика_1** / **Добавить** контекстного меню / значок **Лист** / **ОК** / переместите ярлычок созданного листа за ярлычок **Статистика_1**) и переименуйте ярлычок следующего рабочего листа в **Тренд 1**.

2. Выделите и скопируйте (**Правка** / **Копировать**) в буфер гистограмму **Гистограмма 1**. Вставьте график из буфера в начало рабочего листа **Тренд 1**: щелчок в ячейке **A1** листа **Тренд 1** и **Правка** / **Вставить**.



3. Добавьте линейный тренд для ряда **Частоты** на гистограмму: выделите **Гистограмму 1** / щелчок правой кнопкой мыши по одному из столбиков ряда **Частоты** / **Добавить линию тренда** контекстного меню / в окне **Линия тренда** на вкладке **Тип** выберите образец **Линейная** / **ОК**.

4. Скопируйте **Гистограмму 1** с линейным трендом на рабочем листе **Тренд 1** в позицию под нижней границей рамки уже имеющейся диаграммы с линейным трендом, измените тип линии тренда на логарифмический: щелчок правой кнопкой мыши по линии тренда во второй копии гистограммы / **Формат линии тренда** контекстного меню / в окне **Формат линии тренда** на вкладке **Тип** выберите образец **Логарифмическая** / **ОК**.

5. Аналогично постройте еще 4 версии **Гистограммы 1** с остальными типами линий тренда (полиномиальная степени 2, степенная, экспоненциальная и скользящее среднее), сравните варианты и выберите два наиболее соответствующих данной эмпирической гистограмме.



6. Проанализируйте поведение полиномиального тренда при изменении степени полинома: выделите гистограмму с полиномиальным трендом и вставьте две ее копии на новый рабочий лист, предварительно переименованный в **Полиномиальный тренд_1** / при помощи контекстного меню полиномиального тренда на второй копии гистограммы откройте диалоговое окно **Формат линии тренда** / на вкладке **Тип** для образца **Полиномиальная** в поле со списком **Степень** установите значение **3** (вместо предыдущего **2**) / на вкладке **Параметры** установите флажки **показывать уравнение на диаграмме**, **поместить на диаграмму величину достоверности аппроксимации** / **ОК** / сформируйте аналогичные версии для значений степеней полиномиальной аппроксимации **4** и **5** / сравните полученные графики, уравнения аппроксимирующих полиномов и значения достоверности аппроксимации.

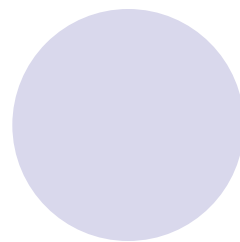
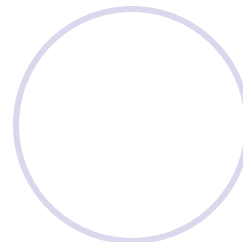
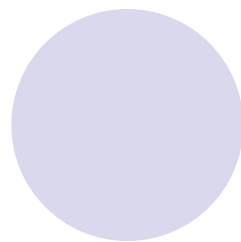
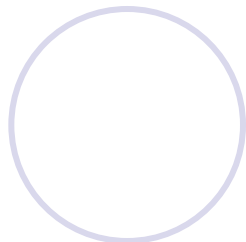
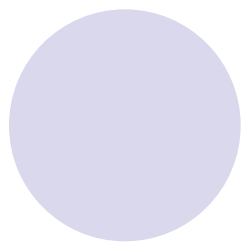



Задание 3. Выполните процедуру генерации случайных чисел и визуализации данных с использованием средств **Анализ данных** и **Мастер диаграмм MS Excel**.


1. Перейдите на свободный рабочий лист книги **Примеры_стат** (при отсутствии такового вставьте новый и перетащите его ярлычок в конец книги) и переименуйте ярлычок рабочего листа в **Генерация данных**.


2. Подключите надстройку **Пакет анализа MS Excel: Сервис / Надстройки /** в окне **Надстройки** установите флажок **Пакет анализа / ОК**.

3. Выполните генерацию 100 случайных чисел, распределенных в соответствии с нормальным законом с нулевым средним и дисперсией 1: в ячейку **A1** введите заголовок столбца с данными **$x, N\{0,1\}, 100$** / установите курсор в ячейку **A2 / Сервис / Анализ данных /** в окне **Анализ данных** в списке поля **Инструмент анализа** выберите **Генерация случайных чисел / ОК /** в окне **Генерация случайных чисел** в поле **Число переменных** введите **1**, в поле **Число случайных чисел** введите **100**, в списке поля **Распределение** выберите **Нормальное**, введите в полях **Среднее – 0**, **Стандартное отклонение – 1**, в разделе **Параметры вывода** включите переключатель **Выходной интервал /** кнопка сворачивания  / щелкните ячейку **A2 /** кнопка разворачивания  / **ОК**.



4. Измените разрядность данных, уменьшив число знаков после запятой до четырех: выделите диапазон ячеек **A2:A101** / инструмент **Уменьшить разрядность**  .

5. Отобразите сгенерированные данные графически: выделите диапазон ячеек **A2:A101** / инструмент **Мастер диаграмм**  / на вкладке **Стандартные** в поле **Тип** выберите **График**, в поле **Вид** – первый образец / кнопка **Далее>** / в окне **...источник данных диаграммы** на вкладке **Диапазон данных** включите переключатель **в столбцах** / кнопка **Далее>** / в окне **...параметры диаграммы** на вкладке **Легенда** снимите флажок **Добавить легенду** / кнопка **Далее>** / в окне **...размещение диаграммы** включите переключатель **имеющемся** / **Готово** / увеличьте размеры диаграммы и снимите заливку области построения для лучшего восприятия графика.




6. Действуя по схеме п.п. 3-5, на листе **Генерация данных** в столбце **В** сгенерируйте и отобразите нормально распределенные данные с нулевым средним и стандартным отклонением 2, количеством чисел 100, заголовок столбца **x,N{0,2},100**;

Задание 4. При помощи средства **Анализ данных** выполните расчет описательной статистики по сгенерированным в **Задании 3** данным.

1. Рассчитайте описательную статистику по данным столбца **А**: на листе **Генерация данных** выполните **Сервис / Анализ данных /** в окне **Анализ данных** выберите **Описательная статистика / ОК /** в окне **Описательная статистика** в поле **Входной интервал** введите ссылку на диапазон ячеек **A1:A101 /** в разделе **Группирование** включите переключатель **по столбцам** и установите флажок **Метки в первой строке /** в разделе **Параметры вывода** включите переключатель **Новый рабочий лист** и в поле ввода справа напечатайте текст **Статистика N{0,1},100** названия ярлычка листа для размещения бланка результатов, установите флажок **Итоговая статистика / ОК.**

2. Аналогично проведите расчеты описательной статистики для столбца **В**, размещая бланки результатов на листе **Статистика N{0,2},100**. При необходимости скорректируйте ширину столбцов итоговых бланков так, чтобы читались все записи в таблицах.



Задание 5. При помощи средства **Анализ данных** выполните процедуру построения гистограммы по сгенерированным в **Задании 3** данным.

1. Постройте гистограмму по данным столбца **A**: на листе **Генерация данных** выполните **Сервис / Анализ данных /** в окне **Анализ данных** выберите **Гистограмма / ОК /** в окне **Гистограмма** в разделе **Входные данные** в поле **Входной интервал** укажите диапазон ячеек **A1:A101** и установите флажок **Метки /** в разделе **Параметры вывода** включите переключатель **Выходной интервал**, установите курсор в поле ввода справа (при необходимости предварительно очистите его) / щелчок по ярлычку листа **Статистика N{0,1},100** и затем в ячейке **D1** для указания размещения гистограммы / установите флажки **Интегральный процент** и **Вывод графика / ОК /** при необходимости скорректируйте ширину столбцов таблицы частот, размеры и параметры гистограммы для лучшего восприятия результатов.

2. Аналогично проведите построение гистограммы для столбца **B** размещая результаты начиная с ячейки **D1** на листе **Статистика N{0,2},100**.

3. Добавьте на построенные гистограммы к ряду частот линии полиномиального (см. **Задание 2**) со степенью 2,3,4 тренда. Используя команду контекстного меню **Формат линии тренда**, при помощи вкладки **Вид** соответствующего диалогового окна оформите линии тренда различным цветом.



4. Содержание отчета

- 4.1. Цель работы.
- 4.2. Методики расчета определяемых величин.
- 4.3. Результаты выполнения работы в виде таблиц исходных данных и результатов расчетов определяемых величин, построенных гистограмм и графиков.
- 4.4. Выводы по работе.

5. Контрольные вопросы

- 5.1. Что такое варианта?
- 5.2. Как составляется вариационный ряд?
- 5.3. В каких случаях строится интервальный вариационный ряд?
- 5.4. В каких случаях строится полигон распределения?
- 5.5. Что такое квартили?
- 5.6. Что такое медиана?
- 5.7. Что такое мода?
- 5.8. Что такое амплитуда вариационного ряда?

