**Лабораторная работа №1  
по курсу «Компьютерный анализ статистических данных»  
на тему «Описательная статистика»**

## Цель работы

Ознакомиться с основными числовыми характеристиками, которые используются в описательной статистике: показателями середины выборки, разброса значений, формы распределения. Научиться рассчитывать их в MS Excel.

## Теоретические сведения

### Основы статистики

Раздел математики, посвященный методам сбора, анализа и обработки статистических данных для научных и практических целей, называется **математической статистикой**. Данный раздел математики имеет дело с массовыми явлениями и тесно связан с *теорией вероятностей*, так как базируется на ее математическом аппарате.

Методы статистики используются в различных науках и дисциплинах – физике, технике, биологии, медицине, экономике, социологии, психологии, юриспруденции и даже в лингвистике. С ее помощью можно изучать поведение потребителей на рынке, надежность технических систем и вероятность ошибок в сочинениях школьников.

**Статистика** – более широкое понятие, чем математическая статистика. Математическая статистика не изучает вопросы сбора исходных данных, проведения опросов и т.п., а только методы их обработки. Методы математической статистики не зависят от содержательного смысла изучаемого явления, но выводы по ее результатам – зависят.

*Цель* статистического исследования – исследование соотношений между статистическим данными (**описательная статистика**) и использование результатов исследований для прогнозирования и принятия решений (**аналитическая статистика**).

*Статистические данные* представляют собой данные, полученные в результате обследования большого число объектов или явлений.

**Статистическая совокупность** – множество обследуемых объектов (людей, семей, предприятий, автомобилей, животных, растений, ...).

По охвату статистической совокупности исследование может быть сплошное или не сплошное. При сплошном исследовании группа формируется из всех единиц изучаемого явления (**генеральная совокупность**), а при не сплошном – только часть этих единиц (**выборка**).

Пример

Исследуется средний рост студентов в России. Генеральная совокупность – рост абсолютно всех студентов всех вузов в конкретный момент времени, например на 1 сентября 2014г. Выборка – только для части студентов. Примеры выборок:

а) студенты только одного вуза

б) студенты Москвы, Самары и Томска города;

в) случайно выбранная 1000 студентов по всей стране.

Цель изучения выборочной совокупности (выборки) – получение информации не о ней самой, а о генеральной совокупности. Выборка – это неполная информация обо всех возможных представителях генеральной совокупности. В идеале, лучше исследовать генеральную совокупность, но на практике это не всегда удается (дорого, долго, физически невозможно).

Пример

Данные экзит-поллов на политических выборах – практический пример обследования выборочной совокупности (только часть избирателей опрашивается). Но это выборка позволяет предварительно судить о результатах голосования всех избирателей.

Поэтому обычно стремятся сделать так, чтобы выборка наилучшим образом представляла генеральную совокупность, т.е. была *репрезентативной* (*представительной*). Репрезентативная выборка – та, которая достаточно точно отражает свойства генеральной совокупности.

Пример

Для исследования среднего роста жителей города, было выбрано случайное здание и измерен рост находящихся в нем людей. Получили средний рост 2,14м. Причина – выбранное здание оказалось спортивным залом, в котором занимались баскетболисты. Такая выборка была *нерепрезентативной*.

Получить выборку можно разными способами. Самый простой – случайный выбор представителей генеральной совокупности.

При исследовании нас интересует не каждый объект целиком, а только какие-то его свойства (**признаки**).

Признаки могут быть:

1. *Количественные* (числовые) и *качественные* (нечисловые).

Пример

У объекта «человек» рассматриваются признаки: год рождения, пол, возраст, вес, рост, цвет глаз.

Количественные: год рождения, возраст, вес, рост.

Качественные: пол, цвет глаз.

1. *Моментные* и *интервальные*.

Пример

У объекта «фирма» рассматриваются признаки: общее число сотрудников; число сотрудников, принятых на работу; число уволенных сотрудников.

Моментные: общее число сотрудников (указать, на какое число).

Интервальные: число сотрудников, принятых на работу; число уволенных сотрудников (указать, за какой период времени – за месяц, за год, за день).

1. *Абсолютные* (в натуральных или денежных единицах) и *относительные* (в процентах или долях).

Пример

У объекта «семья» рассматриваются признаки: ежемесячный доход, количество членов семьи, доля затрат на продукты.

Абсолютные: наличие ежемесячный доход (руб.), количество членов семьи (чел.).

Относительные: доля затрат на продукты (% от общей суммы доходов).

1. *Случайные* и *неслучайные* (детерминированные), зависит от совокупности.

Пример

У объекта «студент» рассматриваются признаки: наличие зачетной книжки, курс, средний балл.

Случайные: курс, средний балл.

Неслучайные: наличие зачетной книжки.

Но если объект – «студент 1-го курса», то признак «курс» тоже будет неслучайным.

Статистика имеет дело со случайными количественными признаками, поэтому все данные в выборке должны быть числовыми. Обозначают выборки заглавными латинскими буквами *X*, *Y*, *Z*. Каждое значение в выборке называют *наблюдением* и обозначают *xi*, *yi*, *zi*, где *i* – номер наблюдения.

Пример

Исследуется *X* возраст студентов вуза (очной формы обучения) и *Y* средний балл. Всего опросили *n* = 100 студентов.

Исходная выборка имеет вид:

|  |  |  |  |
| --- | --- | --- | --- |
| *i* | *X* | *Все оценки* | *Y* |
| 1 | 18 | 4453 | 4 |
| 2 | 22 | 34433345444344555434444 | 3,9 |
| 3 | - | 5555555 | 5 |
| 4 | 17 | - | - |
| ... |  |  |  |
| 100 | 20 | 55443444444454 | 4,1 |

*x*1 = 18, *x*2 = 22, ... *y*1 = 4,0, *y*2 = 3,9, ...

Из нее получили сводную таблицу:

|  |  |  |
| --- | --- | --- |
| *X* | Количество студентов | Y |
| 18 и менее | 19 | 4,4 |
| 19 | 17 | 4,2 |
| 20 | 25 | 4,0 |
| 21 | 14 | 3,8 |
| 22 | 23 | 3,7 |
| 23 и более | 2 | 4,3 |
| *Всего* | *100* |  |

Можно заметить, что в среднем, в этом вузе с возрастом оценки ухудшаются.

Качественные признаки тоже «переводят» в числа.

Примеры

Пол: 1 - муж, 2 - жен.

Наличие машины: 0 - нет, 1 - есть.

Качество жилищных условий: в баллах от 0 до 10.

### Сведения из теории вероятностей

**Случайное событие** – это любое событие, о котором заранее неизвестно, произойдет оно или нет, или утверждение, о котором заранее неизвестно, истинно оно или ложно.

Примеры

Тело без опоры падает на поверхность земли (*неслучайное* событие).

При броске игральной кости на ней выпадет число 6 (*случайное*).

Завтра курс валюты снизится (*случайное*).

Тело человека состоит из клеток (*неслучайное*).

Прием лекарства поможет выздороветь (*случайное*).

При запросе к сайту он будет доступен (*случайное*).

Случайное событие определяется некоторой **вероятностью**. Вероятность события – интуитивно понятная характеристика, определяющая шанс на то, что событие произойдет.

*Формальное определение вероятности p* – отношение числа положительных исходов *m* (когда событие произошло) к общему числу исходов *n*.



Вероятность никогда не бывает отрицательной или больше 1.

**Случайная величина (СВ)** – число, значение которого заранее неизвестно. Для случайных величин задается вероятность появления конкретного значения или диапазона значений.

Случайные величины могут быть дискретными и непрерывными.

**Дискретная СВ** – множество отдельных значений (конечное или бесконечное). Каждому значению СВ можно сопоставить целое число. Сами значения СВ не обязательно целые.

**Непрерывная СВ** может принимать любое значение из некоторого диапазона. Количество возможных значений непрерывной СВ всегда бесконечно.

Примеры

Дискретные СВ: число попаданий в мишень, количество краж за неделю, число заказов за день, количество бракованных изделий в партии, оценка на экзамене.

Непрерывные СВ: температура воздуха на улице, масса буханки хлеба, яркость освещения, мощность сигнала радиостанции.

Многие дискретные по сути СВ для простоты считают непрерывными: цена (точность до копеек, центов), численность населения.

**Закон распределения** СВ характеризует вероятность появления каждого значения СВ. Сумма всех вероятностей должна быть равна 1, иначе какие-то пропустили.



***Для дискретных СВ*** закон распределения задается:

* таблично;

|  |  |  |  |
| --- | --- | --- | --- |
| *x1* | *x2* | … | *xn* |
| *p1* | *p2* | … | *pn* |

* графически;

Pn

P2

P1

x

x1

x2

…

xn

p

* формулой, например, если все значения равновероятны:



Примеры

Бросание игральной кости (СВ – выпавшее число).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 1 | 2 | 3 | 4 | 5 | 6 | Сумма |
| *p* | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1 |



Количество законченных высших образований в некоторой стране:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *x* | 0 | 1 | 2 | 3 | 4 | Сумма |
| *p* | 0,2 | 0,5 | 0,2 | 0,099 | 0,001 | 1 |

***Для непрерывных СВ*** нельзя задать вероятность каждого отдельного значения, ведь их бесконечно много. Поэтому для них задают вероятность попадания в некоторый *интервал значений*.

Пример

Вероятность температуры тела здорового человека от 36 до 37 значительно выше, чем вероятность температуры >40.

**Функция распределения (закон распределения, интегральная функция)** *F*(*x*) непрерывной СВ – вероятность того, что значение случайной величины *X* меньше числа *x*:



Задается формулой (часто очень сложной).

Примеры графиков функции распределения:

0

1

x

F(x)

Пример

На следующем графике:

F(2) = p(X < 2) = около 0

F(4) = p(X < 4) = 0,2

F(5) = p(X < 5) = 0,5

F(6) = p(X < 6) = 0,8

F(8) = p(X < 8) = почти 1

Отсюда:

p(X > 4) = 1 – p(X < 4) = 1 – 0,2 = 0,8

p(4 < X < 8) = p(X > 4) – p(X < 8) = 0,8 – 0,2 = 0,6

Но по функции распределения не очень удобно судить о том, какие значения более вероятны, а какие менее вероятны.

**Плотность распределения (дифференциальная функция**) *f*(*x*) – это производная интегральной функции распределения.



Сами значения плотности распределения не являются вероятностями и могут быть больше 1. Но сравнивая между собой два значения *f*(*x*), можно судить, какое из них более вероятное, а какое – менее вероятное. Чем выше график, тем больше вероятность.

Примеры графиков плотности функции:

0

x

f(x)

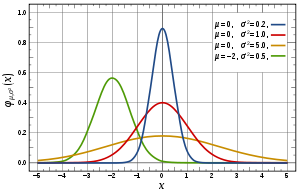
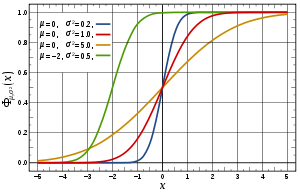
Для предыдущего примера:

Самое вероятное значение = 5. Значения < 2 и >8 почти никогда не появляются. С одинаковой вероятностью могут попасться числа > 5 и < 5.

Существует много разных законов распределения.

**Нормальный (Гауссов) закон распределения** считается одним из наиболее распространенных на практике. Зачастую, если закон распределения неизвестен, его принимают нормальным.

Функция и плотность нормального закона распределения имеют вид:



В предыдущих примерах закон распределения тоже был нормальным.

### Обобщающие числовые характеристики

Делать выводы по отдельным значениям выборки трудно, особенно, когда их много. Поэтому используют различные числовые характеристики, которые позволяют судить **в целом о выборке** (и генеральной совокупности).

**Мощность (объем) выборки** *n* – это число наблюдений в выборке. Чем оно больше, тем ближе выборка к генеральной совокупности. Реальные выборки обычно содержат сотни, тысячи наблюдений. Выборки в 30-50 наблюдений и менее считаются малыми.

**Выборочное среднее** *m* – показывает усредненное значение всех наблюдений. Одна из основных характеристик выборки.



Пример

X: 4; 35; 7; 8; 16

n = 5

*mX* = (4 + 35 + 7 + 8 + 16) / 5 = 70 / 5 = 14

**Медиана** *Me* – серединное значение упорядоченного ряда вариантов значений. Условный центр выборки.

Чтобы найти медиану, нужно упорядочить значения выборки и взять то, которое находится в середине. Например, если *n* = 15, то нужно взять 8-ое по значению наблюдение *x*8. Если *n* четно, то берется среднее между двумя значениями в центре. Например, если n = 20, то *Me* = (*x*10 + *x*11)/2.

Пример1

Выборка: 6;29; 5; 25; 11; 8; 5

n = 7

Упорядоченная выборка: 5; 5; 6; **8;** 11; 25; 29

*Me* = 8

Пример2

Выборка: 240; 104; 302; 300;160; 440

n = 6

Упорядоченная выборка: 104; 160; **240;** **300;** 302; 440

*Me* = (240 + 300) / 2 = 270

**Мода** *Mo* – наиболее вероятное, часто встречающееся значение выборки.

Пример

Модная вещь – это та, которую чаще всего можно увидеть на ком-нибудь еще.

Чтобы найти моду дискретной СВ, нужно посчитать, сколько раз встречается каждое значение в выборке. Мод может быть несколько, если какие-то значения встречаются одинаково часто. Если все значения встречаются одинаковое число раз (один раз), то мода не определена.

Пример

Выборка: 2; 2; 9; 9; 9; 10; 11; 12; 12; 13

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Значение | 2 | **9** | 10 | 11 | 12 | 13 |
| Частота | 2 | **3** | 1 | 1 | 2 | 1 |

*Mo* = 9,

*Me* = ?,

*m* = ?

Для непрерывных СВ мода определяется с помощью плотности распределения (самая высокая точка графика).

У ***однородной*** выборки (близкой к нормальному закону распределения) среднее значение, мода и медиана близки между собой. Медиана сильно отличается от среднего, когда в выборке присутствуют аномально большие или маленькие значения. В таком случае медиана предпочтительнее среднего для описания выборки.

Пример

Представьте, что в одной комнате собрали 9 бедняков и одного богача. У бедняков есть по 10 рублей, а у богача – миллион. Сколько у них в среднем денег?

Выборка: 10; 10; 10; 10; 10; 10; 10; 10; 10; 1 000 000

*n* = 10

*m* = 1 000 090 / 10 = 100 009

*Me* = 10

Другая группа показателей характеризуют ***разброс значений*** выборки – насколько далеко находятся ее значения от центра и друг от друга.

Пример

В двух регионах средняя заработная плата одинакова и составляет 1000 д.е. Однако в первом регионе 90% населения получают зарплату от 900 до 1200 д.е., а в другом 90% населения получают зарплату от 50 до 20 000 д.е. Очевидно, что хотя средние зарплаты в регионах равны, разброс у них разный.

**Размах вариации** *V* – разность между максимальным и минимальным значениями выборки. Показывает максимальную амплитуду вариации, в этот диапазон укладываются все значения выборки.

Пример

5; 12; 14; 17; 19; 25; 27

*V* = 27 – 5 = 22

Размах вариации – простая, но неудобная величина: два крайних значения могут быть очень далеки ото всех остальных.

Пример

–4; 1; 105; 106; 106; 107; 110; 112; 125; 126; 528

*V* = 528 – (–4) = 532

Размах вариации получился очень большим, хотя большинство наблюдений лежит в пределах 100-130.

**Дисперсия (вариация)** *D* – условная величина из теории вероятностей, описывающая разброс значений относительно среднего.





Пример

15; 45; 60

*m* = (15 + 45 + 60) / 3 = 40

(*xi* – *m*): –25; 5; 20

(*xi* – *m*)2: 625; 25; 400

*D* = (625 + 25 + 400) / 2 = 1050 / 2 = 525

Дисперсия измеряется в квадратах исходных единиц измерения (руб.2, чел.2 и т.п.) и плохо поддается интерпретации. Поэтому оценивают **стандартное отклонение** *S* – корень из дисперсии. Смысл тот же, но единицы измерения обычные.

Пример (продолжение)

*D* = 525



**Коэффициент вариации** *KV* позволяет измерить вариацию в процентах.



Считается, что:

если *KV* ≤ 40% – вариация выборки небольшая;

40% <*KV* ≤ 60% – вариация средняя (умеренная);

*KV* > 60% – вариация сильная.

Пример (продолжение)

*m* = 40 *S* = 22,9

*KV* = 22,9 / 40 ∙ 100% = 57,25% (умеренная вариация)

**Квартили** похожи на медиану, но разделяют не половины, а четверти выборки. Четверть всех значений выборки меньше нижнего квартиля *Qн*. Четверть значений больше верхнего квартиля *Qв*. И половина значений выборки находится между квартилями.

Пример

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *i* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* |
| *X* | 10 | 10 | 13 | **14** | 17 | 18 | 18 | 18 | 20 | 21 | **22** | 22 | 25 | 29 |

*Qн* = 14 *Qв* = 22

Половина всех значений находится между 14 и 22.

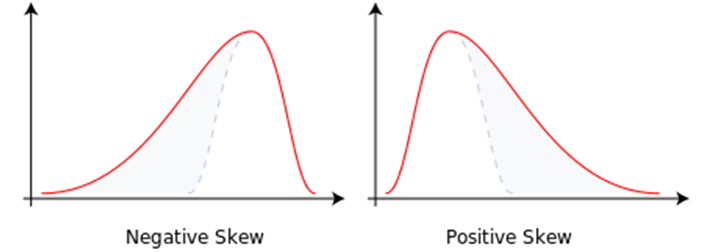
Еще два показателя характеризуют форму распределения в сравнении с нормальным законом.

**Коэффициент асимметрии** KA показывает, с какой стороны от среднего больше значений.

Если он положительный (KA < 0), то больше значений справа, т.е. чаще встречаются значения больше среднего. Медиана больше среднего.

Если отрицательный (KA < 0) – то больше слева, т.е. чаще встречаются значения меньше среднего. Медиана меньше среднего.

Если равен 0, то распределение симметрично относительно среднего (слева и справа от среднего одинаковое количество значений). Медиана равна среднему.



Отрицательная (левая) асимметрия

Положительная (правая) асимметрия

Примеры

В двух регионах средняя заработная плата одинакова и составляет 1000 д.е. Однако в первом регионе коэффициент асимметрии равен –0,5, а во втором +3,4. Это означает, что в первом регионе тех, кто получает зарплату меньше средней и больше средней почти поровну, но тех, кто получает «маленькую» зарплату все же чуть-чуть больше, чем тех, кто получает «большую». А во втором регионе гораздо больше тех, кто получает «большую» заплату (больше средней).

В примере с 9 бедняками и 1 богачом коэффициент асимметрии отрицательный (из 10 наблюдений 9 меньше среднего и только 1 больше, медиана гораздо меньше среднего).

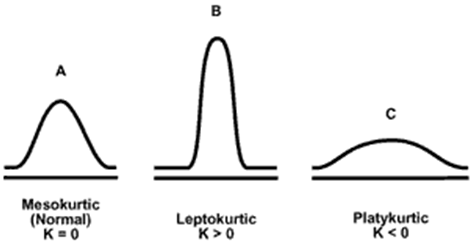
|  |  |  |  |
| --- | --- | --- | --- |
| KA > 0 | правая | *Me > m* | много «богатых», «больших» |
| KA = 0 | симметрично | *Me = m* | «богатых» и «бедных» поровну |
| KA < 0 | левая | *Me < m* | много «бедных», «маленьких» |

**Коэффициент эксцесса** E характеризует остроту пика распределения.

Если E > 0, то пик острый (островершинное), т.е. большинство значений выборки близко к моде, а одно или несколько находятся где-то далеко.

Если E < 0, то значения выборки равномерно распределяются по всему диапазону (плосковершинное), т.е. много и «больших», и «маленьких», и «средних» значений.

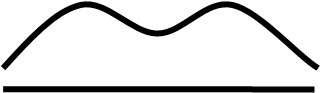
Если E = 0, то распределение близко к нормальному, т.е. чем дальше от среднего, тем меньше таких значений, убывание плавное.



E = 0

E > 0

E < 0

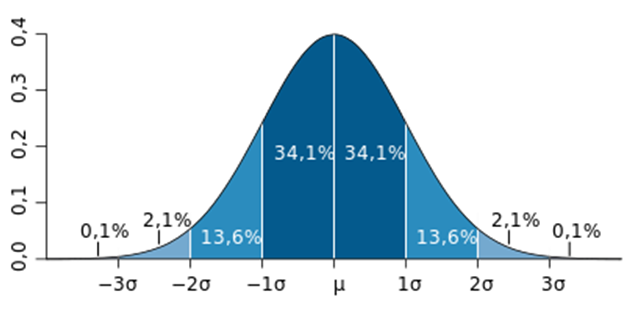


E << 0

Примеры

В четырех регионах средняя заработная плата одинакова и составляет 1000 д.е. Стандартное отклонение тоже одинаковое и равно 200 д.е. Но в первом регионе E = 0,01, во втором E = –1,2, в третьем E = –5,1, а в четвертом E = 3,2.

Это означает, что в первом регионе большинство людей все же получает зарплату, близкую к средней (68% от 800 до 1200 д.е.), а тех, кто получает очень большую или очень маленькую меньше, причем, чем дальше зарплата от среднего, тем меньше людей с такой зарплатой (по 13,5% получают от 600 до 800 или от 1200 до 1400, по 2% от 400 до 600 и по 1400 до 1600, а тех, кто получает меньше 400 или больше 1600 – менее 1%).



–3*S*

–2*S*

–1*S*

1*S*

2*S*

3*S*

*m*

400

600

800

1200

1400

1600

1000

**Правило трех *S***(только для нормального распределения, E = 0 и KA = 0)

В других регионах правило 3*S* не работает.

Во втором регионе почти поровну людей и с большой, и со средней, и с маленькой зарплатой. Правило 3*S* не работает.

В третьем регионе многие получают большую или маленькую зарплату, а людей со средней зарплатой мало.

В четвертом регионе большинство (80-90%) получают среднюю зарплату, и лишь немногие получают очень большую или очень маленькую (но они есть!).

## Задание

Имеется статистическая выборка с наблюдениями некоторого показателя *X*. Необходимо сформировать описательную статистику показателя, рассчитав следующие обобщающие характеристики выборки:

* + мощность выборки *n*;
  + выборочное среднее *m*;
  + медиану *Me*;
  + моду *Mo*;
  + размах вариации *V*;
  + дисперсию *D*;
  + стандартное отклонение *S*;
  + коэффициент вариации *KV*;
  + верхний и нижний квартили *Q*в и *Q*н.

Охарактеризовать выборку на основе полученных значений.

### Исходные данные

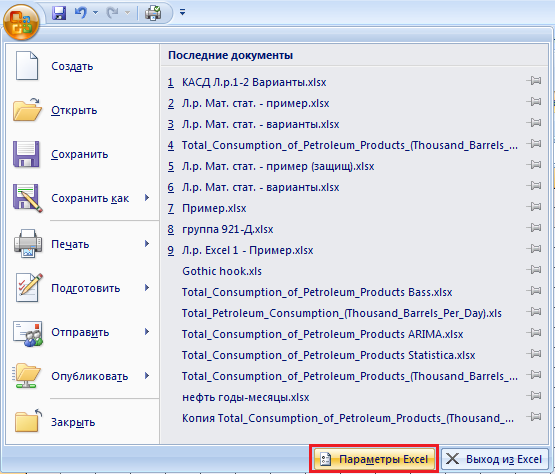
Исходные выборки находятся в файле «КАСД Л.р.1 Варианты.xlsx» Вариант назначается преподавателем.

## Подготовка к работе

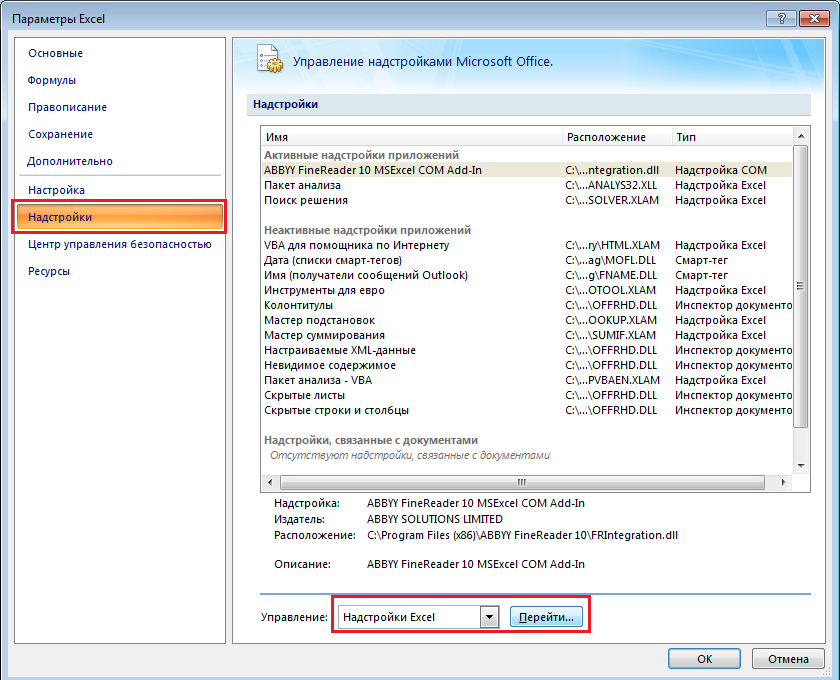
### Подготовка программы

Работа выполняется в MS Excel 2007 или выше. Для выполнения этой и последующих работ потребуется пакет «Анализ данных». По умолчанию он отключен.

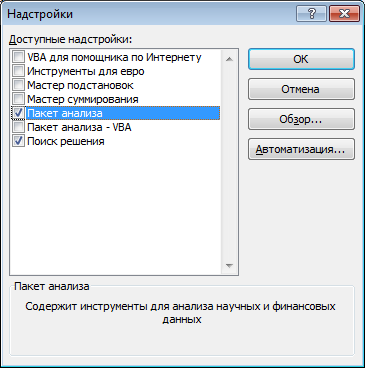
1. В меню Office нажать кнопку «Параметры Excel».



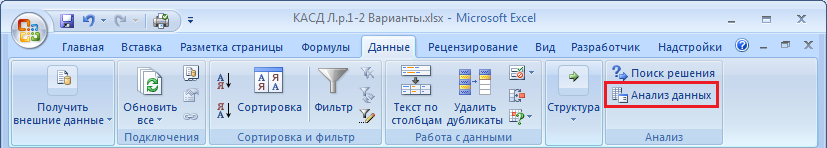
1. В окне параметров Excel выбрать «Надстройки», в выпадающем списке внизу выбрать «Надстройки Excel» и нажать кнопку «Перейти...»



3. В окне «Надстройки» поставить галочку «Пакет анализа» и нажать «OK».



4. В результате на Ленте на вкладке «Данные» должна появиться кнопка «Анализ данных».

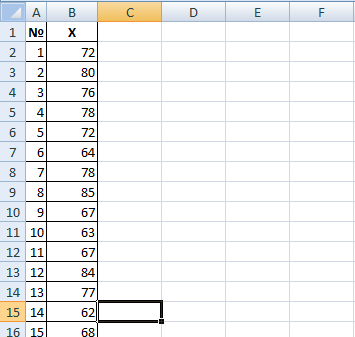


### Подготовка исходных данных

Создайте новый файл Excel и сохраните под именем вида «КАСД Л.р. ФИО группа.xlsx».

Переименуйте «Лист 1» в «Л.р.1».

Скопируйте данные для своего варианта на лист «Л.р.1», начиная со второй строки и столбца B. Добавьте в столбец A нумерацию наблюдений. Укажите названия столбцов.



|  |  |
| --- | --- |
|  | Используйте автозаполнение. |

### Пример

Все расчеты показаны на примере выборки:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 72 | 85 | 68 | 68 | 83 |
| 80 | 67 | 67 | 71 | 72 |
| 76 | 63 | 78 | 64 | 66 |
| 78 | 67 | 67 | 75 | 75 |
| 72 | 84 | 69 | 64 | 77 |
| 64 | 77 | 66 | 64 | 84 |
| 78 | 62 | 66 | 69 | 77 |

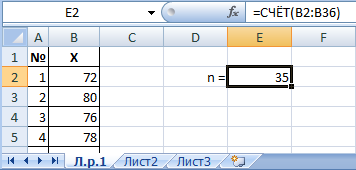
## Указания к выполнению работы

Требуемые в задании расчеты можно выполнить несколькими способами. Мы рассмотрим каждый из них и сравним результаты.

1. Определим количество значений в выборке, т.е. ее **мощность**. Фактически, пронумеровав наблюдения, мы уже узнали их количество. Но есть и формула СЧЁТ, по которой можно узнать количество ячеек в диапазоне.

Все результаты расчетов будем располагать в столбик справа от выборки.

|  |  |
| --- | --- |
|  | Функция СЧЁТ(диапазон) вычисляет количество непустых ячеек в выделенном диапазоне. Диапазон можно ввести вручную или выделить мышкой во время набора формулы. |

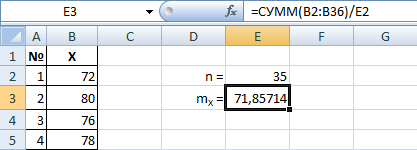


В примере выборка содержит 35 наблюдений.

1. Вычислим **выборочное среднее** *m*X. Сначала сделаем это в соответствии с формулой:

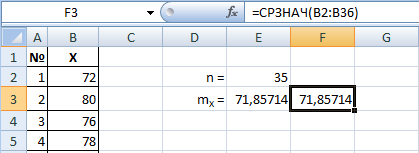


|  |  |
| --- | --- |
|  | Используйте функцию СУММ(*значения*). |



|  |  |
| --- | --- |
|  | Чтобы проставить нижний индекс x, при вводе текста выделите только x, правый клик – Формат ячеек – на вкладке Шрифт поставить галочку «Нижний индекс». |

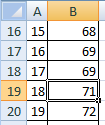
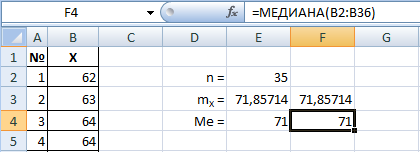
Существует и встроенная формула СРЗНАЧ(диапазон). Она должна дать такой же результат.



1. Найдем **медиану**. Это значение должно находиться в середине упорядоченной выборки. В данном примере середина – это 18-ое значение. Необходимо лишь упорядочить выборку.

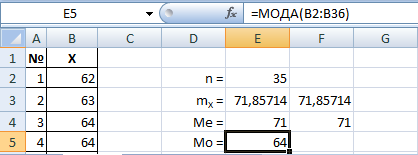
|  |  |
| --- | --- |
|  | Выделите всю выборку без заголовка и на ленте, на вкладке «Главная» нажмите «Сортировка и фильтр» – «Сортировка от минимального к максимальному», или кликните по выборке правой кнопкой и найдите этот пункт во всплывающем меню.    Примечание. В некоторых реальных случаях порядок значений в исходной выборке может быть важен. Тогда лучше сделать копию исходной выборки и отсортировать ее. В данном примере это не нужно. |

В примере 18-ое наблюдение равно 71. Скопируем его столбец результатов. Встроенная формула для вычисления – МЕДИАНА(диапазон) – должна вернуть то же значение.

В данном примере медиана мало отличается от среднего значения, что свидетельствует об однородности выборки (нет очень больших или очень маленьких значений, или их поровну).

1. Найдем **моду**. Чтобы найти моду вручную, нужно посчитать, сколько раз встречается каждое значение в выборке. То, которое встречается чаще всех и есть мода. Но поскольку это достаточно долго, даже по упорядоченной выборке, сразу воспользуемся формулой МОДА(диапазон).

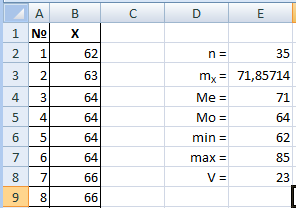
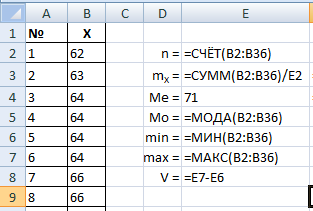


Т.е. наиболее вероятное значение в данной выборке – 64, оно появляется чаще других. Мода немного меньше среднего значения.

**Отчет**

Добавьте в отчет скриншот с полученными результатами. Поясните, что означают полученные числа (расшифруйте обозначения, приведите формулы). В выводе укажите, большая ли у вас выборка, охарактеризуйте соотношение между средним, модой и медианой.

1. Найдем **размах вариации**. Это разница между максимальным и минимальным значениями выборки. Поскольку мы уже отсортировали выборку, это первое и последнее значения, но мы воспользуемся формулами МИН(диапазон) и МАКС(диапазон).



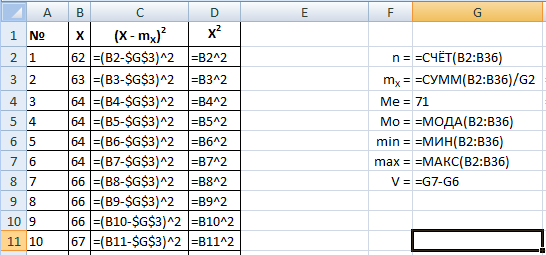
Таким образом, амплитуда колебаний рассматриваемого показателя равна 23.

1. Вычислим дисперсию по двум формулам:
2. 
3.  

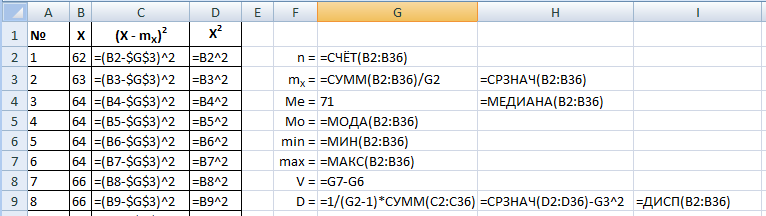
В этих формулах нам понадобятся промежуточные расчеты:  и . Добавьте два столбца справа от X.

|  |  |
| --- | --- |
|  | Правый клик по названию столбца – «Вставить» (без пиктограммы). |

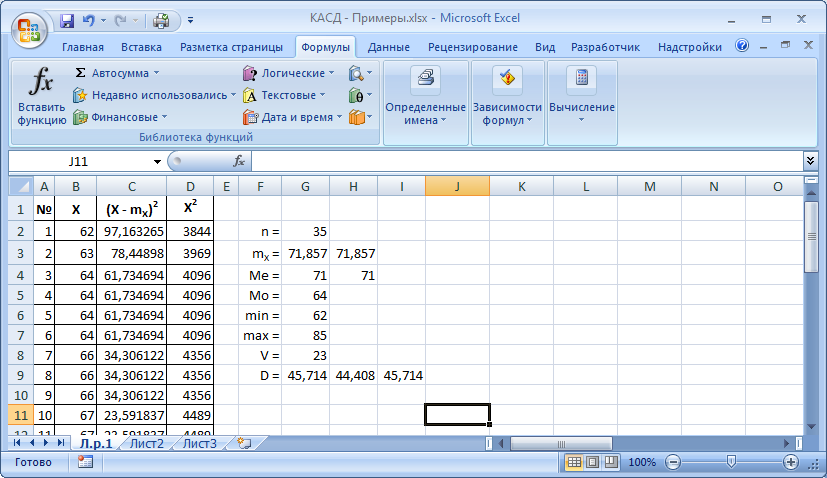
Добавьте названия столбцов и заполните их формулами с помощью «растягивания».



Добавляем обе формулы дисперсии и встроенную формулу – ДИСП(диапазон).



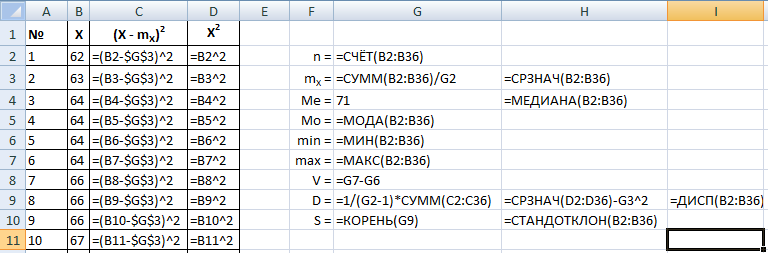
Результат:



Как видим, результат по второй формуле отличается, потому что она приближенная. В исходной формуле мы делим на (*n*–1), а в приближенной при расчете средних – на *n*.

Делать выводы по значению дисперсии трудно, поскольку она возводится в квадрат.

1. Вычислим стандартное отклонение двумя способами: извлечем корень из дисперсии (дисперсию берем по первой формуле) и воспользуемся функцией СТАНДОТКЛОН(диапазон):

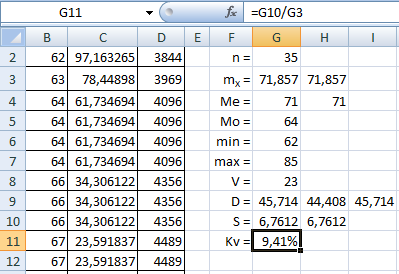




Таким образом, стандартное отклонение равно 6,76, что не очень много по сравнению со средним.

1. Вычислим коэффициент вариации как отношение S/m:



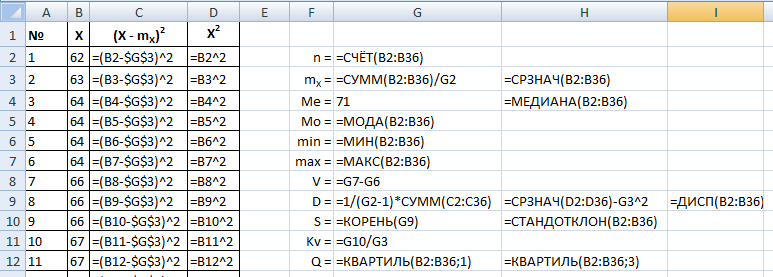


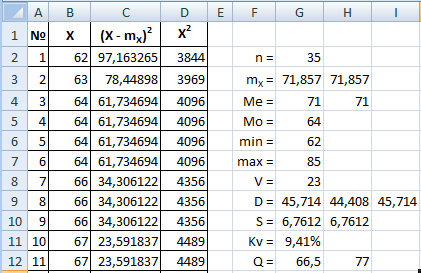
|  |  |
| --- | --- |
|  | Чтобы показать результат в процентах – правый клик по ячейке – «Формат ячейки» – вкладка «Число» – «Процентный» – «ОК». |

Коэффициент вариации составляет менее 10%, разброс значений исследуемой величины небольшой.

1. Квартили вычислим с помощью встроенной формулы КВАРТИЛЬ(диапазон;часть).

|  |  |
| --- | --- |
|  | Параметр «часть» показывает, какой квартиль мы хотим получить: 1 - нижний (25%), 2 - медиана (50%); 3 - верхний (75%). |





Нижний квартиль равен 66,5, верхний – 77, т.е. половина значений выборки лежит в диапазоне от 66,5 до 77.

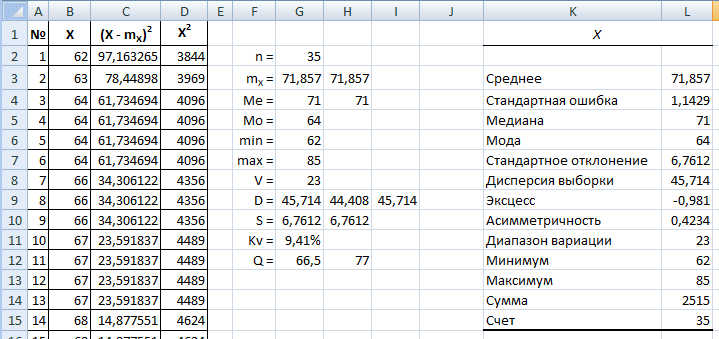
**Отчет**

Добавьте в отчет скриншот с расчетами, определяющими разброс значений выборки. Поясните смысл полученных значений, приведите формулы. В выводе охарактеризуйте разброс значений (слабый, средний, сильный), укажите в каких пределах находится половина значений выборки.

1. Те же результаты можно получить с использованием пакета «Анализ данных» (как его включить, см. раздел Подготовка программы).

|  |  |
| --- | --- |
|  | На вкладке «Данные» нажмите «Анализ данных». В окне выберите «Описательная статистика»    Укажите диапазон исходной выборки вместе с заголовком “X”, отметьте галочкой «Итоговая статистика» и «Метки в первой строке». Выходной диапазон укажите справа от ваших расчетов. Нажмите ОК. |

Результат:



|  |  |
| --- | --- |
|  | Из-за ошибки в Excel вместо «Диапазон вариации» может быть написано ССЫЛКА#. Исправьте текст вручную. |

Коэффициент эксцесса немного меньше нуля, выборка слегка пологая. Коэффициент асимметрии близок к нулю, распределение можно считать симметричным.

**Отчет**

Добавьте скриншот расчетов через «Анализ данных». Сравните результаты с расчетами, выполненными вручную. В выводе охарактеризуйте асимметрию и эксцесс.

В выводе должны присутствовать ответы на следующие вопросы:

* 1. Какова мощность выборки?
  2. Чему равно среднее значение? наиболее ожидаемое значение? Близки ли они между собой? Симметрична выборка по значениям среднего и медианы?
  3. Велик ли разброс значений? Чему он равен?
  4. В каких пределах находится половина значений выборки?
  5. Какова форма распределения показателя? Симметричная, остро- или плосковершинная?